Bounded-load calibration splits for differentially private graph conformal prediction

Patrick Indri

Machine Learning Research Unit, TU Wien, Vienna, Austria
patrick.indri@tuwien.ac.at

Abstract. We investigate differentially private conformal prediction on graph-structured data. Conformal prediction offers reliable uncertainty estimates, but assumes that the calibration scores can be treated as isolated records. In transductive node classification, scores are obtained from individual nodes, which are not isolated records. Many graph learning methods, such as message-passing graph neural networks, propagate information across edges and thus a node can affect the score of every calibration node within its receptive field. *Private* conformal prediction must therefore consider the entire set of affected nodes jointly. We show that private graph calibration can be framed as a structural problem: choosing a calibration set that is large enough to provide useful calibration, yet sparse enough that the privacy noise remains manageable. We formalize this problem and show that it is NP-hard. We propose a greedy heuristic that greatly reduces the graph-dependent sensitivity term on benchmark graph datasets.

Keywords: Conformal Prediction · Privacy · Graph Neural Networks

1 Introduction

Conformal prediction provides prediction sets with finite-sample coverage guarantees after calibrating a score threshold on held-out data. Differentially private (DP) conformal calibration makes this thresholding step compatible with privacy constraints by replacing exact calibration statistics with private ones. These mechanisms usually treat the calibration scores as isolated quantities. In transductive graph learning this treatment is incomplete: message-passing graph neural networks (GNNs) propagate information across edges, so one private node record can perturb the conformity scores of several nearby calibration nodes. Thus, the unit protected by DP and the unit counted by conformal calibration no longer coincide. This paper studies the post-hoc private calibration problem for fixed transductive GNNs. Despite active work on both private and graph conformal prediction, their intersection has remained underexplored. Our main observation is that the privacy cost of calibration is controlled by the *calibration load* of the calibration split, that is, the largest number of calibration scores that a single record can jointly perturb. Hence calibration selection is also a privacy-relevant graph selection problem. We make three main contributions.

First, we prove that the sensitivity of noisy-count private conformal calibration scales with the calibration load, and that this bound is tight. Second, we recast calibration-set selection as a bounded-degree induced-subgraph problem, and show it is NP-hard in general. Third, we introduce a privacy-calibration ratio that captures the size–load trade-off, derive upper and lower bounds, and give a greedy splitting heuristic that approaches these bounds in practice.

2 Related work

Our work sits at the intersection of private conformal prediction, graph conformal prediction, and private graph learning. Differentially private conformal prediction combines split-conformal calibration with private quantile estimation and conservative corrections [1,12,18], building on a line of work on private quantile estimation via the exponential mechanism and its refinements [14,6,10,3]. The only prior treatment of private conformal prediction on structured data we are aware of studies the local differential privacy setting with an untrusted aggregator [11], which does not address the cross-node dependencies induced by message passing. Graph conformal prediction, in turn, studies prediction sets for transductive node classification and uses graph structure to improve efficiency through score diffusion, topology-aware correction, or neighbourhood-based aggregation [19,8,15], but in the non-private setting. Privacy for graph learning has instead focused on private *training* of GNNs, where the per-layer sensitivity of message passing drives the privacy cost [16,20]. However, the privacy of the conformal *calibration* step on graphs is unexplored, and, to our knowledge, no existing private conformal method addresses the transductive setting we consider. Rather than proposing a new private quantile mechanism or conformity score, we isolate the sensitivity term that message passing introduces into private calibration: one changed record perturbs a whole neighbourhood of scores which affects the noise scale.

3 Preliminaries

We recall here the notation and notions used throughout the paper. Additional, more extensive preliminaries are available in Appendix A.

Conformal prediction [13]. Consider a classification problem with input space $\mathcal{X}$ and label space $\mathcal{Y}$. A score function $R : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ measures how well a candidate label $y \in \mathcal{Y}$ conforms to an input $x \in \mathcal{X}$. For a threshold τ, the corresponding prediction set is $C_\tau(x) := \{y \in \mathcal{Y} : R(x, y) \leq \tau\}$. In the standard split-conformal setting, one computes calibration scores $(R_1, \ldots, R_m)$ on a held-out calibration split. If $(R_1, \ldots, R_m, R_{m+1})$ are exchangeable, i.e., their joint distribution is invariant under permutation, and if $R_{(1)} \leq \cdots \leq R_{(m)}$ are the ordered calibration scores, then the threshold $\tau = R_{(k)}$ with $k = \lceil \gamma(m + 1) \rceil$ satisfies $\mathbb{P}(R_{m+1} \leq \tau) \geq \gamma$. Thus, the resulting prediction set satisfies $\mathbb{P}(Y \in C_\tau(X)) \geq \gamma$, and we refer to $\gamma \in (0, 1)$ as the target coverage level. We use this monotone threshold form of split-conformal prediction.

Graphs. Let $G = (V, E)$ be an undirected graph with node set V, $n = |V|$ nodes, and edge set E. For nodes $u, v \in V$, let $d_G(u, v)$ denote the shortest-path distance between u and v in G. For an integer $K \geq 0$, the K-hop ball around a node $u \in V$ is $B_K(u) := \{v \in V : d_G(u, v) \leq K\}$. We use G^K to denote the K-th graph power, with an edge between distinct nodes $u, v \in V$ whenever $d_G(u, v) \leq K$. For $S \subseteq V$, we write $G^K[S]$ for the subgraph of G^K induced by S, and write $\Delta(H)$ for the maximum degree of a graph H. A set $S \subseteq V$ is K-separated if $d_G(u, v) > K$ for all distinct $u, v \in S$. Equivalently, S is an independent set in G^K. We use K to denote the effective graph radius of a local score map. The formal locality assumption is stated in Assumption 1.

Differential privacy [5]. We consider standard (ϵ, δ)-differential privacy [4,5], denoted (ϵ, δ)-DP, with adjacency specified in Section 4. Here $\epsilon \geq 0$ is the *privacy budget*, where smaller values give stronger privacy, and $\delta \geq 0$ is the *failure parameter* for DP; the formal definition is recalled in Appendix A. For adjacent datasets $D \sim D'$ and a deterministic query f, the ℓ_1 *global sensitivity* of f is $GS_{f,\ell_1} = \max_{D \sim D'} \|f(D) - f(D')\|_1$. The Laplace mechanism [5] adds independent Laplace noise calibrated to this sensitivity. In particular, releasing $f(D) + Z$, with independent coordinates $Z_j \sim \mathrm{Lap}\left(\frac{GS_{f,\ell_1}}{\epsilon}\right)$ satisfies *pure* ϵ-DP, i.e., (ϵ, δ)-DP for $\delta = 0$.

4 Private calibration in the transductive setting

In this section, we present our analysis of private conformal calibration for transductive graph learning. We specify the privacy model, analyze the sensitivity of the noisy cumulative counts used for private calibration, and use this bound to define a private threshold with a coverage guarantee. Throughout, the graph topology is fixed; nevertheless, changing one calibration record, i.e., one calibration node, may affect more than one calibration score. The following locality assumption makes this dependence precise.

Assumption 1 (K-local score map). *For every node $i \in V$, the score $R_i(G)$ depends only on the data contained in $B_K(i)$. Equivalently, if two graphs differ only at a node u, then R_i can change only when $i \in B_K(u)$.*

Assumption 1 holds for K-layer message-passing GNNs when the model parameters are fixed. If a GNN is trained on the same private graph, changing one node can also change the learned parameters and the training procedure itself needs to be private; here we isolate the post-hoc private calibration problem instead. Architectures whose receptive field is not K-local—i.e., models with global readout or pooling layers, virtual nodes, or full self-attention as in graph transformers—violate Assumption 1; the effective radius is, in this case, the graph diameter.

4.1 Privacy model and noisy-count query

In our setting, we consider a fixed public graph $G = (V, E)$ with a fixed public calibration index set $S \subseteq V$ with $m = |S|$ and a fixed trained classifier. We focus

on protecting the privacy of the calibration nodes. For each calibration node $i \in S$, the private record is the tuple $z_i := (x_i, y_i)$, where x_i denotes the node attributes used by the fixed classifier and y_i denotes the calibration label; if only the calibration labels are private, one may instead take $z_i = y_i$. Two calibration datasets $Z_S = (z_i)_{i \in S}$ and $Z'_S = (z'_i)_{i \in S}$ are adjacent if they differ in exactly one calibration record. That is, there exists $u \in S$ such that $z_u \neq z'_u$ and $z_i = z'_i$ for all $i \neq u$. With fixed topology, adjacent calibration datasets are viewed as graphs G and G' with the same node and edge sets, differing only in one private calibration record $u \in S$. We refer to this notion of adjacency as *calibration-node attribute adjacency*. Although the graph topology is fixed, the score map is transductive and K-local: changing z_u can change not only the score R_u, but also the score of every calibration node i whose receptive field contains u, which for an undirected graph is equivalent to $i \in B_K(u)$. Thus, one adjacent change can affect all the calibration scores in $B_K(u) \cap S$, and private calibration depends on the local geometry of the calibration split.

Given calibration scores on a held-out set of nodes, we aim to obtain *private* conformal quantiles by releasing noisy cumulative counts on a fixed public grid, the graph analogue of the standard noisy-quantile mechanism of Angelopoulos et al. [1]. For each calibration node $i \in S$, let $R_i \in [0, 1]$ denote its score. For some integer $B > 0$, a publicly available grid is a sequence $0 < t_1 < \cdots < t_B \leq 1$ in the $[0, 1]$ interval. From the grid, we define the cumulative count vector $N(G) := (N_1(G), \ldots, N_B(G))$, with $N_b(G) := \sum_{i \in S} \mathbf{1}\{R_i(G) \leq t_b\}$ for $b \in \{1, \ldots, B\}$ where $\mathbf{1}$ is the indicator function. The private calibration method under analysis releases a noisy version of $N(G)$, and then post-processes this noisy vector to compute conformal thresholds. The key research question is therefore: *How can we determine the sensitivity of $N(G)$ under graph adjacency and therefore calibrate the amount of noise to achieve DP?*

4.2 Calibration load and graph-local sensitivity

The graph quantity which controls the behaviour of private calibration is the maximum number of calibration scores that can be affected by one private calibration node. We refer to this quantity as the *calibration load* of the procedure.

Definition 1 (Calibration load). *For a calibration set $S \subseteq V$, define the calibration load as*

$$\Delta_K^{\mathrm{cal}}(S) := \max_{u \in S} |B_K(u) \cap S|. \tag{1}$$

Figure 1 illustrates the notion captured by Definition 1. Our main analysis protects the privacy of calibration nodes only; if the attributes of *all* nodes are to be protected, the same analysis applies with the maximum in Definition 1 taken over $u \in V$ instead of $u \in S$. Under Assumption 1, we bound a local notion of the sensitivity as follows.

Theorem 1 (Graph-local sensitivity). *Under Assumption 1, the cumulative count vector $N(G)$ has ℓ_1 sensitivity at most $\mathrm{GS}_1(N) \leq B\Delta_K^{\mathrm{cal}}(S)$ under calibration-node attribute adjacency.*

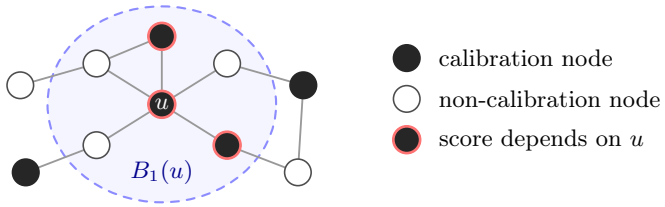

Fig. 1. Calibration load for $K = 1$. The black nodes form the calibration set S; the shaded region is the ball $B_1(u)$. Changing the private record of u can perturb the scores of the $|B_1(u) \cap S| = 3$ calibration nodes inside the ball; $\Delta_1^{\mathrm{cal}}(S)$ is the maximum of this count over calibration nodes.

Proof. Consider adjacent graphs G and G' differing at one calibration node $u \in S$. By Assumption 1, the only calibration scores that can change are those with indices in $B_K(u) \cap S$. Thus, for each grid point t_b,

$$|N_b(G) - N_b(G')| \le |B_K(u) \cap S| \le \Delta_K^{\mathrm{cal}}(S). \tag{2}$$

Summing over $b \in \{1, \dots, B\}$ gives $\|N(G) - N(G')\|_1 \le B\Delta_K^{\mathrm{cal}}(S)$. Taking the maximum over adjacent graphs proves the claim. □

The sensitivity bound identified in Theorem 1 is, in the worst case, tight.

Proposition 1 (Worst-case tightness). *Assume that $t_B < 1$. The dependence on $B\Delta_K^{\mathrm{cal}}(S)$ in Theorem 1 is worst-case tight, i.e., there exist K-local score maps and adjacent graphs G, G' such that $\|N(G) - N(G')\|_1 = B\Delta_K^{\mathrm{cal}}(S)$.*

Proof. See Appendix B.

4.3 Graph-adjusted private calibration

Theorem 1 gives the noise scale for private calibration on transductive graphs. Compared with the non-graph case, the privacy noise is inflated by the graph-local calibration load.

Corollary 1 (Graph-adjusted noisy-count calibration). *Let $\epsilon_{\mathrm{cal}} > 0$ be the privacy budget allocated to the calibration step, and let*

$$\tilde{N}_b = N_b + Z_b, \qquad Z_b \overset{\mathrm{iid}}{\sim} \mathrm{Lap}\left(\frac{B\Delta_K^{\mathrm{cal}}(S)}{\epsilon_{\mathrm{cal}}}\right). \tag{3}$$

Then releasing $(\tilde{N}_1, \dots, \tilde{N}_B)$ is ϵ_{cal}-DP under calibration-node attribute adjacency. Consequently, any threshold computed from these noisy counts is also ϵ_{cal}-DP by post-processing.

Proof. This follows directly from Theorem 1 and the Laplace mechanism [5]. Post-processing preserves DP [5]. □

The factor B in Corollary 1 comes from releasing all B cumulative counts with the Laplace mechanism, and more refined private quantile mechanisms can improve it; the graph-dependent part, however, is the calibration load $\Delta_K^{\mathrm{cal}}(S)$, so the split-design results below apply independently of the private quantile primitive used. We use the noisy counts to choose a conservative, private conformal threshold. Let $k = \lceil \gamma(m+1) \rceil$ be the usual split-conformal rank, where $m = |S|$. Non-private grid calibration would select the smallest t_b such that $N_b \geq k$; with noisy counts, positive noise can make a grid point appear to have enough calibration mass while its true count is below the target rank. We thus add a conservative rank correction offset $\lambda(S)$ and select the first grid point satisfying

$$\tilde{N}_b \geq k + \lambda(S). \tag{4}$$

If no grid point satisfies the inequality, we use the maximum possible threshold, so that the predictor returns all labels. For failure probability $\beta \in (0, 1)$, define now an *offset* $\lambda(S)$ as

$$\lambda(S) := \frac{B\Delta_K^{\mathrm{cal}}(S)}{\epsilon_{\mathrm{cal}}} \log\left(\frac{B}{\beta}\right). \tag{5}$$

By a union bound over the B Laplace variables, $\Pr[\max_b |Z_b| \leq \lambda(S)] \geq 1-\beta$, so with probability at least $1-\beta$ any grid point satisfying Equation (4) also has true count at least k: $\lambda(S)$ is the rank correction that compensates for the privacy noise. The dependence on the split is now explicit: the size $m = |S|$ improves conformal resolution, while the load $\Delta_K^{\mathrm{cal}}(S)$ increases the privacy correction. This motivates the following definition.

Definition 2 (Privacy–calibration ratio). *For a calibration set $S \subseteq V$, the privacy–calibration ratio is*

$$\rho_K(S) := \frac{\Delta_K^{\mathrm{cal}}(S)}{|S|}. \tag{6}$$

Indeed, the private rank correction per calibration point satisfies

$$\frac{\lambda(S)}{|S|} = \frac{B \log(B/\beta)}{\epsilon_{\mathrm{cal}}} \rho_K(S). \tag{7}$$

Thus, the graph split objective is not to minimize $\Delta_K^{\mathrm{cal}}(S)$ alone, but to trade calibration load against calibration size.

4.4 Coverage of the private threshold

The noisy thresholds of Section 4.3 yield an end-to-end DP *coverage* guarantee. Let $\tilde{N}_b$ be the noisy counts in Corollary 1 and $k = \lceil \gamma(m+1) \rceil$. We define the private threshold as

$$\hat{t} := \min\{t_b : \tilde{N}_b \geq k + \lambda(S)\}, \tag{8}$$

for an offset $\lambda(S)$ as per Equation (5), and with the convention $\hat{t} := 1$ if no grid point meets the inequality, i.e., the predictor then returns all labels. We can now provide guarantees on the coverage of the private thresholds.

Theorem 2 (Private graph-conformal coverage). *Consider a calibration set S and an unseen test point X_{test} with score R_{test}. Assume that $k \leq m$, that the scores $\{R_i\}_{i\in S} \cup \{R_{\text{test}}\}$ are exchangeable, and consider noisy counts with Laplace scale $B\Delta_K^{\text{cal}}(S)/\epsilon_{\text{cal}}$ as in Corollary 1. The threshold $\hat{t}$ in Equation (8) defines the prediction set $\widehat{C}(X_{\text{test}}) := \{y : R(X_{\text{test}}, y) \leq \hat{t}\}$ which satisfies*

$$\Pr[R_{\text{test}} \leq \hat{t}] \geq \gamma - \beta. \tag{9}$$

Proof. Consider the event $\mathcal{E} := \{\max_b |Z_b| \leq \lambda(S)\}$. This is the event that all Laplace noise variables are small in absolute value. For $Z_b \sim \text{Lap}(B\Delta_K^{\text{cal}}(S)/\epsilon_{\text{cal}})$ it holds that $\Pr[|Z_b| > \lambda(S)] = \beta/B$ by Equation (5). Taking a union bound we obtain that $\Pr[\mathcal{E}] \geq 1-\beta$. On $\mathcal{E}$, any grid point with $\tilde{N}_b \geq k+\lambda(S)$ has true count $N_b \geq \tilde{N}_b - \lambda(S) \geq k$. Hence whenever $\hat{t}$ is finite it satisfies $|\{i \in S : R_i \leq \hat{t}\}| \geq k$, i.e. $\hat{t} \geq R_{(k)}$, the k-th smallest calibration score; the fallback $\hat{t} = 1 \geq R_{(k)}$ also gives $\hat{t} \geq R_{(k)}$. The event $\mathcal{E}$ depends only on the privacy noise and is independent of the data. Exchangeability gives $\Pr[R_{\text{test}} \leq R_{(k)}] \geq k/(m+1) \geq \gamma$. Therefore

$$\Pr[R_{\text{test}} \leq \hat{t}] \geq \Pr[\{R_{\text{test}} \leq R_{(k)}\} \cap \mathcal{E}] \geq \gamma - \beta,$$

which concludes the proof. □

Rounding $\hat{t}$ up to a grid point can only enlarge the prediction set, so the grid resolution B costs efficiency but never coverage, and the split trade-off of Equation (7) reappears: a larger $|S|$ shrinks the relative slack $\lambda(S)/|S|$, while a larger load inflates it. While the sensitivity results hold for any fixed calibration set, the coverage guarantee additionally assumes that the selected calibration scores are exchangeable with the test score. Topology-aware split selection can bias the calibration population—purely structurally, since our split rules (Section 5) use only the graph and never labels or scores—so bounded-load splitting should be read as controlling the privacy cost of a chosen split, with coverage following whenever that split is exchangeable with the intended test population.

5 Calibration split selection as subgraph selection

The previous section shows that the private rank correction depends on the split through $\rho_K(S)$: larger calibration sets improve conformal resolution, while a larger load increases privacy noise. We now show that this trade-off is a subgraph selection problem on the K-th graph power G^K. Choosing a calibration set $S \subseteq V$ selects the induced subgraph $G^K[S]$, whose edges connect calibration nodes that one private record can jointly affect; controlling the privacy load thus amounts to controlling the maximum degree of $G^K[S]$. For a load budget $D \geq 1$, we consider the split selection problem

$$\max_{S\subseteq V} |S| \qquad \text{subject to} \qquad \Delta_K^{\text{cal}}(S) \leq D.$$

Equivalently, changing one private calibration record can affect at most D calibration scores.

Proposition 2 (Influence graph characterization). *For every calibration set $S \subseteq V$,*

$$\Delta_K^{\mathrm{cal}}(S) = 1 + \Delta(G^K[S]), \tag{10}$$

where $\Delta(G^K[S])$ is the maximum degree of the induced graph power.

Proof. For $u \in S$, the quantity $|B_K(u) \cap S|$ counts u itself plus all calibration nodes within distance at most K. These nodes are exactly u and the neighbors of u in $G^K[S]$. Taking the maximum over $u \in S$ gives the result. □

By Proposition 2, the split-selection problem is equivalent to

$$\max_{S \subseteq V} |S| \qquad \text{subject to} \qquad \Delta(G^K[S]) \leq D - 1. \tag{11}$$

Thus, privacy-aware calibration splitting is a bounded-degree induced-subgraph problem on G^K. The case $D = 1$ is especially simple.

Corollary 2 (K-separated splits). *A calibration set S satisfies $\Delta_K^{\mathrm{cal}}(S) = 1$ if and only if S is an independent set in G^K. Equivalently, $d_G(u, v) > K$ for all distinct $u, v \in S$. For such splits, graph-adjusted private calibration recovers the ordinary non-graph noise scale $B/\epsilon_{\mathrm{cal}}$.*

Proof. The condition $\Delta_K^{\mathrm{cal}}(S) = 1$ means that no calibration node has another in its K-hop ball, i.e., that no two nodes of S are adjacent in G^K; the final claim follows from Corollary 1. □

The graph-power formulation shows that optimal split selection is hard.

Proposition 3 (Computational hardness). *For $D = 1$, finding the largest calibration set with $\Delta_K^{\mathrm{cal}}(S) \leq D$ is NP-hard.*

Proof. See Appendix B.

The reduction in Proposition 3 is straightforward: for $K = 1$ and $D = 1$ the feasible sets are exactly the independent sets of G. This equivalence also transfers approximation results: maximum independent set is inapproximable on general graphs [9,21], but is constant-factor approximable by a greedy algorithm on bounded-degree graphs [7], the relevant regime for sparse networks. Propositions 5 and 9 give the corresponding guarantees on G^K. Proposition 3 motivates a simple ordered greedy heuristic. The parameter D controls the privacy load. $D = 1$ gives a greedy K-separated split; a larger D relaxes separation and can retain more calibration nodes. We describe one possible approach in Algorithm 1.

Proposition 4 (Feasibility of the greedy split). *Algorithm 1 returns a calibration set S satisfying $\Delta_K^{\mathrm{cal}}(S) \leq D$.*

Algorithm 1: Greedy bounded-load calibration split

Input: $G = (V, E)$, candidate pool $V_{\text{pool}} \subseteq V$, radius K, load budget D.
Output: Calibration set S.
$S \leftarrow \emptyset$.
Choose an ordering $(v_1, \ldots, v_{|V_{\text{pool}}|})$ of V_{pool}.
for $r = 1, \ldots, |V_{\text{pool}}|$ **do**
  $S' \leftarrow S \cup \{v_r\}$.
  $\Delta_K^{\text{cal}}(S') \leftarrow \max_{u \in S'} |B_K(u) \cap S'|$.
  if $\Delta_K^{\text{cal}}(S') \leq D$ **then**
    Set $S \leftarrow S'$.
return S.

Proof. The algorithm starts from the empty set and accepts a node only when the load constraint still holds, so the invariant $\Delta_K^{\text{cal}}(S) \leq D$ is maintained throughout and holds for the returned set. □

Proposition 5 (Expected size for $D = 1$). *For $D = 1$, run Algorithm 1 with a uniformly random ordering of $V_{\text{pool}} = V$. Then the expected size of the returned K-separated set is at least $\sum_{v \in V} \frac{1}{|B_K(v)|}$.*

Proof. For $D = 1$, the algorithm is the greedy independent-set algorithm on G^K. A node v is certainly selected if it appears before every node in its closed neighbourhood in G^K. This event has probability $1/(\deg_{G^K}(v) + 1) = 1/|B_K(v)|$. Therefore the selection probability of v is at least $|B_K(v)|^{-1}$. Summing over $v \in V$ gives the claim. □

Allowing $D > 1$ is useful only when the increase in calibration-set size compensates for the larger allowed load, as measured by the ratio $\rho_K(S) = \Delta_K^{\text{cal}}(S)/|S|$.

6 Privacy–calibration trade-off

The achievable trade-off summarized by $\rho_K(S)$ is controlled by graph structure: a good split contains many calibration nodes without placing too many in the same K-hop neighbourhood. In this section we show that random splits can be expensive on graphs with hubs, i.e., nodes with large degree or large K-hop balls, while separated splits control the load by construction. First, a lower bound: local clustering among selected calibration nodes necessarily increases ρ_K.

Proposition 6 (Counting lower bound). *For every calibration set $S \subseteq V$ it holds that*

$$\rho_K(S) \geq \frac{1}{|S|^2} \sum_{s \in S} |B_K(s) \cap S|. \tag{12}$$

Proof. We average the number of selected calibration nodes contained in each selected node's K-hop ball. Since the maximum load is at least the average load,

$$\Delta_K^{\text{cal}}(S) = \max_{u \in S} |B_K(u) \cap S| \geq \frac{1}{|S|} \sum_{u \in S} |B_K(u) \cap S|. \tag{13}$$

Dividing by $|S|$ yields the claim. □

We next compare random and separated splits: a random split is simple, but its load is controlled by the largest K-hop ball.

Proposition 7 (Random split baseline). *Let S be obtained by including each node independently with probability p. Let $b_K^{\max} := \max_{u\in V}|B_K(u)|$. Then, with probability at least $1-\eta$,*

$$\max_{u\in V}|B_K(u)\cap S| \le pb_K^{\max} + \sqrt{2pb_K^{\max}\log(n/\eta)} + \frac{2}{3}\log(n/\eta). \tag{14}$$

Proof. See Appendix B.

Thus, on graphs with hubs or large K-hop balls, the private calibration noise is driven by the worst local load rather than by the average calibration size. The opposite extreme is to enforce K-separation, which gives $\Delta_K^{\mathrm{cal}}(S) = 1$; the following bounds show how large such sets can be.

Proposition 8 (Separated split upper bound). *Assume G has maximum degree at most Δ and $K \ge 1$. Let S be any maximal K-separated set. Then $|S| \ge \frac{n}{M_K(\Delta)}$, where $M_K(\Delta) := 1 + \Delta\sum_{r=0}^{K-1}(\Delta-1)^r$. Consequently, $\rho_K(S) \le \frac{M_K(\Delta)}{n}$.*

Proof. Since S is maximal, every node in V lies within distance at most K of some node in S. Thus the K-hop balls around nodes in S cover V. Each such ball has size at most $M_K(\Delta)$. Therefore $n \le |S|M_K(\Delta)$. Since S is K-separated, $\Delta_K^{\mathrm{cal}}(S) = 1$. Thus $\rho_K(S) = 1/|S| \le M_K(\Delta)/n$. □

The bound in Proposition 8 depends on the worst degree; a sharper guarantee follows from the actual K-hop ball sizes.

Proposition 9 (Caro–Wei achievability). *There exists a K-separated calibration set S with $|S| \ge \sum_{v\in V}\frac{1}{|B_K(v)|}$. Consequently, there exists a split with $\rho_K(S) \le \left(\sum_{v\in V}\frac{1}{|B_K(v)|}\right)^{-1}$.*

Proof. A K-separated set is an independent set in G^K. Furthermore, $\deg_{G^K}(v) = |B_K(v)| - 1$. The Caro–Wei bound [2,17] gives an independent set of size at least

$$\sum_{v\in V}\frac{1}{\deg_{G^K}(v)+1} = \sum_{v\in V}\frac{1}{|B_K(v)|}.$$

Such a set has $\Delta_K^{\mathrm{cal}}(S) = 1$, so $\rho_K(S) = 1/|S|$. □

Together, these results explain why graph-aware splitting can matter. Random splits are controlled by worst-case local concentration, whereas separated splits exploit the distribution of local ball sizes. The bounded-load heuristic of Algorithm 1 interpolates between these extremes by allowing some local overlap while keeping the maximum load controlled.

6.1 Effective radius and topology-aware scores

The previous bounds depend on the effective graph radius of the conformity score, which matters because several graph conformal methods improve efficiency by using graph structure in the score. For example, DAPS smooths node-wise conformity scores over the graph [19], CF-GNN uses topology-aware correction [8], and SNAPS aggregates information from structurally or feature-similar nodes [15]. Our analysis gives a privacy-side interpretation of such operations: using a larger graph neighbourhood can improve non-private efficiency, but it also increases the number of calibration scores affected by one private record. For a K-layer local GNN without graph post-processing the effective radius is K; diffusing base scores over L additional graph steps makes it $K+L$. Consequently, the cumulative-count sensitivity satisfies $\mathrm{GS}_1(N) \leq B\Delta^{\mathrm{cal}}_{K+L}(S)$ and, for every fixed split S, it holds that $\Delta^{\mathrm{cal}}_K(S) \leq \Delta^{\mathrm{cal}}_{K+L}(S)$ and $\rho_K(S) \leq \rho_{K+L}(S)$. Thus, topology-aware score smoothing creates a privacy–efficiency trade-off: it may reduce prediction-set size, but it enlarges the graph neighbourhood over which one private calibration record can influence the calibration statistic.

7 Experiments

We include a minimal experimental setup that focuses only on graph structure to illustrate the split-dependent sensitivity term. Since no prior private graph-conformal method addresses our setting, we compare against a matched-size random-split baseline. The sensitivity bound in Theorem 1 suggests a simple empirical pattern: among splits of comparable size, random splits should have large calibration load when they concentrate calibration nodes in the same K-hop neighbourhood, while bounded-load greedy splits with uniformly random node orderings (Algorithm 1) keep $\Delta^{\mathrm{cal}}_K(S)$ small by construction, with larger D potentially retaining more nodes at the cost of a larger allowed load. Thus, we measure $\Delta^{\mathrm{cal}}_K(S)$ and $\rho_K(S)$, the graph quantities that determine the privacy noise scale in Theorem 1, on the Cora, CiteSeer and PubMed datasets; Table 1 reports the results. Code to reproduce all experiments is available on GitHub[1].

As Table 1 shows, matched-size random splits have substantially larger calibration load than bounded-load splits. For $K = 2$, the random matched-size split has ρ_K roughly 56, 60, and 133 times larger than greedy $D = 1$ on Cora, CiteSeer, and PubMed, respectively, supporting the main message that calibration splitting is not a neutral implementation detail under private graph calibration.

8 Discussion and conclusions

In this work, we discussed how private conformal calibration on graphs depends on the graph structure. For a fixed GNN, the calibration split is not a neutral implementation choice: one changed record perturbs every calibration score

[1] `https://github.com/pindri/private_graph_split_conformal`

Table 1. Topology-only comparison of calibration split strategies for $K = 2$. Each random split is matched to the size of the corresponding greedy $D = 1$ split. Bounded-load splits explicitly control $\Delta_K^{\mathrm{cal}}(S)$ and yield much smaller privacy-calibration ratio $\rho_K(S) = \Delta_K^{\mathrm{cal}}(S)/|S|$. Results are means and standard deviations over 20 random node orderings and matched random splits.

Dataset	Split	$\lvert S\rvert$	$\Delta_K^{\mathrm{cal}}(S)$	$\rho_K(S) \cdot 10^3$
Cora	Random matched-size	510 ± 8	56 ± 15	109.4 ± 28.1
	Greedy $D = 1$	510 ± 8	1	1.96 ± 0.03
	Greedy $D = 2$	733 ± 10	2	2.73 ± 0.04
	Greedy $D = 4$	964 ± 18	4	4.15 ± 0.08
CiteSeer	Random matched-size	939 ± 9	60 ± 13	63.5 ± 14.2
	Greedy $D = 1$	939 ± 9	1	1.06 ± 0.01
	Greedy $D = 2$	1514 ± 19	2	1.32 ± 0.02
	Greedy $D = 4$	1947 ± 19	4	2.06 ± 0.02
PubMed	Random matched-size	2374 ± 19	132 ± 54	55.9 ± 22.7
	Greedy $D = 1$	2374 ± 19	1	0.42 ± 0.00
	Greedy $D = 2$	3658 ± 32	2	0.55 ± 0.01
	Greedy $D = 4$	4969 ± 53	4	0.81 ± 0.01

in its K-hop ball, so the split sets both the number of calibration scores and the local privacy load at once. Balancing these two aspects—keeping calibration nodes spread out to limit the influence of individual records, while retaining enough nodes to estimate thresholds accurately—rather than optimizing either alone, is what makes bounded-load split selection a natural primitive for private graph conformal prediction. One limitation of our work is that we bound only the *privacy cost* of a split: the coverage guarantee in Theorem 2 is inherited from standard split-conformal prediction and holds only when the chosen split is exchangeable with the test population. Because our approach uses only the graph, it introduces no label- or score-dependent bias, but it may still shift the calibration population in ways we do not characterize; a better characterization is left to future work. An additional limitation concerns our experimental evaluation: our experiments isolate the structural quantities that drive the noise scale rather than measuring end-to-end private coverage and efficiency. A more thorough empirical investigation, as well as the inclusion of more refined private quantile estimators, is also a promising direction for future work.

Acknowledgements

This work was funded in part by the Austrian Science Fund (FWF), project NanOX-ML (6728).

References

1. Angelopoulos, A.N., Bates, S., Zrnic, T., Jordan, M.I.: Private prediction sets. Harvard Data Science Review **4**(2) (2022)
2. Caro, Y.: New results on the independence number. Tech. rep., Tel-Aviv University (1979)
3. Durfee, D.: Unbounded differentially private quantile and maximum estimation. Advances in Neural Information Processing Systems **36** (2023)
4. Dwork, C.: Differential privacy. In: International Colloquium on Automata, Languages, and Programming. Springer (2006)
5. Dwork, C., Roth, A.: The algorithmic foundations of differential privacy. Foundations and trends in theoretical computer science **9**(3-4) (2014)
6. Gillenwater, J., Joseph, M., Kulesza, A.: Differentially private quantiles. In: ICML. PMLR (2021)
7. Halldórsson, M., Radhakrishnan, J.: Greed is good: Approximating independent sets in sparse and bounded-degree graphs. In: Proceedings of the twenty-sixth annual ACM symposium on Theory of computing. pp. 439–448 (1994)
8. Huang, K., Jin, Y., Candes, E., Leskovec, J.: Uncertainty quantification over graph with conformalized graph neural networks. Advances in Neural Information Processing Systems **36** (2023)
9. Håstad, J.: Clique is hard to approximate within $n^{1-\epsilon}$. In: Proceedings of 37th Conference on Foundations of Computer Science. pp. 627–636. IEEE (1996)
10. Kaplan, H., Schnapp, S., Stemmer, U.: Differentially private approximate quantiles. In: ICML. PMLR (2022)
11. Penso, C., Mahpud, B., Goldberger, J., Sheffet, O.: Privacy-preserving conformal prediction under local differential privacy. In: Fourteenth Symposium on Conformal and Probabilistic Prediction with Applications (COPA 2025). PMLR (2025)
12. Romanus, O.M., Molinari, R.: Differentially private conformal prediction via quantile binary search. Transactions on Machine Learning Research (2026)
13. Shafer, G., Vovk, V.: A tutorial on conformal prediction. Journal of machine learning research **9**(3) (2008)
14. Smith, A.: Privacy-preserving statistical estimation with optimal convergence rates. In: Proceedings of the forty-third annual ACM symposium on Theory of computing (2011)
15. Song, J., Huang, J., Jiang, W., Zhang, B., Li, S., Wang, C.: Similarity-navigated conformal prediction for graph neural networks. Advances in Neural Information Processing Systems **37** (2024)
16. Wei, J., Zhu, Y., Xiao, X., Bao, E., Yang, Y., Cai, K., Ooi, B.C.: GCON: Differentially private graph convolutional network via objective perturbation. In: 2025 IEEE 41st International Conference on Data Engineering (ICDE). IEEE (2025)
17. Wei, V.K.: A lower bound on the stability number of a simple graph. Bell Laboratories Technical Memorandum Murray Hill, NJ, USA (1981)
18. Wu, J., Zhang, C., Cai, Z., Kong, J., Jiang, B., Kong, L., Kong, L.: Differentially private conformal prediction. arXiv preprint arXiv:2604.14621 (2026)
19. Zargarbashi, S.H., Antonelli, S., Bojchevski, A.: Conformal prediction sets for graph neural networks. In: ICML. PMLR (2023)
20. Zhang, Q., Lee, H.k., Ma, J., Lou, J., Yang, C., Xiong, L.: Dpar: Decoupled graph neural networks with node-level differential privacy. In: Proceedings of the ACM Web Conference 2024 (2024)
21. Zuckerman, D.: Linear degree extractors and the inapproximability of max clique and chromatic number. In: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing. pp. 681–690 (2006)

A Additional preliminaries

We report here extended preliminary notions used throughout the paper.

Conformal prediction. In classification, conformal prediction constructs set-valued predictors, which output sets of labels rather than single labels, with finite-sample coverage guarantees [13]. Conformal prediction relies on a score function $R(x, y)$ that measures how well a label y conforms to an input x. From the score function, in the standard *split-conformal* setting one computes calibration scores $(R_1, \ldots, R_m)$ on a held-out data split and uses them to compute a threshold τ, i.e., a quantile of these scores. Commonly, split-conformal guarantees are obtained relying on the *exchangeability* assumption, where a sequence of random variables is *exchangeable* if its joint distribution is invariant under permutation. With this assumption, the rank of a new score R_{m+1} among $(R_1, \ldots, R_m, R_{m+1})$ is uniform on $\{1, \ldots, m+1\}$. Let $R_{(1)} \leq \cdots \leq R_{(m)}$ denote the sorted calibration scores, where the parenthetical subscript denotes the index of a score in the sorted list of scores. By choosing $k = \lceil \gamma(m+1) \rceil \leq m$ and $\tau = R_{(k)}$, it then follows that $\mathbb{P}(R_{m+1} \leq \tau) \geq \gamma$. The resulting prediction set $C_\tau(x) = \{y : R(x, y) \leq \tau\}$ thus satisfies $\mathbb{P}(Y \in C_\tau(X)) \geq \gamma$. The quantity γ is therefore commonly referred to as the *coverage* guaranteed by the conformal prediction. A number of different possible score functions can be defined [13]: here we rely on the commonly satisfied assumption that the chosen score function defines prediction sets $C_\tau(x) = \{y : R(x, y) \leq \tau\}$ for which $R(x, y) \leq \tau \implies y \in C_\tau(x)$, and where therefore C_τ is monotone in τ.

Graphs. Let $G = (V, E)$ be an undirected graph with node set V and edge set E. We write $n = |V|$ to denote the number of nodes of a node set V. For nodes $u, v \in V$, let $d_G(u, v)$ denote the shortest-path distance between u and v in G. For an integer $K \geq 0$, the K-hop ball around u is $B_K(u) := \{v \in V : d_G(u, v) \leq K\}$. Thus, $B_0(u) = \{u\}$. When the graph is clear from context, we omit the subscript G from the distance. We use G^K to denote the K-th graph power, that is, the graph on the same node set V with an edge between distinct nodes u and v whenever $d_G(u, v) \leq K$. For a subset $S \subseteq V$, we write $G^K[S]$ for the subgraph of G^K induced by S. The maximum degree of a graph H is denoted by $\Delta(H)$. A set $S \subseteq V$ is called K-separated if $d_G(u, v) > K$ for all distinct $u, v \in S$. Equivalently, S is an independent set in G^K. We consider transductive node classification with a fixed trained classifier that relies on message-passing, such as a message-passing GNN. We use K to denote the effective graph radius of a local score map. The formal locality assumption is stated in Assumption 1.

Differential privacy. Differential privacy (DP) [4] is a formal notion of privacy that protects sensitive information from privacy leakage. In this work, we focus on protecting the privacy of the nodes used in the computation of the conformal thresholds. DP is defined in terms of *adjacent datasets*, where a dataset is a collection of *records*. We view a graph dataset as a collection of node records. A randomized mechanism is said to be private if its output on adjacent datasets (see Section 4.1) does not leak sensitive information on the records.

Definition 3 (**(ϵ, δ)-DP, [5]**). *Let $\epsilon \geq 0$ and $\delta \geq 0$. A randomized mechanism $\mathcal{M} : \mathcal{X} \to \mathcal{Y}$ satisfies (ϵ, δ) differential privacy, denoted as (ϵ, δ)-DP, if, for all adjacent $D, D' \in \mathcal{X}$ and all $S \subseteq \mathrm{Range}(\mathcal{M})$ $\Pr[\mathcal{M}(D) \in S] \leq e^{\epsilon} \Pr[\mathcal{M}(D') \in S] + \delta$, where probabilities are taken over the randomness of $\mathcal{M}$. When $\delta = 0$, we say $\mathcal{M}$ satisfies* pure *ϵ-DP.*

We refer to ϵ as the *privacy budget* of a mechanism and to δ as the failure parameter. Larger values of ϵ provide less privacy, while $\epsilon = \delta = 0$ corresponds to identical output distributions on adjacent D, D'. For a given deterministic function f, one can build a private mechanism by means of additive noise calibrated to its global sensitivity $GS_{f,p} = \max_{D \sim D'} \|f(D) - f(D')\|_p$, where $\|\cdot\|_p$ is an ℓ_p-norm. A common approach is to add Laplace noise to the output of a given function to obtain DP.

Theorem 3 (Laplace mechanism for pure DP, [4,5]). *Let $f : \mathcal{X} \to \mathbb{R}$ have ℓ_1 sensitivity GS_{f,ℓ_1}. The randomized mechanism $\mathcal{M}(D) = f(D) + \mathrm{Lap}\left(\frac{GS_{f,\ell_1}}{\epsilon}\right)$ satisfies ϵ-DP, where $\mathrm{Lap}(b)$ denotes Laplacian noise with mean 0 and scale b.*

The Laplace mechanism extends directly to vector-valued queries $f : \mathcal{X} \to \mathbb{R}^d$; adding independent noise $Z_j \sim \mathrm{Lap}\left(GS_{f,\ell_1}/\epsilon\right)$ to each coordinate, where GS_{f,ℓ_1} is the ℓ_1 sensitivity of the entire vector, yields an ϵ-DP release. This is the form used for the cumulative count vector in Section 4.

B Missing proofs

Proposition 1 (Worst-case tightness). *Assume that $t_B < 1$. The dependence on $B\Delta_K^{\mathrm{cal}}(S)$ in Theorem 1 is worst-case tight, i.e., there exist K-local score maps and adjacent graphs G, G' such that $\|N(G) - N(G')\|_1 = B\Delta_K^{\mathrm{cal}}(S)$.*

Proof. Let $u \in S$ attain the maximum in the definition of $\Delta_K^{\mathrm{cal}}(S)$, and write $A := B_K(u) \cap S$ and $|A| = \Delta_K^{\mathrm{cal}}(S)$. Fix adjacent datasets that differ only in the record of u. We construct a K-local score map for which all scores indexed by A change from below the whole grid to above the whole grid, while all other calibration scores remain unchanged. Assume $t_B < 1$ and choose $a, b \in [0, 1]$ with $a \leq t_1$ and $b > t_B$. This is possible since $0 < t_1 < \cdots < t_B \leq 1$. Define a K-local score map such that, on the first dataset, $R_i = a$ for every $i \in A$, while on the adjacent dataset, $R_i = b$ for every $i \in A$. For $i \notin A$, keep the score fixed across the two datasets. This score map is K-local because only nodes whose K-hop ball contains u are allowed to change, and these include precisely the affected calibration nodes in A. For every $i \in A$ and every grid point t_b, the indicator $\mathbf{1}\{R_i \leq t_b\}$ changes from 1 to 0. Hence, for every $b \in \{1, \ldots, B\}$,

$$|N_b(G) - N_b(G')| = |A| = \Delta_K^{\mathrm{cal}}(S).$$

Summing over the B grid points gives

$$\|N(G) - N(G')\|_1 = \sum_{b=1}^{B} |N_b(G) - N_b(G')| = B\Delta_K^{\mathrm{cal}}(S).$$

This proves worst-case tightness. □

Proposition 3 (Computational hardness). *For $D = 1$, finding the largest calibration set with $\Delta_K^{\mathrm{cal}}(S) \le D$ is NP-hard.*

Proof. We reduce from maximum independent set. Consider the special case $K = 1$ and $D = 1$. Then $\Delta_1^{\mathrm{cal}}(S) \le 1$ means that, for every $u \in S$, $|B_1(u) \cap S| \le 1$. Since $u \in B_1(u)$, this condition holds if and only if no node $u \in S$ has a distinct neighbour $v \in S$. Equivalently, S is an independent set in G. Therefore, in the case $K = 1$ and $D = 1$, the problem

$$\max_{S \subseteq V} |S| \qquad \text{subject to} \qquad \Delta_1^{\mathrm{cal}}(S) \le 1$$

is exactly the maximum independent set problem on G. Since maximum independent set is NP-hard, the calibration split-selection problem is NP-hard. □

Proposition 7 (Random split baseline). *Let S be obtained by including each node independently with probability p. Let $b_K^{\max} := \max_{u \in V} |B_K(u)|$. Then, with probability at least $1 - \eta$,*

$$\max_{u \in V} |B_K(u) \cap S| \le p b_K^{\max} + \sqrt{2 p b_K^{\max} \log(n/\eta)} + \frac{2}{3} \log(n/\eta). \tag{14}$$

Proof. For each fixed $u \in V$, write $X_u := |B_K(u) \cap S| = \sum_{v \in B_K(u)} \xi_v$, where $\xi_v = \mathbf{1}\{v \in S\}$. Since nodes are included independently with probability p, the variables ξ_v are independent Bernoulli random variables. Thus, it holds that $X_u \sim \mathrm{Binomial}(|B_K(u)|, p)$ with mean $\mu_u = \mathbb{E}X_u = p|B_K(u)| \le p b_K^{\max}$. A Chernoff bound guarantees that, for every $t > 0$, $\Pr[X_u \ge \mu_u + t] \le \exp\left(-\frac{t^2}{2(\mu_u + t/3)}\right)$. Setting the exponent equal to a and solving $t^2 - \frac{2a}{3} t - 2a\mu_u = 0$, the positive root satisfies $t \le \sqrt{2\mu_u a} + \frac{2}{3} a$. It follows that, for every $a > 0$,

$$\Pr\left[X_u \ge \mu_u + \sqrt{2\mu_u a} + \frac{2}{3} a\right] \le e^{-a}.$$

Since the function $x \mapsto x + \sqrt{2xa}$ is increasing for $x \ge 0$, and since $\mu_u \le p b_K^{\max}$, we obtain

$$\Pr\left[X_u \ge p b_K^{\max} + \sqrt{2 p b_K^{\max} a} + \frac{2}{3} a\right] \le e^{-a}.$$

Set $a = \log(n/\eta)$. Then $e^{-a} = \eta/n$, so the displayed bound fails for this fixed u with probability at most η/n. Taking a union bound over all $u \in V$ gives

$$\Pr\left[\exists u \in V : X_u > p b_K^{\max} + \sqrt{2 p b_K^{\max} \log(n/\eta)} + \frac{2}{3} \log(n/\eta)\right] \le \eta.$$

Equivalently, with probability at least $1 - \eta$,

$$\max_{u \in V} |B_K(u) \cap S| \le p b_K^{\max} + \sqrt{2 p b_K^{\max} \log(n/\eta)} + \frac{2}{3} \log(n/\eta)$$

which concludes the proof. □