TURBOCHARGING GRAPH NEURAL NETWORKS WITH ZENTROPY

Anonymous authors
Paper under double-blind review

ABSTRACT

Graph neural networks (GNNs) learn by aggregating feature representations across connected nodes, a mechanism that excels when graph topology aligns with label consensus. In heterophilic graphs, however, the homophily inductive bias of GNNs often yields suboptimal predictions, as the learning dynamics overemphasize relational smoothness. The popular driving force of GNN is cross-entropy, an objective that minimizes information uncertainty via a *zero-temperature* energy descent. Because cross-entropy does not account for internal disparities during feature aggregation, current GNN architectures remain trapped in rigid energy regimes that cannot dynamically adapt to node-level structural divergence. To address this limitation, we reformulate GNN optimization through *zentropy*, a multi-scale thermodynamic framework under a *free-energy landscape* that explicitly balances internal energy (the likelihood associated with downstream tasks) with intrinsic disparities (a prior on data heterogeneity). In the spirit of thermodynamic principle, we devise a physics-informed neural network to discover the *critical temperature* for each graph node, creating a new degree of freedom that mitigates representational collapse caused by feature over-smoothing in GNNs. Together, we present a model-agnostic zentropy loss function in the regime of Helmholtz free energy that serves as a plug-and-play module compatible with standard GNN architectures. We evaluate the zentropy-based readout against matched cross-entropy baselines on seven encoders, three citation networks, five heterophilic benchmarks, three OGB graphs and five graph-classification datasets. Our proposed zentropy loss rescues deep encoders in GNN where standard cross-entropy optimization collapses to chance-level performance. Across all evaluated combinations of backbone architectures and network configurations, zentropy loss yields promising performance gains, most notably on the heterophilic graph datasets.

1 INTRODUCTION

Graph Neural Networks (GNNs) process graph data by iteratively passing messages across edges to refine node representations (Gilmer et al., 2017; Wu et al., 2020). This mechanism acts as a low-pass filter under an implicit assumption of *homophily* that connected nodes share similar latent states (McPherson et al., 2001). However, many real-world graphs exhibit *heterophily*, i.e., adjacent nodes routinely belong to distinct classes (Pei et al., 2020; Zhu et al., 2020). In heterophilic graph learning, accurate classification requires connected nodes to remain *separable*, a property standard aggregation operators actively erase. This fundamental tension worsens with network depth: repeated message passing inevitably collapses local signals, driving representations toward a uniform, non-discriminative fixed point in a phenomenon known as *over-smoothing* (Li et al., 2018). To mitigate these constraints, existing work spans diverse architectural paradigms: from foundational neighborhood aggregation in GCN (Kipf & Welling, 2017), GraphSAGE (Hamilton et al., 2017), and attention-based weighting in GAT/GATv2 (Veličković et al., 2018; Brody et al., 2022), to decoupled propagation frameworks like APPNP (Gasteiger et al., 2018), linearized message passing in SGC (Wu et al., 2019), and initial residual connections in GCNII (Chen et al., 2020). A more comprehensive review of related work is provided in Appendix A.1.

Following message passing, a GNN's linear readout assigns scalar scores to candidate labels using the *softmax* operator. This formulation is mathematically equivalent to a Gibbs–Boltzmann distribution (Grathwohl et al., 2019; LeCun et al., 2006): negative scores act as energies E, casting

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.

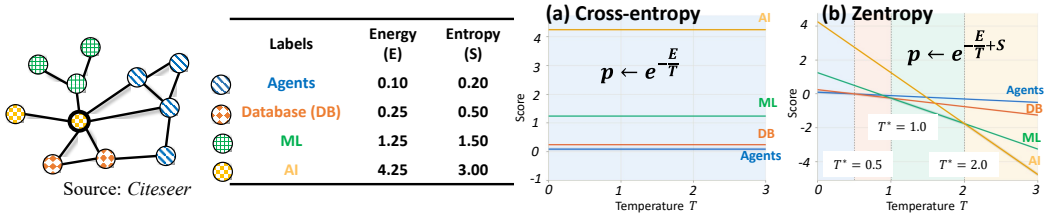

Figure 1: **Cross-entropy *vs* zentropy in GNNs.** For one node in a heterophilic CiteSeer neighborhood, each of four candidate labels carries an energy E (likelihood of label assignment) and an entropy S (prior on label variance). **(a)** Cross-entropy implicitly sets $S = 0$, rendering scores invariant to temperature T and freezing the prediction at paper category "*Agents*". **(b)** Zentropy evaluates candidate labels via a free energy $F = E - TS$. Since temperature scales E relative to S, score trajectories intersect at critical temperatures $T^* = \Delta E/\Delta S$, dynamically transitioning the predicted label across temperature regimes (*Agents* $\xrightarrow{T^*=0.5}$ *DB* $\xrightarrow{T^*=1.0}$ *ML* $\xrightarrow{T^*=2.0}$ *AI*).

classes as wells within a discrete energy landscape that training reshapes toward a ground-truth global minimum. Under standard cross-entropy, thermodynamic temperature is implicitly pinned to unity ($T = 1$) and intrinsic entropy is assumed to vanish ($S = 0$) (Guo et al., 2017). As shown in Fig. 1(a), this leaves class rankings invariant to temperature, where adjusting T only sharpens or flattens the probability distribution without ever altering the winning prediction (e.g., "*Agents*").

Zentropy theory (Liu et al., 2022; Wang et al., 2026) resolves this limitation by modeling target states as coarse-grained configurations with internal degrees of freedom. By attributing an *intrinsic entropy* to each class, entropy operates at two scales: within individual configurations and across the label distribution. Class rankings are thus governed by the Helmholtz free energy $F = E - TS$, discounting energy by internal freedom at rate T. Because each class carries a distinct entropy, their free-energy trajectories evolve non-uniformly with T, intersecting to trigger discrete rank inversions. As shown in Figure 1(b) on CiteSeer, while standard cross-entropy remains locked to *Agents*, zentropy induces sequential phase transitions (Agents $\rightarrow$ DB $\rightarrow$ ML $\rightarrow$ AI). Temperature thus transcends its role as a passive confidence scalar, becoming an active decision variable tailored for local neighborhoods lacking reliable evidence.

Translated to GNNs, energy measures label cost derived from message-passed evidence, while intrinsic entropy quantifies *local disparity*, the structural ambiguity induced by class heterogeneity. Temperature controls the balance, weighing neighborhood evidence against residual disorder. In homophilic regions, low entropy allows energy to dominate, recovering standard GNN optimization. Under heterophily, expanded entropy reorders class rankings, enabling predictions to break free from over-smoothed consensus.We operationalize this framework via a model-agnostic zentropy loss alongside a physics-informed neural network that predicts node-specific critical temperatures. A detailed summary of our contributions and distinctions from prior work is provided in Appendix A.1. As a drop-in replacement for cross-entropy, zentropy consistently enhances diverse GNN backbones, mitigating depth-induced over-smoothing and achieving up to a 9% accuracy gain on heterophilic benchmarks like Squirrel.

2 Method

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected graph with $N = |\mathcal{V}|$ nodes, feature matrix $X \in \mathbb{R}^{N\times F}$, adjacency matrix $A \in \mathbb{R}^{N\times N}$, and node neighborhood $\mathcal{N}(v) = \{u : (u, v) \in \mathcal{E}\}$. Each node $v \in \mathcal{V}$ carries a label $y_v \in \{1, \ldots, C\}$. For a GNN-derived logit vector $z \in \mathbb{R}^C$, the standard softmax cross-entropy loss is $\ell(z, y_v) = -z_{y_v} + \log \sum_{c=1}^{C} e^{z_c}$, where, under the energy mapping $E_c = -z_c$, the log-partition term $\log \sum_{c=1}^{C} e^{z_c}$ corresponds precisely to the negative Helmholtz free energy of the label system but at unit temperature (Liu et al., 2020).

2.1 The Thermodynamic Mechanism of Zentropy

Definition of zentropy. Suppose the entire system is heterogeneous, consisting of K configurations. $p^{(k)}$ is the probability of the k-th configuration and $E^{(k)}$ is the internal energy within each configuration. Zentropy theory (Liu et al., 2022) constructs a coarse-grained representation of entropy

by $S = \sum_{k=1}^{K} p^{(k)} S^{(k)} - k_B \sum_{k=1}^{K} p^{(k)} \ln p^{(k)}$, where $S^{(k)} (S^{(k)} \geq 0)$ is the entropy of the k-th configuration and k_B is the Boltzmann constant. For a system at temperature T, the *Helmholtz free energy* (Callen & Scott, 1998) $F = E - TS$ measures usable energy after accounting for entropic dissipation (TS). Intuitively, F captures the fundamental competition between minimizing energy $E = \sum_{k=1}^{K} E^{(k)}$ and maximizing disorder S, with equilibrium reached at its minimum.

Following the maximum entropy principle, $p^{(k)}$ is formulated as the solution to the following constrained optimization problem:

$$\max_{p^{(k)}} S, \qquad s.t. \qquad \sum_{k}^{K} p^{(k)} = 1, \qquad \sum_{k}^{K} p^{(k)} E^{(k)} = C_1, \qquad \sum_{k}^{K} p^{(k)} \left(S^{(k)}\right)^2 = C_2, \tag{1}$$

where the first constraint guarantees probability normalization, while the second and third constrain the expected internal energy and entropic fluctuations under C_1 and C_2, respectively. By using Lagrange multipliers and introducing temperature T and a tunable hyperparameter γ into the optimization [1], the solution of $p^{(k)}$ becomes (Details shown in Appendix A.2):

$$p^{(k)} = \frac{1}{Z} \exp\left(-\frac{E^{(k)} - TS^{(k)}}{k_B T} - \left(\frac{S^{(k)}}{\gamma k_B} \right)^2 \right), \quad Z = \sum_{k=1}^{K} \exp\left(-\frac{E^{(k)} - TS^{[k]}}{k_B T} - \left(\frac{S^{(k)}}{\gamma k_B} \right)^2 \right). \tag{2}$$

Discussion. Zentropy reduces to standard Shannon entropy when $S^{(k)} = 0$ for all k. Under this limit, the system is assumed to be entirely homogeneous. To that end, maximizing entropy subject to a fixed expected energy constraint yields $p^{(k)} \propto \exp\left(-\frac{E^{(k)}}{k_B T}\right)$, which corresponds precisely to the *softmax* function with $-E^{(k)}$ acting as the logit. The negative log-likelihood of the ground-truth class is essentially the cross-entropy loss.

2.2 TURBOCHARGE GNN WITH ZENTROPY

The cross-entropy loss is $\ell_{CE} = \frac{1}{|\mathcal{V}|} \sum_v \langle y_v, -\ln p_v \rangle$. [2] By replacing p_v in ℓ_{CE} using Eq. 2, we transit from cross-entropy to temperature-dependent zentropy as:

$$\ell_Z(T) = \frac{1}{|\mathcal{V}|} \sum_v \left\langle y_v, \frac{\tilde{E} - T_v \tilde{S}}{k_B T_v} + \left(\frac{\tilde{S}}{\gamma k_B} \right)^2 + \ln Z \right\rangle, \tag{3}$$

where $\tilde{E} = [E^{(1)}, E^{(2)}, ..., E^{(C)}]$ and $\tilde{S} = [S^{(1)}, S^{(2)}, ..., S^{(C)}]$ denote the internal energy and entropy of each class c. $T = [T_v]_{v \in \mathcal{V}}$ is the collection of temperatures throughout graph nodes.

Parameterization of energy and entropy. Compared to cross-entropy, zentropy ℓ_{CE} requires the estimation of internal energy $\tilde{E}$ and intrinsic entropy $\tilde{S}$. Suppose the final message-passing layer in GNN produces the node representation $h_v \in \mathbb{R}^d$. In the traditional cross-entropy baseline, the linear readout inherently acts as an energy function. Following this notion, we parameterize the prediction of internal energy for node v as $\tilde{E}_v = W_E h_v + b_E$ (a.k.a. energy head), where $W_E \in \mathbb{R}^{C \times d}$ and $b_E \in \mathbb{R}^C$ denote the learnable weight matrix and bias vector, respectively.

In addition to node embedding h_v, the entropy head receives an augmented descriptor $\phi_v = \left[h_v, \log(1 + \deg(v)), \delta_v\right]$, where $\delta_v = \|h_v - \deg(v)^{-1} \sum_{u \in \mathcal{N}(v)} h_u\|_2$ acts as a local Dirichlet term capturing neighborhood disagreement. We introduce these explicit structural features because standard message passing acts as a low-pass filter, inherently attenuating the local variations that $S^{(k)}$ aims to quantify. Consequently, the node representation h_v cannot reliably retain this information on its own, particularly since operations like the symmetric normalization in GNNs actively wash out node degrees. Importantly, both structural augmentations depend strictly on the input features X and adjacency matrix A; requiring no label information, they are fully accessible during inference. Finally, the entropy head is parameterized as $\tilde{S} = W_S \phi_v + b_S$ with learnable parameters $W_S \in \mathbb{R}^{C \times (d+2)}$ and $b_S \in \mathbb{R}^C$.

[1] We set one Lagrange multiplier to $\frac{1}{T}$ and another one to $\frac{1}{\gamma^2 k_B}$.

[2] Cross-entropy loss is computed only over the training nodes rather than the entire graph. We use $\mathcal{V}$ to denote this training subset for notational simplicity.

Physical Intuition of Zentropy in label assignment. The log-odds of the probability $p^{(c)}$ in Eq. 2 can be reformulated as: $\xi^{(c)} = \log p^{(c)} = -\frac{E_c}{T} + \left[S_c - \frac{S_c^2}{k_B\gamma^2}\right] - \log Z$, where the first two terms represent the log-likelihood and log-prior, and $\log Z$ is the log-partition function. Multiplying by T, the zentropy-driven label assignment corresponds directly to minimizing the Helmholtz free energy $\mathbb{F}^{(c)} = E^{(c)} - T\mathbb{S}^{(c)} = E^{(c)} - T\left(S^{(c)} - \frac{(S^{(c)})^2}{k_B\gamma^2}\right)$. The quadratic term increases monotonically from $S^{(c)} = 0$, peaks at $S^\star = \gamma^2 k_{\mathrm{B}}/2$ regardless of labels, and decreases thereafter. Because the intrinsic entropy S is independent of the temperature T, γ acts as a shape parameter determining the exact threshold at which entropy caseses to be rewarded during label assignment. Specifically, within the interval $[0, S^\star]$, a larger intrinsic entropy $S^{(c)}$ lowers the effective free energy $F^{(c)}$ (thereby boosting the class logit). Beyond $S^\star$, this relationship strictly inverts, actively penalizing excessively high entropy. In contrast, the temperature parameter T induces differential scaling between the log-likelihood and log-prior terms, breaking the invariance of class rankings. As shown in Fig. 1(b), classes c and $d\,(d \neq c)$ swap rank at a critical temperature $T^\star = \frac{\Delta E_{cd}}{\Delta \mathbb{S}_{cd}} = \frac{E^{(c)} - E^{(d)}}{\mathbb{S}^{(c)} - \mathbb{S}^{(d)}}$, which geometrically corresponds to the slope of the chord connecting the two states in the $(\mathbb{S}, E)$ plane.

Separability bounds for Zentropy readout. Under the assumption that the log-odds mapping $\xi^{(c)}$ is L_ξ-Lipschitz continuous, Proposition 1 shows that temperature-dependent zentropy alone cannot fully overcome over-smoothing, as pairwise log-odds differences remain bounded by the representation-level Dirichlet energy.

Proposition 1. *For node representations h and an L_ξ-Lipschitz log-odds mapping $\xi^{(c)}$, the average edge-wise variation satisfies $\frac{1}{|\mathcal{E}|}\sum_{(u,v)\in\mathcal{E}} \|\xi(h_u) - \xi(h_v)\|_2^2 \leq L_\xi^2 \cdot D(h)$, where $D(h) = \frac{1}{|\mathcal{E}|}\sum_{(u,v)\in\mathcal{E}} \|h_u - h_v\|_2^2$ is the Dirichlet energy. (Proof shown in Appendix A.3)*

2.3 Learning Temperature T for GNNs

Heterophily calls for a node-wise temperature. A graph is rarely homogeneous in its local topology. Some nodes reside in strongly homophilic neighborhoods where the energy term alone might be sufficient to identify the correct class. On the contrary, others lie in heterophilic or sparsely connected regions where multiple label assignments remain less consistent with the observed topology, requiring the entropy term to carry more weight in the decision boundary. However, a single global temperature imposes a rigid compromise across these distinct regimes. We therefore assign each node v an individualized temperature T_v, ensuring that the minimized free energy adaptively reflects the local trade-off between likelihood and multiplicity.

Coarse-graining the temperature axis. Predicting T_v as an unconstrained positive scalar induces a degenerate optimization trajectory. As soon as the energy head correctly ranks the ground-truth label, the model can trivially minimize the loss by shrinking T, which amplifies the logit margins. Consequently, gradient descent aggressively drives $T \to 0$, marginalizing the entropic contribution and collapsing back to the standard cross-entropy schema. To prevent this trivial minimum, we coarse-grain the temperature axis into M fixed anchors $\{T_1, \ldots, T_M\}$. Since these anchors remain static during training, they block the gradient-driven collapse to zero.

Learning temperature T. We define a strictly positive temperature range $[T_{\min}, T_{\max}]$ (where $T_{\min} > 0$) and discretize it into M uniformly spaced anchors $\{T_m\}_{m=1}^M$. The node-specific temperature T_v is then computed as a convex combination of these anchors, $T_v = \sum_{m=1}^M q_{vm} T_m$. The assignment probabilities $q_v = [q_{v1}, \ldots, q_{vM}]^\top$ are generated by a two-layer MLP applied to the augmented descriptor ϕ_v:

$$q_v = \mathrm{softmax}\big(W_{T,2}\mathrm{ReLU}(W_{T,1}\phi_v + b_{T,1}) + b_{T,2}\big), \tag{4}$$

where $W_{T,1} \in \mathbb{R}^{H\times(d+2)}$ and $W_{T,2} \in \mathbb{R}^{M\times H}$ are the learnable weight matrices (with H denoting the hidden dimension), and $b_{T,1} \in \mathbb{R}^H$, $b_{T,2} \in \mathbb{R}^M$ are the corresponding bias vectors.

2.4 Why a Node-wise Temperature Can Re-rank Labels

Let $\beta_v = 1/T_v$. For a class pair $c \neq d$, we define $\Delta E_{cd}(v) = E_v^{(c)} - E_v^{(d)}$, $\Delta \mathbb{S}_{cd}(v) = \mathbb{S}_v^{(c)} - \mathbb{S}_v^{(d)}$, and pairwise conditional log-odds $g_{cd}(v) = \log p_v^{(c)} - \log p_v^{(d)} = [\mathbb{S}_v^{(c)} - \beta E_v^{(c)}] - [\mathbb{S}_v^{(d)} - \beta E_v^{(d)}] = \Delta \mathbb{S}_{cd}(v) - \beta_v\, \Delta E_{cd}(v)$.

Proposition 2. *If $\Delta\mathbb{S}_{cd}(v) = 0$ for all v and $c \neq d$, for any two temperatures $T_v, T'_v > 0$ with $g_{cd}(v)\big|_{T'} \neq 0$, $\frac{g_{cd}(v)\big|_T}{g_{cd}(v)\big|_{T'}} = \frac{\beta_v}{\beta'_v}$. Hence the ordering of c and d, is identical for every $T_v > 0$.*

Proof. If $\Delta\mathbb{S}_{cd}(v) = 0$, then $g_{cd}(v) = -\beta_v \Delta E_{cd}(v)$. Since physical temperatures mandate $\beta_v > 0$, the ratio $\beta_v/\beta'_v > 0$ remains strictly positive; thus, $\operatorname{sign}(g_{cd}(v)) = -\operatorname{sign}(\Delta E_{cd}(v))$ holds independently of β_v. □

Proposition 3. *Suppose two different graph nodes/class configurations (v, c, d) and (v', c', d'), $g_{cd}(v)\big|_{T'} \neq 0$, and $g_{c'd'}(v')\big|_{T'} \neq 0$. If $\frac{g_{cd}(v)\big|_T}{g_{cd}(v)\big|_{T'}} \neq \frac{g_{c'd'}(v')\big|_T}{g_{c'd'}(v')\big|_{T'}}$, then no model satisfying $\Delta\mathbb{S} \equiv 0$ can exactly represent these conditional distributions. Exact representation requires $\exists v, c \neq d$ such that $\Delta\mathbb{S}_{cd}(v) \neq 0$.*

Proof. If the true conditional distributions violate this proportionality, enforcing $\Delta\mathbb{S} \equiv 0$ forces both ratios to evaluate to T'_v/T_v, which contradicts the premise. Thus, $\Delta\mathbb{S}_{cd}(v)$ must be non-zero for at least one class pair within the input domain. □

Proposition 4. *Suppose the pairwise decision boundary for temperature T_v is $\mathcal{B}_{cd}\big|_T = \{v : \Delta\mathbb{S}_{cd}(v) = \beta_v \Delta E_{cd}(v)\}$. Thus, a non-zero entropy contrast permits temperature-adaptive decision boundaries through the temperature-specific value β_v. If $\Delta\mathbb{S}_{cd} \equiv 0$, then $\mathcal{B}_{cd}\big|_T = \{v : \Delta E_{cd}(v) = 0\}$, which is independent of temperature T_v.*

Proof. Setting the pairwise decision boundary $g_{cd}\big|_T(v) = 0$ yields $\Delta\mathbb{S}_{cd}(v) = \beta_v \Delta E_{cd}(v)$, making the boundary entropy-dependent via temperature T_v. However, if $\Delta\mathbb{S}_{cd} \equiv 0$, the strict positivity of β_v reduces this to $\Delta E_{cd}(v) = 0$, rendering the boundary independent of T_v. □

Assume $\Delta E_{cd}(x) \neq 0$ and define the critical inverse temperature $\beta^{\text{crit}}_{cd}(v) = \Delta\mathbb{S}_{cd}(v)/\Delta E_{cd}(v)$. By definition, $g_{cd}\big|_T(v) = \Delta\mathbb{S}_{cd}(v) - \beta_v \Delta E_{cd}(v) = -\Delta E_{cd}(x)[\beta_v - \beta^{\text{crit}}_{cd}(v)]$.

Corollary 1. *Two temperatures T_v and T'_v induce opposite pairwise class orderings if and only if β_v and β'_v strictly bound the critical threshold: $\left(\beta_v - \beta^{\text{crit}}_{cd}(v)\right)\left(\beta'_v - \beta^{\text{crit}}_{cd}(v)\right) < 0$.*

Proof. By the definition of $g_{cd}\big|_T(v)$, the product of the log-odds for temperatures T_v and T'_v is $< g_{cd}\big|_T(v), g_{cd}\big|_{T'}(v) >= [\Delta E_{cd}(v)]^2[\beta_v - \beta^{crit}_{cd}(v)][\beta'_v - \beta^{crit}_{cd}(v)]$. Since $[\Delta E_{cd}(v)]^2$ is strictly positive, $g_{cd}\big|_T(v)$ and $g_{cd}\big|_{T'}(v)$ take opposite signs if and only if the product of the inverse temperature differences is negative. □

Discussion. These findings establish that temperature and entropy heads are strictly coupled. Without entropy contrast ($\Delta\mathbb{S}_{cd}(v) = 0$), temperature acts solely as a positive log-odds scaler, leaving decision boundaries invariant (Propositions 2–4). Any conditional heterogeneity under heterophily thus necessitates $\Delta\mathbb{S}_{cd}(v) \neq 0$ (Proposition 3), making the entropy head a necessary prerequisite for temperature scaling. Furthermore, class pair (c, d) re-ranks at node v only when β_v crosses $\beta^{\text{crit}}_{cd}(v)$ (Corollary 1). Since β_v is confined to $[1/T_{\max}, 1/T_{\min}]$, this admissible range defines expressible label shifts, where the fraction of nodes crossing their critical temperature quantifies active re-ranking across the network.

2.5 Training GNNs with Zentropy

In addition to node-specific temperature T_v, we further partition the graph into M subgraphs and allow graph nodes within the same subgraphs share the temperature $\bar{T}_v$. The loss function of training GNNs with Zentropy (called Z-GNN) is:

$$\mathcal{L} = \alpha \mathcal{L}_{\text{Z}}(T) + (1 - \alpha)\mathcal{L}_Z(\bar{T}) + \eta \mathcal{L}_{\text{reg}}, \tag{5}$$

where α and η are hyperparameter. To make the training more stable, the regularization function $\mathcal{L}_{reg}$ includes: (1) ℓ_2-norm difference between T and $\bar{T}$, (2) temperature difference within a local graph neighborhood $\frac{1}{|\mathcal{E}|}\sum_{(u,v)\in\mathcal{E}}(\bar{T}_u - \bar{T}_v)^2$, (3) ℓ_2-norm magnitude of output E_v from energy head, and (4) ℓ_2-norm regularization on the temperature weighting vector q_v in Eq. 4.

3 EXPERIMENTS

3.1 EXPERIMENTAL SETUP

We evaluate seven GNN architectures including GCN (Kipf & Welling, 2017), ResGCN(Chen et al., 2020), GATv2 (Brody et al., 2022), GraphSAGE(Hamilton et al., 2017), APPNP(Gasteiger et al., 2018), SGC(Wu et al., 2019), and GCNII(Chen et al., 2020), implemented via PyTorch Geometric (Fey & Lenssen, 2019), across depths $L \in \{2, 4, 8, 16, 32, 64, 128\}$ (encoding layers or propagation steps). For small/medium node classification, homophilic graphs (Cora, Citeseer, Pubmed) use Planetoid splits (Yang et al., 2016) across 10 seeds, while heterophilic graphs (Cornell, Texas, Wisconsin, Chameleon, Squirrel) use the 10 official `geom-gcn` splits (Pei et al., 2020). All cross-entropy and zentropy runs share identical hyperparameter configurations, splits, and early-stopping protocols. For large-scale OGB benchmarks (Hu et al., 2020) over 10 seeds, `ogbn-arxiv` and `ogbn-mag` are trained full-batch, whereas `ogbn-products` uses neighbor sampling ($L \in \{2, 3\}$) or full propagation ($L \in \{2, 4, 8, 16\}$). Finally, MUTAG, PROTEINS, and NCI1 (TUDataset (Morris et al., 2020)) are evaluated across all depths, and IMDB-BINARY/MULTI at $L = 2 - 16$, under 10-fold cross-validation. Statistical significance is assessed via a two-sided paired sign-flip test (*: $p \leq 0.05$, **: $p \leq 0.01$), uncorrected for multiple comparisons across cells. Detailed hyperparameters, per-backbone weights, and full numerical results corresponding to all figures are provided in Appendices A.4 and B.

3.2 EXPERIMENTAL RESULTS

Homophilic benchmarks Evaluating both cross-entropy and zentropy objective function across three homophilic citation networks (Figure 2(a-c); full results in Table 3), our analysis of 147 paired configurations reveals a stark asymmetry: 48 configurations significantly favor Zentropy, whereas only 7 favor cross-entropy ($p \leq 0.05$). Crucially, major performance gains occur precisely where the cross-entropy baseline undergoes severe representational collapse toward random chance, a regime geometrically corresponding to the flat-gradient "dead-pool". For example, Citeseer with APPNP at $L \geq 8$ recovers dramatically from 20.3% to 67.7% ($+47.4\%$, $p = 0.003$ at $L = 8$), maintaining up to a $+50$-point advantage across deeper layers. Similar robust recoveries span multiple architectures, including Cora/APPNP at $L = 8$ ($+40.7\%$), Cora/SGC at $L = 128$ ($+34.7\%$), Citeseer/GATv2 at $L = 4$ ($+37.2\%$), and Pubmed/SGC at $L \geq 16$ (up to $+11.4\%$; all $p \leq 0.01$).

Note, our empirical observations strictly adhere to the bound in Proposition 1: representational collapse is not universally reversible. Once $D(h)$ dissipates entirely, models reach an absolute performance floor irrespective of the readout objective. This terminal regime is exemplified by deep GCN and GraphSAGE backbones on Cora, where Cora/GCN at $L = 128$ shows only a marginal shift ($23.7\% \rightarrow 24.0\%$). Minor underperformance in zentropy observed in well-behaved settings (e.g., -0.06% to -1.02% drop on high-accuracy Citeseer/SGC) do not contradict our premise. Instead, they reflect the slight optimization overhead of thermodynamic regularization in low-entropy regimes where the baseline loss landscape is already well-conditioned and no representational collapse requires mitigation.

Synthesizing these empirical findings with Proposition 1 delineates two distinct over-smoothing regimes. **(1) Soft collapse** ($D(h) > 0$): Dirichlet energy has decayed significantly without fully vanishing. Standard cross-entropy traps parameters in flat local minima, whereas zentropy's high-temperature regime maintains a well-conditioned loss landscape, enabling the optimizer to exploit residual representation variance before cooling. **(2) Hard collapse** ($D(h) \rightarrow 0$): Node representations collapse to a uniform point, rendering topological signals completely uninformative and precluding thermodynamic recovery. A mechanistic interpretation is provided in Section 4.

Heterophilic benchmarks. Empirical evaluation across five heterophilic graph benchmarks (Figure 2(d-h); Tables 9–10) demonstrates that 54 of 245 paired configurations significantly favor the zentropy objective and that *no* configuration significantly favors cross-entropy ($p \leq 0.05$). The Squirrel architecture paired with GCNII exemplifies this trend: Zentropy yields statistically significant classification accuracy improvements across all seven evaluated depths ($+6.4\%$ to $+9.0\%$ margin, all $p \leq 0.01$). It is worth noting that such advantage persists at shallow network depth ($L = 2$) in the absence of representation collapse, establishing that thermodynamic loss-landscape

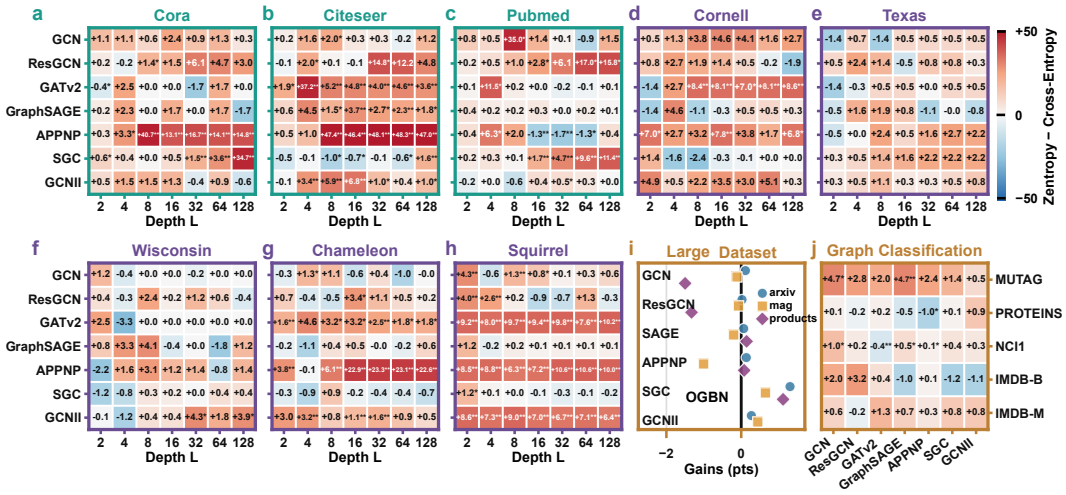

Figure 2: **Results on benchmarks.** Prediction accuracy (in %) improvements (Zentropy − Cross-entropy) across ten paired runs for each backbone and depth, on the homophilic graphs (in green box) **(a)** Cora, **(b)** Citeseer and **(c)** Pubmed, and on the heterophilic graphs (in purple box) **(d)** Cornell, **(e)** Texas, **(f)** Wisconsin, **(g)** Chameleon and **(h)** Squirrel. **(i)** OGB node classification, one marker per dataset and backbone, averaged over depths; **(j)** TUDataset graph classification, each cell averaged over $L \in \{2, 4, 8, 16\}$ over ten folds. Statistical significance is evaluated via a paired sign-flip test ($^*p \leq 0.05$, $^{**}p \leq 0.01$).

smoothing mitigates optimization pathologies induced by topological heterophily independently of depth-induced over-smoothing.

Aggregating heterophilic neighbors severely degrades feature discriminability, frequently trapping standard cross-entropy objectives in sub-optimal local minima even at shallow network depths. Zentropy robustly mitigates this vulnerability across architectures. Empirically, GATv2 on the Squirrel dataset yields depth-independent accuracy improvements of $+7.57\%$ to $+10.21\%$ ($L \in \{2, \ldots, 128\}$, $p \leq 0.01$), while APPNP achieves $+6.27\%$ to $+10.62\%$ gains. Similarly, GATv2 on Cornell exhibits $+7.03\%$ to $+8.65\%$ improvements for deeper networks ($L \geq 8$). As established in Corollary 1, these depth-independent gains emerge precisely when the entropy contrast ΔA_{cd} exhibits a non-uniform correlation with node degree. Under this regime, zentropy functions as a macroscopic mean-field controller. By dynamically scaling the effective temperature within dense, heterophilic subgraphs, it smooths localized structural noise and induces a dynamic equilibrium phase. This thermodynamic regulation enables the optimization trajectory to escape sharp, spurious minima before effectively annealing to resolve fine-grained, low-entropy decision boundaries.

Large graphs. Zentropy vs cross-entropy performance comparison on large graphs is shown in Fig. 2(i), where evaluations on `ogbn-arxiv`, `ogbn-mag`, and `ogbn-products` are shown in blue, orange, and purple, respectively. On `ogbn-arxiv`, sufficient Dirichlet energy $D(h)$ prevents representational collapse (maintaining 63.1%–72.4% prediction accuracy by cross-entropy), allowing zentropy to act as a soft thermodynamic regularizer that yields consistent gains across 20/24 configurations (0.32% improvement in average; SGC up to $+2.42\%$ at $L = 16$, Table 4). On heterogeneous `ogbn-mag`, zentropy outperforms cross-entropy in 9 of 24 configurations ($+0.14\%$ average). On `ogbn-products`, SGC, APPNP, and GraphSAGE gain up to $+1.45\%$, whereas GCN and ResGCN exhibit minor performance decline (min-max: 0.25%–2.53%). This minor divergence arises from train-test degree shifts under standard GCN normalization, which slightly misaligns context-mapped thermal parameters between training and inference regimes.

Graph classification. For graph-level classification (Shown in Fig. 2(j) and Table 8), zentropy following sum pooling yields marginal gains across 84 configurations (average $+0.27\%$ improvement), which remains consistent when employing attention readout, pre-pooling temperature scaling, and evaluating on IMDB benchmarks. We attribute this limitation to *thermodynamic homogenization*: standard global pooling aggregates node embeddings prior to classification, effectively erasing the local topological signals required for node-specific temperature modulation. Lacking the localized structural variance necessary to establish distinct high- and low-entropy optimization basins, uniform graph-level thermal adjustments are insufficient to alleviate representation collapse.

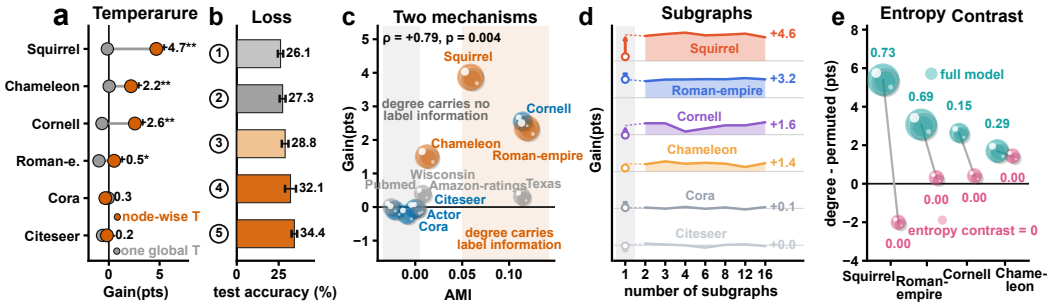

Figure 3: **Understand the performance gain by zentropy through ablation studies.** Performance is valued under paired accuracy differences over 10 seeds per dataset–backbone pair. Significance is assessed via an exact sign-flip test ($^{*}p \leq 0.05$, $^{**}p \leq 0.01$). **(a)** Gain of degree-based partitioning over an unpartitioned control, comparing node-wise (orange) versus global (gray) temperatures. **(b)** Step-wise loss ablation for GCNII on Squirrel: (1) standard cross-entropy, (2) zentropy readout with node-wise temperatures, (3) with temperature-matching regularizer, (4) with subgraph loss, and (5) the full objective. Bars and whiskers indicate means and standard deviations. **(c)** Adjusted Mutual Information (AMI) between training node degrees and labels is a putative indicator for possible performance gain by zentropy. Bubbles represent datasets averaged across 7 backbones; size and color indicate the number of backbones showing individually significant gains. **(d)** Performance gain across varying subgraph counts (M). Rings denote the unpartitioned baseline ($M{=}1$). Arrow lengths isolate the contribution of structural partitioning ($M \geq 2$); flat sparklines demonstrate that gains are insensitive to the exact choice of $M > 1$. Averaged performance gain is shown on the right. **(e)** Entropy contrast ablation, achieved by forcing a shared entropy head across classes ($\Delta S_{cd} \equiv 0$). Lines connect the full (green) and ablated (pink) models per dataset. Bubble sizes and printed values indicate the fraction of nodes whose predicted class changes across temperature extremes, a mechanism the ablation mathematically drives to zero.

4 Mechanistic Analysis of the Performance Gains

Since accuracy alone does not reveal why zentropy works, we next explore the mechanism behind the gains reported in Section 3.2. We design two nested ablations to isolate the impact of the spatial configuration of the node-wise temperature T_v. The first investigates whether temperature partitioning to sub-graphs provides any empirical benefit. The second evaluates whether the specific *degree-based* partition strategy matters, comparing our degree-based sub-graph against a random partition with identical M and matched sub-graph sizes. This sequential evaluation is strictly logical: the second ablation is meaningful only if the first yields positive results, as the composition of a partition is irrelevant if the act of partitioning itself offers no advantage. Figure 3(a) presents the results of the first ablation study, while panels (c)–(e) report the findings of the second.

Zentropy improves feature representations alongside decision boundaries. To isolate the source of the performance gains, we freeze encoders trained with either cross-entropy or zentropy, discard their original readouts, and train new linear, MLP, or zentropy heads on the frozen embeddings. On citation networks prone to severe representation collapse, the zentropy-trained encoders recover performance across all readouts, achieving 61.7%–71.7% accuracy even with a simple linear probe. In contrast, cross-entropy encoders plateau at 22.4%–35.2% regardless of the readout's capacity. Because this performance gap persists even when the readout is restricted to a linear probe, the improvement cannot be attributed merely to a more expressive classification head. Instead, this demonstrates that zentropy fundamentally reshapes the encoder during training, mitigating representation collapse at the source.

Graph partitioning and node-wise temperatures act in synergy (Figure 3a). Compared to the unpartitioned baseline, degree-based partitioning yielded significant gains only when paired with node-wise temperatures T_v (e.g., $+4.7^{**}$ on Squirrel and $+2.6^{**}$ on Cornell). Replacing T_v with a single shared temperature eliminated these gains. In this regard, the two components are complementary rather than redundant. The degree partition supplies the meaningful spatial configuration, while the node-wise temperature executes the boundary shifts prescribed by Propositions 2–4.

Each component in the loss function contributes incrementally (Figure 3b). Decomposing the zentropy objective on Squirrel with GCNII revealed a monotonic improvement over the cross-entropy baseline (26.1%). The zentropy readout with node-wise temperatures alone gave a modest gain (27.3%), and adding the temperature-matching regularizer raised accuracy to 28.8% ($p = 0.006$). The subgraph loss $L_Z(\bar{T})$ contributed the largest single increment (32.1%, $p = 0.012$), and the full objective peaked at 34.4% ($+7.02$ overall, $p = 0.002$). Taken together, the subgraph loss drives most of the gain on collapsed homophilic networks, whereas the full composite objective is required to reach peak performance under severe heterophily.

A training-free indicator for zentropy performance (Figure 3(c)). As established in Propositions 2–4 and Corollary 1, subgraph-conditional temperatures effectively shift decision boundaries only when the topology-induced temperature is informative of the ground-truth label. We can quantify this *a priori* using the Adjusted Mutual Information (AMI) (Vinh et al., 2009) between node degrees and class labels within the training set. Across 11 datasets and 7 backbones, empirical performance gains correlate strongly with this AMI (Spearman $\rho = +0.79$, $p = 0.004$ across datasets; $\rho = +0.46$, $p = 2 \times 10^{-5}$ across all 77 dataset-backbone pairs). For the six datasets with negligible AMI (< 0.01), the performance gain derived strictly from degree partitioning remains flat (-0.22% to $+0.39\%$). In contrast, the full zentropy formulation still yields massive improvements on these datasets (e.g., $+40.7\%$ for APPNP on Cora and $+47.4\%$ on Citeseer at $L = 8$). This demonstrates that performance recovery on citation networks stems from the encoder representation effect rather than the degree mechanism, proving the two effects are separable. However, high AMI is a necessary but not sufficient condition: Texas closely matches Cornell in AMI (0.114 vs. 0.115) yet exhibits no meaningful performance gain ($+0.31\%$ vs. $+2.55\%$, failing to reach statistical significance on any of the seven backbones).

Performance gains are robust to the number of subgraphs (Figure 3d). When we varied the subgraph count M, gains emerged once $M \geq 2$ and plateaued quickly over an eightfold range (e.g., Squirrel remained stable between $+4.1\%$ and $+4.9\%$). This pattern indicates that the advantage stems from the presence of topological conditioning rather than the granularity of degree quantization. Consequently, M is an insensitive hyperparameter in practice. *Roman-empire* dataset is the exception, showing substantial gains already at $M = 1$, which suggests that it benefits mainly from global temperature scaling rather than structural partitioning.

Entropy contrast is required for temperature-induced boundary shifts (Figure 3e). To test whether entropy contrast is necessary, we investigate the role of $\Delta\mathbb{S}_{cd}$ by sharing a single entropy head across classes. This ablation reduced the fraction of nodes that switch class between the two temperature extremes to zero. As a result, the structural gains collapsed (Squirrel dropped from $+5.4\%$ to -2.0%, and Cornell from $+2.6\%$ to $+0.4\%$). These results indicate that a non-zero entropy contrast is a necessary condition for temperature-induced boundary shifts, consistent with our theoretical analysis (Corollary 1).

5 Conclusion

We introduce a novel thermodynamic optimization framework that fundamentally reimagines the training of GNNs by transitioning from the rigid, zero-temperature assumptions of standard cross-entropy to a dynamic, model-agnostic zentropy loss using the Helmholtz free-energy formulation. Our empirical evaluations across diverse backbone architectures demonstrate that this physics-inspired approach successfully rescues deep GNNs from soft representational collapse, yielding substantial performance improvements on challenging heterophilic benchmarks such as Squirrel and Cornell. While bounded by the theoretical limits of complete Dirichlet energy decay, zentropy serves as a highly flexible, plug-and-play regularization mechanism. Ultimately, bridging statistical thermodynamics with graph representation learning establishes a principled and highly extensible pathway for enhancing the expressivity, depth, and robustness of neural architectures operating over complex network topologies.

Limitations. Since models are evaluated at matched depths, the recovery on homophilic graphs may partially stem from implicit depth regularization rather than zentropy alone. Furthermore, these benefits do not transfer to graph-level tasks, as global pooling washes out the local structural signals necessary for node-wise temperature modulation. Finally, our empirical scope is presently restricted to standard message-passing encoders.

AI USE STATEMENT

An LLM-based coding assistant implemented the coding, summarize the results and polished the manuscript. We have not used generative AI tools for figure generation, dataset construction, or ideation. We have reviewed all AI-assisted work. LLM-generated code was verified and tested for correctness by the first author. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This paper uses only public benchmark graphs and reports aggregate predictive performance; no human-subject data are involved.

REPRODUCIBILITY STATEMENT

Appendix A.4 gives every setting. Code will be available on GitHub after acceptance.

REFERENCES

Deyu Bo, Xiao Wang, Chuan Shi, and Huawei Shen. Beyond low-frequency information in graph convolutional networks. In *AAAI Conference on Artificial Intelligence*, 2021.

Shaked Brody, Uri Alon, and Eran Yahav. How attentive are graph attention networks? In *International Conference on Learning Representations (ICLR)*, 2022.

Chen Cai and Yusu Wang. A note on over-smoothing for graph neural networks. *arXiv preprint arXiv:2006.13318*, 2020.

Herbert B Callen and HL Scott. Thermodynamics and an introduction to thermostatistics, 1998.

Benjamin Paul Chamberlain, James Rowbottom, Maria Gorinova, Stefan Webb, Emanuele Rossi, and Michael M. Bronstein. GRAND: Graph neural diffusion. In *International Conference on Machine Learning (ICML)*, 2021.

Ming Chen, Zhewei Wei, Zengfeng Huang, Bolin Ding, and Yaliang Li. Simple and deep graph convolutional networks. In *International Conference on Machine Learning (ICML)*, 2020.

Eli Chien, Jianhao Peng, Pan Li, and Olgica Milenkovic. Adaptive universal generalized pagerank graph neural network. In *International Conference on Learning Representations (ICLR)*, 2021.

Matthias Fey and Jan Eric Lenssen. Fast graph representation learning with pytorch geometric. *arXiv preprint arXiv:1903.02428*, 2019.

Johannes Gasteiger, Aleksandar Bojchevski, and Stephan Günnemann. Predict then propagate: Graph neural networks meet personalized pagerank. *arXiv preprint arXiv:1810.05997*, 2018.

Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International conference on machine learning*, pp. 1263–1272. Pmlr, 2017.

Will Grathwohl, Kuan-Chieh Wang, Jörn-Henrik Jacobsen, David Duvenaud, Mohammad Norouzi, and Kevin Swersky. Your classifier is secretly an energy based model and you should treat it like one. *arXiv preprint arXiv:1912.03263*, 2019.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. On calibration of modern neural networks. In *International conference on machine learning*, pp. 1321–1330. PMLR, 2017.

Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

Hans Hao-Hsun Hsu, Yuesong Shen, Christian Tomani, and Daniel Cremers. What makes graph neural networks miscalibrated? In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

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

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*, 2017.

Yann LeCun, Sumit Chopra, Raia Hadsell, M Ranzato, Fujie Huang, et al. A tutorial on energy-based learning. *Predicting structured data*, 1(0), 2006.

Qimai Li, Zhichao Han, and Xiao-Ming Wu. Deeper insights into graph convolutional networks for semi-supervised learning. In *AAAI Conference on Artificial Intelligence*, 2018.

Derek Lim, Felix Hohne, Xiuyu Li, Sijia Linda Huang, Vaishnavi Gupta, Omkar Bhalerao, and Ser-Nam Lim. Large scale learning on non-homophilous graphs: New benchmarks and strong simple methods. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Weitang Liu, Xiaoyun Wang, John Owens, and Yixuan Li. Energy-based out-of-distribution detection. *Advances in neural information processing systems*, 33:21464–21475, 2020.

Zi-Kui Liu, Yi Wang, and Shun-Li Shang. Zentropy theory for positive and negative thermal expansion. *Journal of Phase Equilibria and Diffusion*, 43(6):598–605, 2022.

Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.

Christopher Morris, Nils M Kriege, Franka Bause, Kristian Kersting, Petra Mutzel, and Marion Neumann. Tudataset: A collection of benchmark datasets for learning with graphs. *arXiv preprint arXiv:2007.08663*, 2020.

Kenta Oono and Taiji Suzuki. Graph neural networks exponentially lose expressive power for node classification. In *International Conference on Learning Representations (ICLR)*, 2020.

Hongbin Pei, Bingzhe Wei, Kevin Chen-Chuan Chang, Yu Lei, and Bo Yang. Geom-gcn: Geometric graph convolutional networks. *arXiv preprint arXiv:2002.05287*, 2020.

Oleg Platonov, Denis Kuznedelev, Michael Diskin, Artem Babenko, and Liudmila Prokhorenkova. A critical look at the evaluation of gnns under heterophily: Are we really making progress? *arXiv preprint arXiv:2302.11640*, 2023.

Yu Rong, Wenbing Huang, Tingyang Xu, and Junzhou Huang. DropEdge: Towards deep graph convolutional networks on node classification. In *International Conference on Learning Representations (ICLR)*, 2020.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations (ICLR)*, 2018.

Nguyen Xuan Vinh, Julien Epps, and James Bailey. Information theoretic measures for clusterings comparison: is a correction for chance necessary? In *Proceedings of the 26th annual international conference on machine learning*, pp. 1073–1080, 2009.

Shun Wang, Shun-Li Shang, Zi-Kui Liu, and Wenrui Hao. ZENN: A thermodynamics-inspired computational framework for heterogeneous data–driven modeling. *Proceedings of the National Academy of Sciences*, 123(1):e2511227122, 2026.

Felix Wu, Amauri Souza, Tianyi Zhang, Christopher Fifty, Tao Yu, and Kilian Weinberger. Simplifying graph convolutional networks. In *International Conference on Machine Learning (ICML)*, 2019.

Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S Yu. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1):4–24, 2020.

Zhilin Yang, William Cohen, and Ruslan Salakhudinov. Revisiting semi-supervised learning with graph embeddings. In *International conference on machine learning*, pp. 40–48. PMLR, 2016.

Lingxiao Zhao and Leman Akoglu. PairNorm: Tackling oversmoothing in GNNs. In *International Conference on Learning Representations (ICLR)*, 2020.

Jiong Zhu, Yujun Yan, Lingxiao Zhao, Mark Heimann, Leman Akoglu, and Danai Koutra. Beyond homophily in graph neural networks: Current limitations and effective designs. *Advances in neural information processing systems*, 33:7793–7804, 2020.

A Appendix

A.1 Related Work

Learning under heterophily. Most message-passing architectures (Gilmer et al., 2017) aggregate neighbor features with a low-pass filter, which implicitly assumes homophily (i.e., adjacent nodes tend to share labels (McPherson et al., 2001)). Under strong heterophily, this aggregation mixes features from different classes and can render them inseparable. A large body of work relaxes the homophily assumption at the architectural level. Zhu et al. (2020) separate ego and neighbor embeddings and aggregate over higher-order neighborhoods, and Pei et al. (2020) route messages through a latent geometric space. Chien et al. (2021) learn signed generalized PageRank weights that permit subtractive propagation, while Bo et al. (2021) retain the high-frequency signals that low-pass filters discard. Lim et al. (2021) further show that embedding node features and adjacency separately yields strong baselines on large non-homophilous graphs. Meanwhile, Platonov et al. (2023) caution that part of the reported progress depends on the choice of benchmarks and data splits. Despite their diversity, these methods share one design principle: they change *how messages are aggregated*. In contrast, we leave the encoder untouched and modify only the readout. Since Zentropy acts solely at the output layer, it is orthogonal to these architectures and can, in principle, be combined with any of them.

Over-smoothing and depth. Repeated aggregation drives node representations toward a low-dimensional invariant subspace (Li et al., 2018; Oono & Suzuki, 2020), and Cai & Wang (2020) quantify this collapse through the exponential decay of the Dirichlet energy $D(h)$. Existing remedies act on the propagation operator, including normalization (Zhao & Akoglu, 2020), random edge dropping (Rong et al., 2020), initial residual and identity mappings (Chen et al., 2020), decoupled propagation (Gasteiger et al., 2018; Wu et al., 2019), and continuous diffusion (Chamberlain et al., 2021). Our readout-side approach is complementary to these remedies, but it also has a clear limit. Proposition 1 shows that the edge-wise variation of a readout is bounded by that of its input, scaled by the readout's Lipschitz constant. Thus, once $D(h)$ has vanished, no readout (thermodynamic or otherwise) can restore class separability. Between intact and fully collapsed representations lies an intermediate regime, in which features are heavily compressed but still carry residual signal. In this regime, the node-wise boundary shifts of Zentropy can still exploit that signal, as our experiments on deep citation networks show (Section 4). Taken together, our theory and experiments delineate where a readout can recover accuracy and where no readout can.

Energy-based views of classification. Interpreting logits as negative energies has a long history in deep learning (LeCun et al., 2006). For example, Liu et al. (2020) show that the log-partition function of a softmax classifier acts as a free energy that separates in-distribution from out-of-distribution inputs. Zentropy theory (Liu et al., 2022; Wang et al., 2026) extends this view by treating a system as a mixture of macroscopic configurations rather than a collection of structureless states. Each configuration carries its own internal degrees of freedom, which gives rise to a two-level entropy. The maximum-entropy solution of this formulation reduces to the standard softmax only when the configurational entropy vanishes. To our knowledge, this work is the first to bring the multi-level zentropy formulation to graph representation learning. On top of this, we identify the entropy contrast $\Delta\mathbb{S}_{cd}$ as the quantity that turns temperature from a mere rescaling factor into a driver of decision boundaries (Propositions 3–4).

Temperature and calibration. Temperature scaling is widely used for post-hoc calibration (Guo et al., 2017) and knowledge distillation (Hinton et al., 2015), and Hsu et al. (2022) extend it to GNNs with node-specific temperatures. In all these settings, the temperature rescales every logit of a node by the same factor. This softens or sharpens the output distribution but preserves the rank order of the logits, so it cannot change the predicted label (Proposition 3). Our temperature differs in two respects. First, it is conditioned on graph structure and active during training, rather than fitted post hoc. Second, it enters the free energy $F = E - TS$ asymmetrically. In the Boltzmann weight $\exp(-F/T) = \exp(-E/T + S)$, temperature rescales the energy but leaves the entropy term unchanged. Therefore, whenever classes differ in entropy ($\Delta\mathbb{S}_{cd} \neq 0$), changing T can reorder class probabilities and move a node across a decision boundary. In this regard, temperature becomes a structure-dependent control variable rather than a passive calibration scalar.

A.2 Derivation of probability from zentropy

Zentropy theory formulates a coarse-grained representation of entropy as:

$$S = \sum_{k=1}^{K} p^{(k)} S^{(k)} - k_{\mathrm{B}} \sum_{k=1}^{K} p^{(k)} \ln p^{(k)}, \tag{6}$$

where $S^{(k)} \geq 0$ denotes the intrinsic entropy of the k-th configuration. Originating from statistical mechanics, this term captures internal degrees of freedom, such as the vibrational frequencies of atoms (Liu et al., 2022). Here, k_{B} is the Boltzmann constant and $p^{(k)}$ is the probability of the k-th configuration. We infer the optimal distribution $p^{(k)}$ via the maximum entropy principle, subject to the following optimization problem:

$$\begin{aligned} \max_{\{p^{(k)}\}} \quad & \sum_{k=1}^{K} p^{(k)} S^{(k)} - k_{\mathrm{B}} \sum_{k=1}^{K} p^{(k)} \ln p^{(k)} \\ \text{s.t.} \quad & \sum_{k=1}^{K} p^{(k)} = 1 \\ & \sum_{k=1}^{K} p^{(k)} E^{(k)} = C_1 \\ & \sum_{k=1}^{K} p^{(k)} \big(S^{(k)}\big)^2 = C_2, \end{aligned} \tag{7}$$

where C_1 and C_2 are constants. While the first two constraints enforce standard probability normalization and total internal energy conservation, the third explicitly bounds the fluctuation of intrinsic entropy across individual configurations.

By introducing Lagrange multipliers λ_1, λ_2, and λ_3, we define the Lagrangian as:

$$\begin{aligned} \mathcal{L}\big(p^{(k)}\big) = & \sum_{k=1}^{K} p^{(k)} S^{(k)} - k_{\mathrm{B}} \sum_{k=1}^{K} p^{(k)} \ln p^{(k)} \\ & - \lambda_1 \left(\sum_{k=1}^{K} p^{(k)} - 1 \right) - \lambda_2 \left(\sum_{k=1}^{K} p^{(k)} E^{(k)} - C_1 \right) \\ & - \lambda_3 \left(\sum_{k=1}^{K} p^{(k)} \big(S^{(k)}\big)^2 - C_2 \right). \end{aligned} \tag{8}$$

Setting the partial derivative with respect to $p^{(k)}$ to zero yields the stationarity condition:

$$\frac{\partial \mathcal{L}}{\partial p^{(k)}} = S^{(k)} - k_{\mathrm{B}} \big(\ln p^{(k)} + 1\big) - \lambda_1 - \lambda_2 E^{(k)} - \lambda_3 \big(S^{(k)}\big)^2 = 0. \tag{9}$$

Solving for $p^{(k)}$, we obtain a factored, Gibbs-like distribution:

$$p^{(k)} = \exp\left(-\frac{\lambda_1 + k_{\mathrm{B}}}{k_{\mathrm{B}}} \right) \exp\left(\frac{S^{(k)} - \lambda_2 E^{(k)}}{k_{\mathrm{B}}} \right) \exp\left(-\frac{\lambda_3 \big(S^{(k)}\big)^2}{k_{\mathrm{B}}} \right). \tag{10}$$

The three exponential factors correspond respectively to a global normalization constant, the canonical thermodynamic weighting (balancing intrinsic entropy and internal energy), and a penalty on entropy fluctuation.

To ensure dimensional consistency within the exponential arguments, we parameterize the Lagrange multipliers as $\lambda_2 = \frac{1}{T}$ and $\lambda_3 = \frac{1}{\gamma^2 k_{\mathrm{B}}}$, where T denotes the temperature and γ is a tunable hyperparameter. By introducing the Helmholtz free energy for the k-th configuration, $F^{(k)} = E^{(k)} - TS^{(k)}$,

we can rewrite the optimal probability $p^{(k)}$ as a generalized Boltzmann distribution:

$$p^{(k)} = \frac{1}{Z} \exp\left(-\frac{F^{(k)}}{k_{\mathrm{B}} T} - \left(\frac{S^{(k)}}{\gamma k_{\mathrm{B}}} \right)^2 \right), \qquad Z = \sum_{k=1}^{K} \exp\left(-\frac{F^{(k)}}{k_{\mathrm{B}} T} - \left(\frac{S^{(k)}}{\gamma k_{\mathrm{B}}} \right)^2 \right). \tag{11}$$

A.3 Proof for Proposition 1

Let $\mathcal{H} \subset \mathbb{R}^d$ be a bounded region containing the representations h, and on it let B_S, B_E bound $\|S(h)\|$ and $\|E(h)\|$ and $L_{\bar{T}}$ be the Lipschitz constant of temperature mapping $h \mapsto \bar{T}(h)$.

Lemma 1 (Lipschitz constant of the zentropy readout)**.** *Under the stated assumptions, ξ is Lipschitz on $\mathcal{H}$ with constant L_ξ as given in*

$$L_\xi \le \|W_S\|_{\mathrm{op}}(1 + 2\gamma B_S) + \frac{\|W_E\|_{\mathrm{op}}}{T_{\min}} + \frac{B_E L_{\bar{T}}}{T_{\min}^2}, \tag{12}$$

where $\|\cdot\|_{\mathrm{op}}$ is the operator norm. Note, W_E and W_S are the project matrices used in the energy and entropy head.

Proof. Fix $h_1, h_2 \in \mathcal{H}$ and write $E_i = E(h_i) = W_E h_i$, $S_i = S(h_i) = W_S h_i$ and $T_i = \bar{T}(h_i) \ge T_{\min}$. Splitting the readout into its entropic and energetic parts,

$$\xi(h_1) - \xi(h_2) = \big[\mathbb{S}(h_1) - \mathbb{S}(h_2)\big] - \Big[\tfrac{E_1}{T_1} - \tfrac{E_2}{T_2}\Big].$$

For the first term in Eq. 12, $\mathbb{S}(h) = S(h) - \gamma S(h)^2$ acts componentwise, so

$$\mathbb{S}(h_1) - \mathbb{S}(h_2) = (S_1 - S_2) \odot \big[\mathbf{1} - \gamma(S_1 + S_2)\big],$$

and since $\|S_i\| \le B_S$,

$$\big\|\mathbb{S}(h_1) - \mathbb{S}(h_2)\big\| \le (1 + 2\gamma B_S)\,\|S_1 - S_2\| \le \|W_S\|_{\mathrm{op}}(1 + 2\gamma B_S)\,\|h_1 - h_2\|.$$

For the second and third terms in Eq. 12, adding and subtracting E_2/T_1 gives

$$\frac{E_1}{T_1} - \frac{E_2}{T_2} = \frac{E_1 - E_2}{T_1} + E_2\Big(\frac{1}{T_1} - \frac{1}{T_2}\Big),$$

so that, using $T_i \ge T_{\min} > 0$, $\|E_2\| \le B_E$ and $|T_1 - T_2| \le L_{\bar{T}}\|h_1 - h_2\|$,

$$\Big\|\frac{E_1}{T_1} - \frac{E_2}{T_2}\Big\| \le \frac{\|E_1 - E_2\|}{T_{\min}} + B_E\frac{|T_1 - T_2|}{T_{\min}^2} \le \Big(\frac{\|W_E\|_{\mathrm{op}}}{T_{\min}} + \frac{B_E L_{\bar{T}}}{T_{\min}^2}\Big)\|h_1 - h_2\|.$$

The triangle inequality combines the two bounds into Eq. 12. □

Proposition 1 *For node representations h and an L_z-Lipschitz log-odds mapping $\xi^{(c)}$, the average edge-wise variation satisfies $\frac{1}{|\mathcal{E}|}\sum_{(u,v)\in\mathcal{E}} \|\xi(h_u) - \xi(h_v)\|_2^2 \le L_\xi^2 \cdot D(h)$, where $D(h) = \frac{1}{|\mathcal{E}|}\sum_{(u,v)\in\mathcal{E}} \|h_u - h_v\|_2^2$ is the Dirichlet energy.*

Proof. Let $(u, v) \in \mathcal{E}$. Since $h_u, h_v \in \mathcal{H}$ and ξ is L_ξ-Lipschitz on $\mathcal{H}$,

$$\|\xi(h_u) - \xi(h_v)\|_2 \;\le\; L_\xi\,\|h_u - h_v\|_2.$$

Both sides are non-negative, so squaring preserves the inequality: $\|\xi(h_u) - \xi(h_v)\|_2^2 \le L_\xi^2\|h_u - h_v\|_2^2$. Summing over all edges and dividing by $|\mathcal{E}| > 0$,

$$\frac{1}{|\mathcal{E}|}\sum_{(u,v)\in\mathcal{E}} \|\xi(h_u) - \xi(h_v)\|_2^2 \;\le\; L_\xi^2 \cdot \frac{1}{|\mathcal{E}|}\sum_{(u,v)\in\mathcal{E}} \|h_u - h_v\|_2^2 \;=\; L_\xi^2 \cdot D(h).$$ □

Corollary 2 (Unrecoverability under complete collapse)**.** *In the limit of severe over-smoothing where Dirichlet energy $D(h) \to 0$, the mean edge-wise variation of ξ tends to zero for any readout exhibiting a finite Lipschitz constant L_ξ. Thus, if the graph encoder collapses neighboring representations into indistinguishable points, no readout mechanism, including zentropy, can recover their separation.*

A.4 EXPERIMENTAL PROTOCOL

All runs use Adam at learning rate 0.01 with gradient-norm clipping at 5.0. Citation runs allow 400 epochs (minimum 80, patience 60); heterophilic runs 600 (minimum 100, patience 80); evaluation every five epochs. The readout uses 6 temperature anchors confined to $[T_{\min}, T_{\max}] = [0.3, 2.5]$ and $M = 6$ degree-quantile subgraphs at every dataset and depth; the temperature network is a two-layer MLP whose hidden width H equals the encoder's, and attention backbones use a single head. Dataset hyperparameters are in Tables 1–2. ogbn-arxiv and ogbn-mag use depth-stabilized encoders (LayerNorm and residuals), hidden width 256 and 128 respectively. ogbn-products uses fanouts $[15, 10, 5]$, batch 1024, and exact layer-wise inference; GCN/ResGCN use the importance-reweighted normalization of Section 3.2.

B COMPLETE RESULT TABLES

All detailed numerical tables are collected here; the main text reports the figure-based view.

Table 1: Dataset-specific hyperparameters, shared across all seven backbones.

Dataset	Hidden	Dropout	Weight decay	$1/(\gamma^2 k_{\mathrm{B}}^2)$	μ
Cora	64	0.5	5×10^{-4}	0.00	1×10^{-3}
Citeseer	128	0.6	1×10^{-3}	0.01	1×10^{-2}
Pubmed	128	0.6	1×10^{-3}	0.01	1×10^{-2}

Table 2: Dataset-specific hyperparameters for the heterophilic benchmarks, shared across all seven backbones.

Dataset	Hidden	Dropout	Weight decay	$1/(\gamma^2 k_{\mathrm{B}}^2)$	μ
Cornell	64	0.5	5×10^{-4}	0.01	1×10^{-2}
Texas	64	0.5	5×10^{-4}	0.01	1×10^{-2}
Wisconsin	64	0.5	5×10^{-4}	0.01	1×10^{-2}
chameleon	128	0.6	1×10^{-3}	0.01	1×10^{-2}
squirrel	128	0.6	1×10^{-3}	0.01	1×10^{-2}

Table 3: Test accuracy (%) on citation networks with the zentropy readout of Section 2.5 (Z-GNN). Values are mean ± standard deviation over paired seeds. Arrows give the change relative to Cross-entropy. * denotes $p \leq 0.05$ and ** denotes $p \leq 0.01$ under a two-sided paired sign-flip test.

Dataset	Model	$L=2$	$L=4$	$L=8$	$L=16$	$L=32$	$L=64$	$L=128$
Cora	GCN	80.51 ± 1.03	73.83 ± 2.87	23.73 ± 5.53	24.95 ± 5.14	22.99 ± 5.89	22.56 ± 5.24	23.67 ± 5.36
	GCN+Z	$\mathbf{81.59 \pm 0.60}^{\sharp}_{1.08\uparrow}$	$\mathbf{74.90 \pm 2.22}_{1.07\uparrow}$	$\mathbf{24.31 \pm 9.17}_{0.58\uparrow}$	$\mathbf{27.31 \pm 4.03}^{\sharp}_{2.36\uparrow}$	$\mathbf{23.90 \pm 5.28}_{0.91\uparrow}$	$\mathbf{23.90 \pm 5.12}_{1.34\uparrow}$	$\mathbf{23.97 \pm 4.06}_{0.30\uparrow}$
	RESGCN	80.71 ± 1.09	79.49 ± 1.45	76.80 ± 1.90	71.21 ± 4.50	55.24 ± 22.25	56.83 ± 18.12	49.38 ± 17.79
	RESGCN+Z	$\mathbf{80.94 \pm 0.76}_{0.23\uparrow}$	$79.29 \pm 0.95^{\sharp}_{0.20\downarrow}$	$\mathbf{78.23 \pm 1.52}^{\uparrow}_{1.43\uparrow}$	$\mathbf{72.69 \pm 2.30}_{1.48\uparrow}$	$\mathbf{61.30 \pm 14.58}_{6.06\uparrow}$	$\mathbf{61.54 \pm 13.24}_{4.71\uparrow}$	$\mathbf{52.40 \pm 19.50}_{3.02\uparrow}$
	GATV2	80.38 ± 1.02	73.87 ± 3.36	31.90 ± 0.00	31.90 ± 0.00	31.90 ± 0.00	30.20 ± 5.38	31.90 ± 0.00
	GATV2+Z	$79.99 \pm 0.87^{\sharp}_{0.39\downarrow}$	$\mathbf{76.37 \pm 1.94}_{2.50\uparrow}$	$\mathbf{31.90 \pm 0.00}_{0.00\uparrow}$	$\mathbf{31.90 \pm 0.00}_{0.00\uparrow}$	$30.20 \pm 5.38_{1.70\downarrow}$	$\mathbf{31.90 \pm 0.00}_{1.70\uparrow}$	$\mathbf{31.90 \pm 0.00}_{0.00\uparrow}$
	SAGE	79.96 ± 0.97	75.90 ± 2.06	31.90 ± 0.00	30.20 ± 5.38	30.20 ± 5.38	30.20 ± 5.38	30.20 ± 5.38
	SAGE+Z	$\mathbf{80.18 \pm 0.86}_{0.22\uparrow}$	$\mathbf{78.19 \pm 1.86}_{2.29\uparrow}$	$\mathbf{31.90 \pm 0.00}_{0.00\uparrow}$	$\mathbf{31.90 \pm 0.00}_{1.70\uparrow}$	$\mathbf{30.20 \pm 5.38}_{0.00\uparrow}$	$\mathbf{31.90 \pm 0.00}^{\sharp}_{1.70\uparrow}$	$28.50 \pm 7.17_{1.70\downarrow}$
	APPNP	80.10 ± 1.54	77.34 ± 4.78	30.55 ± 5.21	61.59 ± 11.96	57.06 ± 4.26	58.91 ± 10.35	59.43 ± 10.56
	APPNP+Z	$\mathbf{80.44 \pm 0.71}^{\sharp}_{0.34\uparrow}$	$\mathbf{80.62 \pm 1.52}^{\sharp}_{8.28\uparrow}$	$\mathbf{71.25 \pm 10.69}^{**}_{40.70\uparrow}$	$\mathbf{74.66 \pm 2.51}^{**}_{13.07\uparrow}$	$\mathbf{73.81 \pm 1.91}^{**}_{16.75\uparrow}$	$\mathbf{73.01 \pm 5.01}^{**}_{14.10\uparrow}$	$\mathbf{74.24 \pm 3.59}^{**}_{14.81\uparrow}$
	SGC	81.71 ± 0.70	81.87 ± 1.10	81.23 ± 0.33	79.82 ± 0.48	77.69 ± 0.95	71.33 ± 4.57	33.81 ± 4.19
	SGC+Z	$\mathbf{82.33 \pm 0.55}^{*}_{0.62\uparrow}$	$\mathbf{82.28 \pm 0.66}_{0.41\uparrow}$	$\mathbf{81.27 \pm 0.33}_{0.04\uparrow}$	$\mathbf{80.29 \pm 0.42}_{0.47\uparrow}$	$\mathbf{79.23 \pm 0.61}^{**}_{1.54\uparrow}$	$\mathbf{74.92 \pm 1.02}^{**}_{3.59\uparrow}$	$\mathbf{68.48 \pm 10.26}^{**}_{34.67\uparrow}$
	GCNII	80.25 ± 1.05	79.08 ± 1.72	78.65 ± 1.90	79.95 ± 1.03	84.03 ± 0.55	84.31 ± 0.61	84.74 ± 0.72
	GCNII+Z	$\mathbf{80.73 \pm 0.69}^{\sharp}_{0.48\uparrow}$	$\mathbf{80.61 \pm 1.05}^{\sharp}_{1.53\uparrow}$	$\mathbf{80.15 \pm 0.70}_{1.50\uparrow}$	$\mathbf{81.28 \pm 1.53}_{1.33\uparrow}$	$83.68 \pm 0.89_{0.35\downarrow}$	$\mathbf{85.20 \pm 0.41}^{\sharp}_{0.89\uparrow}$	$84.16 \pm 0.67^{\uparrow}_{0.58\downarrow}$
Citeseer	GCN	66.25 ± 1.68	54.04 ± 9.88	20.38 ± 2.07	21.94 ± 1.71	21.80 ± 1.69	22.14 ± 2.40	21.07 ± 2.44
	GCN+Z	$\mathbf{66.44 \pm 1.62}_{0.19\uparrow}$	$\mathbf{55.60 \pm 9.85}_{1.56\uparrow}$	$\mathbf{22.42 \pm 2.27}^{*}_{2.04\uparrow}$	$\mathbf{22.21 \pm 1.73}_{0.27\uparrow}$	$\mathbf{22.07 \pm 2.42}_{0.27\uparrow}$	$21.95 \pm 1.74_{0.19\downarrow}$	$\mathbf{22.29 \pm 1.52}_{1.22\uparrow}$
	RESGCN	69.08 ± 1.51	65.34 ± 1.57	62.97 ± 1.39	48.71 ± 11.35	34.99 ± 12.08	37.07 ± 15.28	38.78 ± 10.11
	RESGCN+Z	$68.98 \pm 1.27_{0.10\downarrow}$	$\mathbf{67.33 \pm 1.32}^{*}_{1.99\uparrow}$	$\mathbf{63.06 \pm 2.05}_{0.09\uparrow}$	$48.65 \pm 14.46_{0.06\downarrow}$	$\mathbf{49.75 \pm 7.68}^{\uparrow}_{14.76\uparrow}$	$\mathbf{49.26 \pm 6.97}_{12.19\uparrow}$	$\mathbf{43.57 \pm 12.65}_{4.79\uparrow}$
	GATV2	66.46 ± 1.40	19.47 ± 4.33	18.10 ± 0.00	18.10 ± 0.00	18.10 ± 0.00	18.10 ± 0.00	18.10 ± 0.00
	GATV2+Z	$\mathbf{68.34 \pm 1.69}^{\uparrow}_{1.88\uparrow}$	$\mathbf{56.63 \pm 12.04}^{**}_{37.16\uparrow}$	$\mathbf{23.31 \pm 0.82}^{**}_{5.21\uparrow}$	$\mathbf{22.85 \pm 0.46}^{*}_{4.75\uparrow}$	$\mathbf{22.09 \pm 2.22}^{**}_{3.99\uparrow}$	$\mathbf{22.66 \pm 1.84}^{**}_{4.56\uparrow}$	$\mathbf{21.71 \pm 2.01}^{*}_{3.61\uparrow}$
	SAGE	68.24 ± 1.09	59.39 ± 14.61	18.21 ± 0.22	18.17 ± 0.12	18.17 ± 0.22	18.12 ± 0.06	18.60 ± 1.58
	SAGE+Z	$\mathbf{68.82 \pm 1.90}_{0.58\uparrow}$	$\mathbf{63.84 \pm 2.06}_{4.45\uparrow}$	$\mathbf{19.74 \pm 2.15}^{\uparrow}_{1.53\uparrow}$	$\mathbf{21.88 \pm 2.68}^{**}_{3.71\uparrow}$	$\mathbf{20.91 \pm 2.89}^{*}_{2.74\uparrow}$	$\mathbf{20.42 \pm 1.68}^{**}_{2.30\uparrow}$	$\mathbf{20.35 \pm 1.92}^{\uparrow}_{1.75\uparrow}$
	APPNP	68.28 ± 1.36	22.68 ± 4.76	20.32 ± 4.77	20.68 ± 2.71	18.98 ± 2.20	18.85 ± 3.54	19.69 ± 2.94
	APPNP+Z	$\mathbf{68.83 \pm 1.15}^{\sharp}_{0.55\uparrow}$	$\mathbf{23.70 \pm 1.81}_{1.02\uparrow}$	$\mathbf{67.72 \pm 1.15}^{**}_{47.40\uparrow}$	$\mathbf{67.05 \pm 1.29}^{**}_{46.37\uparrow}$	$\mathbf{67.10 \pm 1.34}^{**}_{48.12\uparrow}$	$\mathbf{67.19 \pm 1.94}^{**}_{48.34\uparrow}$	$\mathbf{66.69 \pm 1.49}^{**}_{47.00\uparrow}$
	SGC	71.98 ± 0.54	71.53 ± 0.56	72.16 ± 0.41	72.65 ± 0.45	72.07 ± 0.47	70.62 ± 0.52	67.93 ± 1.09
	SGC+Z	$71.46 \pm 0.57^{\sharp}_{0.52\downarrow}$	$71.47 \pm 0.35^{\sharp}_{0.06\downarrow}$	$71.14 \pm 0.68^{\uparrow}_{1.02\downarrow}$	$71.96 \pm 0.40^{*}_{0.69\downarrow}$	$71.99 \pm 0.41^{\uparrow}_{0.08\downarrow}$	$70.06 \pm 0.20^{\sharp}_{0.56\downarrow}$	$\mathbf{69.50 \pm 0.63}^{**}_{1.57\uparrow}$
	GCNII	66.36 ± 1.86	63.06 ± 2.82	61.53 ± 4.85	64.86 ± 3.08	72.69 ± 1.00	72.95 ± 0.80	73.10 ± 0.96
	GCNII+Z	$66.30 \pm 1.65_{0.06\downarrow}$	$\mathbf{66.43 \pm 1.38}^{**}_{5.37\uparrow}$	$\mathbf{67.40 \pm 1.86}^{**}_{5.87\uparrow}$	$\mathbf{71.64 \pm 0.81}^{**}_{6.78\uparrow}$	$\mathbf{73.70 \pm 0.51}^{**}_{1.01\uparrow}$	$\mathbf{73.31 \pm 0.61}_{0.36\uparrow}$	$\mathbf{74.12 \pm 0.63}^{*}_{1.02\uparrow}$
Pubmed	GCN	78.04 ± 1.10	75.60 ± 1.37	39.91 ± 1.79	39.58 ± 1.99	39.57 ± 2.47	40.36 ± 1.64	39.06 ± 1.47
	GCN+Z	$\mathbf{78.81 \pm 0.53}_{0.77\uparrow}$	$\mathbf{76.13 \pm 0.80}_{0.53\uparrow}$	$\mathbf{74.94 \pm 1.33}^{**}_{35.03\uparrow}$	$\mathbf{41.00 \pm 2.78}_{1.42\uparrow}$	$\mathbf{39.65 \pm 2.20}_{0.08\uparrow}$	$39.41 \pm 0.94_{0.95\downarrow}$	$\mathbf{40.59 \pm 2.28}_{1.53\uparrow}$
	RESGCN	78.32 ± 0.88	77.31 ± 0.86	75.36 ± 1.04	71.32 ± 3.87	58.58 ± 15.11	46.05 ± 9.95	54.98 ± 14.68
	RESGCN+Z	$\mathbf{78.55 \pm 0.76}_{0.23\uparrow}$	$\mathbf{77.76 \pm 0.91}_{0.45\uparrow}$	$\mathbf{76.39 \pm 1.50}_{1.03\uparrow}$	$\mathbf{74.07 \pm 3.66}^{*}_{2.75\uparrow}$	$\mathbf{64.69 \pm 14.25}_{6.11\uparrow}$	$\mathbf{63.02 \pm 15.86}^{\uparrow}_{16.97\uparrow}$	$\mathbf{70.75 \pm 10.29}^{\uparrow}_{15.77\uparrow}$
	GATV2	77.79 ± 0.69	65.50 ± 17.15	40.70 ± 0.00	40.70 ± 0.00	40.70 ± 0.00	40.70 ± 0.00	40.70 ± 0.00
	GATV2+Z	$\mathbf{77.87 \pm 0.56}_{0.08\uparrow}$	$\mathbf{77.04 \pm 1.41}^{\uparrow}_{11.54\uparrow}$	$\mathbf{40.88 \pm 0.29}_{0.18\uparrow}$	$\mathbf{40.70 \pm 0.00}_{0.00\uparrow}$	$40.50 \pm 0.42_{0.20\downarrow}$	$40.60 \pm 0.32_{0.10\downarrow}$	$\mathbf{40.76 \pm 0.19}_{0.06\uparrow}$
	SAGE	76.80 ± 0.84	76.45 ± 0.98	40.70 ± 0.00	40.70 ± 0.00	40.70 ± 0.00	40.70 ± 0.00	40.70 ± 0.00
	SAGE+Z	$\mathbf{77.17 \pm 0.44}_{0.37\uparrow}$	$\mathbf{76.66 \pm 1.62}^{\sharp}_{0.21\uparrow}$	$\mathbf{40.89 \pm 0.60}_{0.19\uparrow}$	$\mathbf{40.97 \pm 0.72}^{\sharp}_{0.27\uparrow}$	$\mathbf{40.70 \pm 0.00}_{0.00\uparrow}$	$\mathbf{40.89 \pm 0.60}_{0.19\uparrow}$	$\mathbf{40.76 \pm 0.19}_{0.06\uparrow}$
	APPNP	76.71 ± 0.65	66.17 ± 4.68	74.39 ± 6.34	77.04 ± 0.80	77.46 ± 1.15	77.54 ± 1.47	76.12 ± 2.12
	APPNP+Z	$\mathbf{77.13 \pm 0.95}_{0.42\uparrow}$	$\mathbf{72.47 \pm 3.77}^{\sharp}_{6.30\uparrow}$	$\mathbf{76.35 \pm 1.36}_{1.96\uparrow}$	$75.79 \pm 0.84^{**}_{1.25\downarrow}$	$75.71 \pm 1.03^{**}_{1.75\downarrow}$	$76.28 \pm 1.09^{\uparrow}_{1.26\downarrow}$	$\mathbf{76.55 \pm 1.17}_{0.43\uparrow}$
	SGC	78.52 ± 0.33	78.61 ± 0.32	79.23 ± 0.39	77.22 ± 0.42	71.87 ± 1.43	61.69 ± 4.96	47.60 ± 3.10
	SGC+Z	$\mathbf{78.76 \pm 0.32}_{0.24\uparrow}$	$\mathbf{78.96 \pm 0.63}^{\sharp}_{0.35\uparrow}$	$\mathbf{79.29 \pm 0.34}_{0.06\uparrow}$	$\mathbf{78.89 \pm 0.54}^{**}_{1.67\uparrow}$	$\mathbf{76.54 \pm 0.43}^{**}_{4.67\uparrow}$	$\mathbf{71.29 \pm 0.42}^{**}_{9.60\uparrow}$	$\mathbf{59.02 \pm 8.16}^{**}_{11.42\uparrow}$
	GCNII	78.19 ± 0.82	77.17 ± 0.95	78.43 ± 0.80	79.45 ± 0.76	79.54 ± 0.24	79.55 ± 0.51	79.72 ± 0.36
	GCNII+Z	$77.94 \pm 0.42_{0.25\downarrow}$	$\mathbf{77.19 \pm 0.86}_{0.02\uparrow}$	$77.78 \pm 1.22_{0.65\downarrow}$	$\mathbf{79.89 \pm 0.36}_{0.44\uparrow}$	$\mathbf{80.00 \pm 0.30}^{*}_{0.46\uparrow}$	$\mathbf{79.81 \pm 0.47}_{0.26\uparrow}$	$\mathbf{79.74 \pm 0.31}_{0.02\uparrow}$

Table 4: Test accuracy (%) on ogbn-arxiv (169,343 nodes, official split) with the zentropy readout (Z-GNN).

Model	$L=2$	$L=4$	$L=8$	$L=16$
GCN	71.38 ± 0.10	72.22 ± 0.20	72.23 ± 0.19	72.25 ± 0.24
GCN+Z	$\mathbf{71.49 \pm 0.15}^{\mathbf{8/10}}_{\mathbf{0.11\uparrow}}$	$\mathbf{72.29 \pm 0.20}^{\mathbf{7/10}}_{\mathbf{0.07\uparrow}}$	$\mathbf{72.48 \pm 0.22}^{\mathbf{9/10}\,**}_{\mathbf{0.25\uparrow}}$	$\mathbf{72.26 \pm 0.24}^{\mathbf{6/10}}_{\mathbf{0.01\uparrow}}$
RESGCN	71.32 ± 0.19	72.33 ± 0.28	72.39 ± 0.22	72.09 ± 0.19
RESGCN+Z	$\mathbf{71.51 \pm 0.17}^{\mathbf{9/10}\,*}_{\mathbf{0.19\uparrow}}$	$72.25 \pm 0.33^{4/10}_{0.08\downarrow}$	$72.34 \pm 0.13^{5/10}_{0.05\downarrow}$	$\mathbf{72.11 \pm 0.29}^{\mathbf{4/10}}_{\mathbf{0.02\uparrow}}$
SAGE	71.14 ± 0.23	71.78 ± 0.35	72.05 ± 0.23	71.96 ± 0.32
SAGE+Z	$\mathbf{71.36 \pm 0.18}^{\mathbf{8/10}\,*}_{\mathbf{0.22\uparrow}}$	$\mathbf{71.86 \pm 0.17}^{\mathbf{7/10}}_{\mathbf{0.08\uparrow}}$	$72.03 \pm 0.29^{6/10}_{0.03\downarrow}$	$71.95 \pm 0.22^{4/10}_{0.01\downarrow}$
APPNP	69.88 ± 0.22	70.48 ± 0.29	70.34 ± 0.15	69.97 ± 0.15
APPNP+Z	$\mathbf{70.15 \pm 0.26}^{\mathbf{8/10}\,*}_{\mathbf{0.27\uparrow}}$	$\mathbf{70.66 \pm 0.25}^{\mathbf{8/10}}_{\mathbf{0.18\uparrow}}$	$\mathbf{70.34 \pm 0.24}^{\mathbf{6/10}}_{\mathbf{0.00\uparrow}}$	$\mathbf{70.07 \pm 0.19}^{\mathbf{6/10}}_{\mathbf{0.10\uparrow}}$
SGC	69.01 ± 0.10	68.68 ± 0.12	66.76 ± 0.20	63.09 ± 0.22
SGC+Z	$\mathbf{69.55 \pm 0.12}^{\mathbf{10/10}\,**}_{\mathbf{0.54\uparrow}}$	$\mathbf{69.64 \pm 0.13}^{\mathbf{10/10}\,**}_{\mathbf{0.96\uparrow}}$	$\mathbf{68.07 \pm 0.35}^{\mathbf{10/10}\,**}_{\mathbf{1.30\uparrow}}$	$\mathbf{65.51 \pm 0.22}^{\mathbf{10/10}\,**}_{\mathbf{2.42\uparrow}}$
GCNII	71.39 ± 0.22	71.66 ± 0.28	71.32 ± 0.21	70.82 ± 0.26
GCNII+Z	$\mathbf{71.50 \pm 0.21}^{\mathbf{7/10}}_{\mathbf{0.11\uparrow}}$	$\mathbf{72.04 \pm 0.16}^{\mathbf{9/10}\,**}_{\mathbf{0.38\uparrow}}$	$\mathbf{71.53 \pm 0.18}^{\mathbf{9/10}\,**}_{\mathbf{0.21\uparrow}}$	$\mathbf{71.21 \pm 0.23}^{\mathbf{9/10}\,**}_{\mathbf{0.39\uparrow}}$
GATV2	$69.19 \pm 0.19^{\S}$	$70.06 \pm 0.34^{\S}$	$70.23 \pm 0.18^{\S}$	$69.99 \pm 0.44^{\S}$
GATV2+Z	$\mathbf{69.54 \pm 0.28}^{\mathbf{8/10}\,*\S}_{\mathbf{0.35\uparrow}}$	$\mathbf{70.25 \pm 0.27}^{\mathbf{6/10}\S}_{\mathbf{0.20\uparrow}}$	$\mathbf{70.40 \pm 0.23}^{\mathbf{6/10}\S}_{\mathbf{0.17\uparrow}}$	$\mathbf{70.50 \pm 0.26}^{\mathbf{9/10}\,**\S}_{\mathbf{0.51\uparrow}}$

Table 5: Test accuracy (%) on ogbn-products (2,449,029 nodes, 123,718,280 edges) with the temperature-distillation Z-GNN adapter. GCN, ResGCN, and SAGE use `SampledStack`, a stack of GCNConv/SAGEConv layers trained via neighbor-sampled mini-batches (fanouts $[15, 10, 5]$, hidden dim 256) with exact layer-wise full-neighbor inference, and are only run at $L \in \{2, 3\}$. APPNP and SGC use genuine full-graph sparse-matrix propagation (hidden dim 128) and run at $L \in \{2, 4, 8, 16\}$. $--$ marks depths outside a family's attempted range, not a failure.

Model	$L=2$	$L=3$	$L=4$	$L=8$	$L=16$
GCN	75.32 ± 0.59	77.62 ± 0.79	–	–	–
GCN+Z	$74.85 \pm 0.87^{3/10}_{0.47\downarrow}$	$75.10 \pm 0.68^{1/10\,**}_{2.53\downarrow}$	–	–	–
RESGCN	74.99 ± 0.63	76.62 ± 0.75	–	–	–
RESGCN+Z	$74.73 \pm 0.62^{3/10}_{0.25\downarrow}$	$74.22 \pm 0.85^{0/10\,**}_{2.40\downarrow}$	–	–	–
SAGE	77.63 ± 0.47	78.70 ± 0.53	–	–	–
SAGE+Z	$\mathbf{77.86 \pm 0.27^{7/10}_{0.23\uparrow}}$	$\mathbf{78.75 \pm 0.31^{5/10}_{0.06\uparrow}}$	–	–	–
APPNP	73.42 ± 0.20	–	74.19 ± 0.20	74.03 ± 0.19	73.66 ± 0.21
APPNP+Z	$\mathbf{73.59 \pm 0.15^{8/10\,*}_{0.17\uparrow}}$	–	$\mathbf{74.28 \pm 0.17^{6/10}_{0.09\uparrow}}$	$\mathbf{74.07 \pm 0.16^{5/10}_{0.04\uparrow}}$	$\mathbf{73.68 \pm 0.17^{6/10}_{0.02\uparrow}}$
SGC	65.06 ± 0.14	–	67.75 ± 0.17	68.10 ± 0.24	65.74 ± 0.30
SGC+Z	$\mathbf{66.51 \pm 0.09^{10/10\,**}_{1.45\uparrow}}$	–	$\mathbf{68.59 \pm 0.10^{10/10\,**}_{0.84\uparrow}}$	$\mathbf{68.93 \pm 0.09^{10/10\,**}_{0.83\uparrow}}$	$\mathbf{67.10 \pm 0.06^{10/10\,**}_{1.36\uparrow}}$

Table 6: Five Z-GNN adapter-integration variants for graph classification.

Variant	Overall	MUTAG	PROTEINS	NCI1
Pool-then-adapt (Table 8)	10/84 (4+/6–)	4/28 (4+/0–)	1/28 (0+/1–)	5/28 (0+/5–)
Adapt-then-pool	58/84 (0+/58–)	9/28 (0+/9–)	24/28 (0+/24–)	25/28 (0+/25–)
Pool_norm (ownCross- entropy)†	6/84 (3+/3–)	1/28 (1+/0–)	1/28 (0+/1–)	4/28 (2+/2–)
Attention-pool (ownCross- entropy)†	4/84 (0+/4–)	0/28 (0+/0–)	2/28 (0+/2–)	2/28 (0+/2–)
Hybrid ($\bar{T}$ per-node, pooled)	9/84 (3+/6–)	3/28 (3+/0–)	1/28 (0+/1–)	5/28 (0+/5–)

Table 7: Frozen-encoder probes.

	CE-trained encoder			Zentropy-trained encoder		
Cell	linear	MLP	zentropy	linear	MLP	zentropy
chameleon/APPNP/L=2	38.90	43.38	42.00	42.83	46.78	44.85
chameleon/GATv2/L=2	44.43	44.47	44.71	45.48	44.98	45.92
chameleon/GCNII/L=2	37.15	38.31	37.26	40.77	42.52	41.97
Citeseer/APPNP/L=8	23.82	22.43	25.69	67.23	67.05	67.80
Citeseer/GATv2/L=2	66.76	66.74	66.62	68.16	66.91	68.39
Cora/APPNP/L=8	24.26	28.30	26.70	71.69	73.66	72.52
Cora/SGC/L=128	31.27	34.34	35.17	61.71	67.09	63.87
Cornell/APPNP/L=2	44.86	50.81	47.57	48.11	48.11	51.35
Cornell/GCNII/L=2	46.22	45.95	47.03	46.49	45.95	48.11
squirrel/APPNP/L=2	25.42	30.38	31.39	30.18	33.34	35.65
squirrel/GCN/L=2	30.38	30.45	30.76	31.26	31.57	35.34
squirrel/GCNII/L=2	27.02	28.93	32.47	29.26	31.10	34.32
squirrel/ResGCN/L=2	31.83	32.07	32.37	32.33	32.34	35.73
Texas/SGC/L=2	60.27	61.35	60.27	59.73	61.35	61.08
Wisconsin/APPNP/L=2	54.71	57.06	54.90	52.94	51.37	52.75
Wisconsin/GATv2/L=2	53.53	52.16	52.94	53.53	52.35	52.35

Table 8: Test accuracy (%) on TUDataset graph classification (MUTAG, PROTEINS, NCI1, IMDB-BINARY, IMDB-MULTI) with the zentropy readout (Z-GNN), pool-then-adapt integration, 10-fold stratified cross-validation across the full depth sweep. Values are mean ± standard deviation over paired folds. Arrows give the change relative to Cross- entropy. * denotes $p \leq 0.05$ and ** denotes $p \leq 0.01$ under a two-sided paired sign-flip test.

Dataset	Model	$L = 2$	$L = 4$	$L = 8$	$L = 16$
MUTAG	GCN	73.39 ± 12.20	75.06 ± 12.56	72.84 ± 7.77	73.45 ± 6.89
	GCN+Z	$\mathbf{79.30 \pm 10.01^{*}_{5.91\uparrow}}$	$\mathbf{77.69 \pm 8.88_{2.63\uparrow}}$	$70.23 \pm 8.24_{2.60\downarrow}$	$\mathbf{77.25 \pm 11.10_{3.80\uparrow}}$
	RESGCN	77.69 ± 10.76	76.08 ± 12.29	76.64 ± 7.90	78.19 ± 10.50
	RESGCN+Z	$\mathbf{78.22 \pm 12.03_{0.53\uparrow}}$	$\mathbf{80.32 \pm 9.03_{4.24\uparrow}}$	$\mathbf{78.71 \pm 10.08_{2.08\uparrow}}$	$\mathbf{82.51 \pm 8.87_{4.33\uparrow}}$
	GATV2	75.03 ± 12.21	76.11 ± 11.14	71.84 ± 8.50	67.05 ± 3.01
	GATV2+Z	$\mathbf{76.05 \pm 9.79_{1.02\uparrow}}$	$\mathbf{77.16 \pm 11.38_{1.05\uparrow}}$	$\mathbf{73.95 \pm 10.40_{2.11\uparrow}}$	$\mathbf{70.70 \pm 4.75_{3.65\uparrow}}$
	SAGE	80.35 ± 10.22	76.08 ± 12.53	74.97 ± 8.00	66.49 ± 2.40
	SAGE+Z	$78.16 \pm 8.85_{2.19\downarrow}$	$\mathbf{83.04 \pm 8.43^{*}_{6.96\uparrow}}$	$\mathbf{78.19 \pm 7.26_{3.22\uparrow}}$	$\mathbf{77.16 \pm 8.88^{**}_{10.67\uparrow}}$
	APPNP	77.66 ± 11.95	65.96 ± 3.53	66.49 ± 2.40	66.49 ± 2.40
	APPNP+Z	$\mathbf{80.32 \pm 11.17_{2.66\uparrow}}$	$\mathbf{72.92 \pm 9.75_{6.96\uparrow}}$	$64.94 \pm 6.85_{1.55\downarrow}$	$\mathbf{66.49 \pm 2.40_{0.00\uparrow}}$
	SGC	75.58 ± 12.42	76.11 ± 12.44	78.77 ± 11.04	77.72 ± 12.29
	SGC+Z	$\mathbf{76.11 \pm 12.44_{0.53\uparrow}}$	$\mathbf{77.72 \pm 11.51_{1.61\uparrow}}$	$\mathbf{80.88 \pm 10.27_{2.11\uparrow}}$	$\mathbf{79.27 \pm 9.47_{1.55\uparrow}}$
	GCNII	79.82 ± 10.43	80.35 ± 10.22	80.38 ± 10.75	79.82 ± 11.55
	GCNII+Z	$\mathbf{80.35 \pm 10.22_{0.53\uparrow}}$	$\mathbf{80.38 \pm 10.81_{0.03\uparrow}}$	$79.30 \pm 10.61_{1.08\downarrow}$	$\mathbf{81.40 \pm 9.36_{1.58\uparrow}}$
PROTEINS	GCN	71.98 ± 4.77	72.50 ± 3.90	71.97 ± 4.24	71.70 ± 4.31
	GCN+Z	$\mathbf{73.31 \pm 4.99_{1.33\uparrow}}$	$71.97 \pm 3.63_{0.53\downarrow}$	$70.26 \pm 2.65_{1.71\downarrow}$	$71.34 \pm 3.07_{0.36\downarrow}$
	RESGCN	72.24 ± 3.74	71.16 ± 3.77	72.78 ± 4.40	73.22 ± 3.63
	RESGCN+Z	$71.44 \pm 4.20_{0.81\downarrow}$	$\mathbf{71.52 \pm 4.47_{0.37\uparrow}}$	$72.06 \pm 4.48_{0.72\downarrow}$	$71.42 \pm 4.87_{1.80\downarrow}$
	GATV2	72.68 ± 5.30	72.32 ± 3.85	70.98 ± 3.01	70.98 ± 3.98
	GATV2+Z	$70.44 \pm 3.73_{2.24\downarrow}$	$70.70 \pm 4.47_{1.62\downarrow}$	$69.63 \pm 4.27_{1.34\downarrow}$	$\mathbf{71.25 \pm 3.96_{0.27\uparrow}}$
	SAGE	73.31 ± 5.61	73.14 ± 4.04	71.70 ± 4.25	71.88 ± 4.13
	SAGE+Z	$70.53 \pm 6.36_{2.78\downarrow}$	$72.33 \pm 3.46_{0.81\downarrow}$	$70.80 \pm 3.29_{0.90\downarrow}$	$70.98 \pm 3.53_{0.90\downarrow}$
	APPNP	71.52 ± 4.67	71.61 ± 3.77	64.79 ± 6.89	66.04 ± 4.69
	APPNP+Z	$70.80 \pm 3.98_{0.73\downarrow}$	$71.53 \pm 4.90_{0.08\downarrow}$	$\mathbf{66.32 \pm 3.85_{1.53\uparrow}}$	$60.47 \pm 2.22^{*}_{5.57\downarrow}$
	SGC	72.42 ± 4.15	72.69 ± 3.64	72.51 ± 4.59	73.68 ± 4.59
	SGC+Z	$71.70 \pm 4.85_{0.72\downarrow}$	$71.34 \pm 5.70_{1.35\downarrow}$	$\mathbf{73.95 \pm 3.36_{1.44\uparrow}}$	$\mathbf{74.66 \pm 4.73_{0.98\uparrow}}$
	GCNII	70.44 ± 3.82	70.17 ± 4.65	70.53 ± 5.43	70.08 ± 5.32
	GCNII+Z	$\mathbf{71.43 \pm 4.36_{0.99\uparrow}}$	$\mathbf{71.61 \pm 4.78_{1.44\uparrow}}$	$70.53 \pm 5.77_{0.00\downarrow}$	$\mathbf{70.17 \pm 3.56_{0.10\uparrow}}$
NCI1	GCN	64.09 ± 1.86	62.36 ± 3.16	61.53 ± 3.61	62.21 ± 2.91
	GCN+Z	$\mathbf{64.99 \pm 2.53_{0.90\uparrow}}$	$61.75 \pm 4.28_{0.61\downarrow}$	$56.79 \pm 5.49^{*}_{4.74\downarrow}$	$\mathbf{62.53 \pm 2.52_{0.32\uparrow}}$
	RESGCN	64.21 ± 1.92	64.96 ± 2.77	64.82 ± 2.20	63.94 ± 2.39
	RESGCN+Z	$\mathbf{64.50 \pm 2.16_{0.29\uparrow}}$	$\mathbf{64.99 \pm 1.50_{0.02\uparrow}}$	$64.06 \pm 2.32_{0.75\downarrow}$	$63.43 \pm 2.11_{0.51\downarrow}$
	GATV2	63.38 ± 1.62	59.42 ± 3.79	61.22 ± 4.77	58.25 ± 6.00
	GATV2+Z	$63.11 \pm 1.76_{0.27\downarrow}$	$55.89 \pm 5.83_{3.53\downarrow}$	$54.28 \pm 5.17^{**}_{6.93\downarrow}$	$55.84 \pm 4.09_{2.41\downarrow}$
	SAGE	63.80 ± 2.25	61.12 ± 3.04	62.70 ± 2.98	62.55 ± 2.91
	SAGE+Z	$\mathbf{64.96 \pm 2.45_{1.17\uparrow}}$	$\mathbf{62.46 \pm 3.40_{1.34\uparrow}}$	$56.59 \pm 6.32^{*}_{6.11\downarrow}$	$59.71 \pm 5.27_{2.85\downarrow}$
	APPNP	62.53 ± 2.86	62.26 ± 2.88	59.83 ± 3.35	52.21 ± 2.73
	APPNP+Z	$\mathbf{62.94 \pm 2.61_{0.41\uparrow}}$	$\mathbf{62.60 \pm 3.01_{0.34\uparrow}}$	$\mathbf{61.75 \pm 4.03_{1.92\uparrow}}$	$50.00 \pm 0.17^{*}_{2.21\downarrow}$
	SGC	65.28 ± 1.98	65.18 ± 2.86	65.01 ± 2.58	64.91 ± 1.79
	SGC+Z	$65.16 \pm 1.48_{0.12\downarrow}$	$64.79 \pm 2.15_{0.39\downarrow}$	$\mathbf{65.35 \pm 2.28_{0.34\uparrow}}$	$\mathbf{65.38 \pm 2.43_{0.46\uparrow}}$
	GCNII	63.24 ± 2.18	63.33 ± 2.93	63.77 ± 2.96	62.97 ± 1.92
	GCNII+Z	$\mathbf{64.23 \pm 2.64_{1.00\uparrow}}$	$\mathbf{64.67 \pm 2.17_{1.34\uparrow}}$	$63.19 \pm 3.82_{0.58\downarrow}$	$62.36 \pm 3.59_{0.61\downarrow}$
IMDB-BINARY	GCN	49.00 ± 3.27	48.40 ± 2.88	49.30 ± 3.20	49.30 ± 1.89
	GCN+Z	$\mathbf{49.10 \pm 2.51_{0.10\uparrow}}$	$\mathbf{50.40 \pm 2.95_{2.00\uparrow}}$	$\mathbf{50.00 \pm 0.00_{0.70\uparrow}}$	$\mathbf{50.00 \pm 0.00_{0.70\uparrow}}$
	RESGCN	53.70 ± 4.97	51.50 ± 2.84	52.50 ± 5.93	49.20 ± 3.29
	RESGCN+Z	$52.70 \pm 4.62_{1.00\downarrow}$	$\mathbf{54.70 \pm 4.47_{3.20\uparrow}}$	$51.00 \pm 1.49_{1.50\downarrow}$	$\mathbf{50.90 \pm 4.01_{1.70\uparrow}}$
	GATV2	49.90 ± 1.66	49.50 ± 3.31	49.90 ± 1.20	49.00 ± 1.89
	GATV2+Z	$\mathbf{50.30 \pm 0.67_{0.40\uparrow}}$	$\mathbf{49.90 \pm 0.99_{0.40\uparrow}}$	$\mathbf{50.00 \pm 0.00_{0.10\uparrow}}$	$\mathbf{49.60 \pm 2.07_{0.60\uparrow}}$
	SAGE	50.00 ± 0.00	50.80 ± 1.62	49.50 ± 1.96	50.70 ± 1.25
	SAGE+Z	$\mathbf{50.00 \pm 0.00_{0.00\uparrow}}$	$49.80 \pm 1.55_{1.00\downarrow}$	$\mathbf{49.70 \pm 1.89_{0.20\uparrow}}$	$50.50 \pm 1.78_{0.20\downarrow}$
	APPNP	49.30 ± 3.06	49.50 ± 2.64	50.70 ± 3.83	50.10 ± 0.32
	APPNP+Z	$48.60 \pm 4.45_{0.70\downarrow}$	$\mathbf{49.60 \pm 3.06_{0.10\uparrow}}$	$48.90 \pm 3.73_{1.80\downarrow}$	$50.00 \pm 0.00_{0.10\downarrow}$
	SGC	61.10 ± 3.38	60.10 ± 3.70	58.80 ± 6.65	59.30 ± 6.04
	SGC+Z	$59.30 \pm 5.03_{1.80\downarrow}$	$58.90 \pm 4.51_{1.20\downarrow}$	$\mathbf{59.80 \pm 6.21_{1.00\uparrow}}$	$58.30 \pm 5.17_{1.00\downarrow}$
	GCNII	50.20 ± 3.52	51.10 ± 5.04	48.60 ± 2.37	51.20 ± 4.08
	GCNII+Z	$49.20 \pm 4.49_{1.00\downarrow}$	$50.00 \pm 4.03_{1.10\downarrow}$	$\mathbf{48.90 \pm 3.90_{0.30\uparrow}}$	$\mathbf{51.40 \pm 1.84_{0.20\uparrow}}$
IMDB-MULTI	GCN	33.80 ± 1.99	33.67 ± 2.04	35.53 ± 3.65	33.40 ± 0.58
	GCN+Z	$33.53 \pm 2.01_{0.27\downarrow}$	$\mathbf{34.27 \pm 3.07_{0.60\uparrow}}$	$35.27 \pm 3.36_{0.27\downarrow}$	$\mathbf{34.27 \pm 1.58_{0.87\uparrow}}$
	RESGCN	35.87 ± 3.97	36.13 ± 2.43	35.93 ± 2.50	36.60 ± 2.66
	RESGCN+Z	$35.60 \pm 3.64_{0.27\downarrow}$	$35.93 \pm 4.34_{0.20\downarrow}$	$\mathbf{36.00 \pm 3.40_{0.07\uparrow}}$	$34.20 \pm 2.31^{*}_{2.40\downarrow}$
	GATV2	32.67 ± 2.47	32.33 ± 2.38	35.00 ± 2.74	34.67 ± 2.88
	GATV2+Z	$\mathbf{34.27 \pm 1.78_{1.60\uparrow}}$	$\mathbf{33.67 \pm 3.61_{1.33\uparrow}}$	$32.87 \pm 0.89^{*}_{2.13\downarrow}$	$32.40 \pm 1.86^{*}_{2.27\downarrow}$
	SAGE	33.33 ± 0.00	33.13 ± 2.29	35.33 ± 3.37	34.73 ± 1.55
	SAGE+Z	$\mathbf{33.60 \pm 0.84_{0.27\uparrow}}$	$\mathbf{33.87 \pm 1.69_{0.73\uparrow}}$	$34.07 \pm 3.13_{1.27\downarrow}$	$34.07 \pm 4.23_{0.67\downarrow}$
	APPNP	36.40 ± 4.05	34.93 ± 2.93	35.33 ± 3.11	35.60 ± 3.21
	APPNP+Z	$35.53 \pm 3.10_{0.87\downarrow}$	$\mathbf{35.27 \pm 4.80_{0.33\uparrow}}$	$34.87 \pm 2.27_{0.47\downarrow}$	$33.67 \pm 1.05_{1.93\downarrow}$
	SGC	40.67 ± 3.76	40.27 ± 5.09	39.53 ± 4.39	38.20 ± 2.88
	SGC+Z	$39.67 \pm 3.24_{1.00\downarrow}$	$\mathbf{41.07 \pm 4.16_{0.80\uparrow}}$	$38.47 \pm 3.89_{1.07\downarrow}$	$\mathbf{39.20 \pm 3.19_{1.00\uparrow}}$
	GCNII	36.00 ± 1.89	35.27 ± 3.87	35.27 ± 3.20	34.20 ± 2.67
	GCNII+Z	$35.13 \pm 4.97_{0.87\downarrow}$	$\mathbf{36.07 \pm 2.97_{0.80\uparrow}}$	$\mathbf{35.27 \pm 2.56_{0.00\uparrow}}$	$\mathbf{35.73 \pm 3.50_{1.53\uparrow}}$

Table 9: Test accuracy (%) on WebKB (Cornell/Texas/Wisconsin) with the zentropy readout (Z-GNN).

Dataset	Model	$L=2$	$L=4$	$L=8$	$L=16$	$L=32$	$L=64$	$L=128$
Cornell	GCN	45.68 ± 9.74	42.70 ± 6.08	43.51 ± 7.03	42.70 ± 5.81	42.70 ± 5.81	43.78 ± 6.96	42.70 ± 6.34
	GCN+Z	$\mathbf{46.22 \pm 9.49_{0.54\uparrow}}$	$\mathbf{44.02 \pm 11.11^{\sharp}_{1.31\uparrow}}$	$\mathbf{47.30 \pm 8.48_{3.78\uparrow}}$	$\mathbf{47.30 \pm 7.01_{4.59\uparrow}}$	$\mathbf{46.76 \pm 6.75_{4.05\uparrow}}$	$\mathbf{45.41 \pm 6.34_{1.62\uparrow}}$	$\mathbf{45.41 \pm 7.83_{2.70\uparrow}}$
	RESGCN	45.41 ± 7.62	44.05 ± 5.85	42.16 ± 4.45	41.35 ± 5.85	43.51 ± 7.15	42.70 ± 8.24	44.32 ± 7.45
	RESGCN+Z	$\mathbf{46.22 \pm 7.26_{0.81\uparrow}}$	$\mathbf{46.72 \pm 9.06^{\sharp}_{2.66\uparrow}}$	$\mathbf{44.02 \pm 15.58^{\sharp}_{1.85\uparrow}}$	$\mathbf{42.70 \pm 5.52_{1.35\uparrow}}$	$\mathbf{44.05 \pm 5.41_{0.54\uparrow}}$	$42.47 \pm 6.75^{\sharp}_{0.23\downarrow}$	$42.43 \pm 6.38_{1.89\downarrow}$
	GATV2	49.46 ± 10.35	41.89 ± 8.94	42.70 ± 5.81	42.70 ± 5.81	42.70 ± 5.81	42.70 ± 5.81	42.70 ± 5.81
	GATV2+Z	$48.11 \pm 8.72_{1.35\downarrow}$	$\mathbf{44.59 \pm 8.08_{2.70\uparrow}}$	$\mathbf{51.08 \pm 7.26^{**}_{8.38\uparrow}}$	$\mathbf{50.81 \pm 6.08^{**}_{8.11\uparrow}}$	$\mathbf{49.73 \pm 5.87^{*}_{7.03\uparrow}}$	$\mathbf{50.81 \pm 6.08^{**}_{8.11\uparrow}}$	$\mathbf{51.35 \pm 6.62^{**}_{8.65\uparrow}}$
	SAGE	74.32 ± 4.46	62.97 ± 10.81	50.81 ± 8.72	50.81 ± 8.72	51.08 ± 8.68	50.81 ± 8.72	50.81 ± 8.72
	SAGE+Z	$72.97 \pm 5.70_{1.35\downarrow}$	$\mathbf{67.57 \pm 5.63^{\sharp}_{4.59\uparrow}}$	$49.73 \pm 8.57_{1.08\downarrow}$	$50.54 \pm 9.19_{0.27\downarrow}$	$\mathbf{51.62 \pm 8.96_{0.54\uparrow}}$	$\mathbf{51.35 \pm 8.26_{0.54\uparrow}}$	$\mathbf{51.08 \pm 8.68_{0.27\uparrow}}$
	APPNP	45.14 ± 5.26	42.70 ± 5.81	42.70 ± 5.81	45.14 ± 7.43	45.14 ± 7.96	47.30 ± 6.27	45.95 ± 7.96
	APPNP+Z	$\mathbf{52.16 \pm 7.86^{*}_{7.03\uparrow}}$	$\mathbf{45.41 \pm 6.60_{2.70\uparrow}}$	$\mathbf{45.95 \pm 9.62_{3.24\uparrow}}$	$\mathbf{52.97 \pm 8.28^{**}_{7.84\uparrow}}$	$\mathbf{48.92 \pm 7.90_{3.78\uparrow}}$	$\mathbf{49.03 \pm 7.39^{\sharp}_{1.74\uparrow}}$	$\mathbf{52.70 \pm 7.01^{*}_{5.76\uparrow}}$
	SGC	45.68 ± 5.76	44.32 ± 8.66	46.76 ± 4.94	45.14 ± 5.98	45.68 ± 5.47	45.68 ± 5.17	45.68 ± 5.17
	SGC+Z	$\mathbf{47.03 \pm 7.67_{1.35\uparrow}}$	$42.70 \pm 6.47_{1.62\downarrow}$	$44.40 \pm 6.41^{\sharp}_{2.36\downarrow}$	$44.86 \pm 6.64_{0.27\downarrow}$	$45.56 \pm 6.88^{\sharp}_{0.12\downarrow}$	$\mathbf{45.68 \pm 6.04_{0.00\uparrow}}$	$\mathbf{45.68 \pm 6.04_{0.00\uparrow}}$
	GCNII	45.95 ± 9.45	42.16 ± 8.28	42.97 ± 6.04	42.70 ± 5.81	42.70 ± 5.81	43.51 ± 6.43	42.70 ± 5.81
	GCNII+Z	$\mathbf{50.81 \pm 9.43_{4.86\uparrow}}$	$\mathbf{42.70 \pm 7.94_{0.54\uparrow}}$	$\mathbf{45.14 \pm 7.32_{2.16\uparrow}}$	$\mathbf{46.22 \pm 6.30_{3.51\uparrow}}$	$\mathbf{45.68 \pm 5.62_{2.97\uparrow}}$	$\mathbf{48.65 \pm 6.98^{\sharp}_{5.14\uparrow}}$	$\mathbf{42.97 \pm 6.04_{0.27\uparrow}}$
Texas	GCN	61.89 ± 6.55	58.38 ± 5.28	60.81 ± 5.14	58.92 ± 4.56	58.92 ± 4.56	58.92 ± 4.56	58.92 ± 4.56
	GCN+Z	$60.54 \pm 5.87_{1.35\downarrow}$	$\mathbf{59.07 \pm 3.28^{\sharp}_{0.69\uparrow}}$	$59.46 \pm 3.12^{\sharp}_{1.35\downarrow}$	$\mathbf{59.46 \pm 3.37_{0.54\uparrow}}$	$\mathbf{59.46 \pm 3.37_{0.54\uparrow}}$	$\mathbf{59.46 \pm 3.37_{0.54\uparrow}}$	$\mathbf{59.46 \pm 3.37_{0.54\uparrow}}$
	RESGCN	61.08 ± 9.12	57.57 ± 4.78	57.84 ± 6.14	58.65 ± 5.70	57.03 ± 4.31	56.49 ± 6.80	58.65 ± 4.42
	RESGCN+Z	$\mathbf{61.62 \pm 7.18_{0.54\uparrow}}$	$\mathbf{60.00 \pm 7.73_{2.43\uparrow}}$	$\mathbf{59.19 \pm 5.76_{1.35\uparrow}}$	$58.11 \pm 6.27_{0.54\downarrow}$	$\mathbf{57.84 \pm 4.80_{0.81\uparrow}}$	$\mathbf{57.30 \pm 4.38_{0.81\uparrow}}$	$\mathbf{58.92 \pm 5.22_{0.27\uparrow}}$
	GATV2	61.62 ± 6.34	58.38 ± 4.80	58.92 ± 4.56	58.92 ± 4.56	58.92 ± 4.56	58.92 ± 4.56	58.92 ± 4.56
	GATV2+Z	$60.27 \pm 8.46_{1.35\downarrow}$	$58.11 \pm 5.73_{0.27\downarrow}$	$\mathbf{59.46 \pm 3.37_{0.54\uparrow}}$	$\mathbf{59.46 \pm 3.37_{0.54\uparrow}}$	$\mathbf{58.92 \pm 4.56_{0.00\uparrow}}$	$\mathbf{59.46 \pm 3.37_{0.54\uparrow}}$	$\mathbf{59.46 \pm 3.37_{0.54\uparrow}}$
	SAGE	79.46 ± 4.97	72.97 ± 5.41	59.46 ± 4.77	60.00 ± 4.73	60.81 ± 6.53	60.27 ± 6.87	61.62 ± 6.34
	SAGE+Z	$78.92 \pm 6.08_{0.54\downarrow}$	$\mathbf{74.59 \pm 3.42_{1.62\uparrow}}$	$\mathbf{61.35 \pm 5.10_{1.89\uparrow}}$	$\mathbf{60.81 \pm 6.40_{0.81\uparrow}}$	$59.73 \pm 7.15_{1.08\downarrow}$	$60.23 \pm 6.38^{\sharp}_{0.04\downarrow}$	$60.81 \pm 6.40_{0.81\downarrow}$
	APPNP	59.19 ± 4.67	58.92 ± 4.56	58.92 ± 4.56	60.54 ± 4.26	60.54 ± 5.87	59.46 ± 4.93	60.00 ± 4.38
	APPNP+Z	$58.65 \pm 6.12_{0.54\downarrow}$	$\mathbf{58.92 \pm 4.56_{0.00\uparrow}}$	$\mathbf{61.35 \pm 7.65_{2.43\uparrow}}$	$\mathbf{61.08 \pm 5.87_{0.54\uparrow}}$	$\mathbf{62.16 \pm 4.93_{1.62\uparrow}}$	$\mathbf{62.16 \pm 6.24_{2.70\uparrow}}$	$\mathbf{62.16 \pm 5.55_{2.16\uparrow}}$
	SGC	60.27 ± 6.75	58.65 ± 6.75	58.11 ± 4.98	58.11 ± 5.44	57.57 ± 5.10	57.84 ± 5.44	57.84 ± 5.44
	SGC+Z	$\mathbf{60.54 \pm 5.58_{0.27\uparrow}}$	$\mathbf{59.19 \pm 5.32_{0.54\uparrow}}$	$\mathbf{59.46 \pm 4.23_{1.35\uparrow}}$	$\mathbf{59.73 \pm 4.31_{1.62\uparrow}}$	$\mathbf{59.73 \pm 4.31_{2.16\uparrow}}$	$\mathbf{60.00 \pm 4.56_{2.16\uparrow}}$	$\mathbf{60.00 \pm 4.56_{2.16\uparrow}}$
	GCNII	61.35 ± 6.63	59.19 ± 4.84	60.54 ± 5.13	61.35 ± 4.78	59.46 ± 4.59	59.19 ± 4.67	58.92 ± 4.56
	GCNII+Z	$\mathbf{61.62 \pm 6.08_{0.27\uparrow}}$	$\mathbf{59.73 \pm 5.76_{0.54\uparrow}}$	$\mathbf{61.62 \pm 6.22_{1.08\uparrow}}$	$\mathbf{61.62 \pm 4.38_{0.27\uparrow}}$	$\mathbf{59.73 \pm 4.84_{0.27\uparrow}}$	$\mathbf{59.73 \pm 5.17_{0.54\uparrow}}$	$\mathbf{59.73 \pm 5.17_{0.81\uparrow}}$
Wisconsin	GCN	51.76 ± 5.25	51.37 ± 6.84	48.82 ± 7.76	48.04 ± 6.99	48.24 ± 7.17	48.04 ± 6.99	48.04 ± 6.99
	GCN+Z	$\mathbf{52.94 \pm 4.24_{1.18\uparrow}}$	$50.98 \pm 5.88^{\sharp}_{0.39\downarrow}$	$\mathbf{48.82 \pm 6.57_{0.00\uparrow}}$	$\mathbf{48.04 \pm 6.99_{0.00\uparrow}}$	$48.04 \pm 6.99_{0.20\downarrow}$	$\mathbf{48.04 \pm 6.99_{0.00\uparrow}}$	$\mathbf{48.04 \pm 6.99_{0.00\uparrow}}$
	RESGCN	55.10 ± 5.02	53.53 ± 7.10	50.20 ± 6.08	53.33 ± 6.19	50.59 ± 7.38	49.80 ± 6.80	49.80 ± 5.56
	RESGCN+Z	$\mathbf{55.49 \pm 5.07_{0.39\uparrow}}$	$53.22 \pm 4.45^{\sharp}_{0.31\downarrow}$	$\mathbf{52.55 \pm 4.79_{2.35\uparrow}}$	$\mathbf{53.50 \pm 7.57^{\sharp}_{0.17\uparrow}}$	$\mathbf{51.82 \pm 4.65^{\sharp}_{1.23\uparrow}}$	$\mathbf{50.39 \pm 6.20_{0.59\uparrow}}$	$49.41 \pm 4.11_{0.39\downarrow}$
	GATV2	51.76 ± 7.23	50.98 ± 8.72	48.04 ± 6.99	48.04 ± 6.99	48.04 ± 6.99	48.04 ± 6.99	48.04 ± 6.99
	GATV2+Z	$\mathbf{54.31 \pm 6.60_{2.55\uparrow}}$	$47.65 \pm 9.83_{0.33\downarrow}$	$\mathbf{48.04 \pm 6.99_{0.00\uparrow}}$	$\mathbf{48.04 \pm 6.99_{0.00\uparrow}}$	$\mathbf{48.04 \pm 6.99_{0.00\uparrow}}$	$\mathbf{48.04 \pm 6.99_{0.00\uparrow}}$	$\mathbf{48.04 \pm 6.99_{0.00\uparrow}}$
	SAGE	81.96 ± 5.45	71.76 ± 7.80	50.20 ± 6.74	51.57 ± 5.63	50.78 ± 6.02	52.94 ± 4.53	49.41 ± 7.15
	SAGE+Z	$\mathbf{82.75 \pm 3.90_{0.78\uparrow}}$	$\mathbf{75.10 \pm 7.16_{3.33\uparrow}}$	$\mathbf{54.31 \pm 9.91_{4.12\uparrow}}$	$51.18 \pm 5.58_{0.39\downarrow}$	$\mathbf{50.78 \pm 5.18_{0.00\uparrow}}$	$51.18 \pm 7.19_{1.76\downarrow}$	$\mathbf{50.59 \pm 5.76_{1.18\uparrow}}$
	APPNP	54.90 ± 6.20	48.82 ± 6.16	49.02 ± 6.20	54.12 ± 6.21	53.73 ± 8.43	52.75 ± 8.19	53.53 ± 8.82
	APPNP+Z	$52.75 \pm 5.50_{2.16\downarrow}$	$\mathbf{50.39 \pm 6.54_{1.57\uparrow}}$	$\mathbf{52.10 \pm 7.50^{\sharp}_{3.08\uparrow}}$	$\mathbf{55.29 \pm 7.15_{1.18\uparrow}}$	$\mathbf{55.10 \pm 7.65_{1.37\uparrow}}$	$51.96 \pm 6.87_{0.78\downarrow}$	$\mathbf{54.90 \pm 5.23_{1.37\uparrow}}$
	SGC	53.92 ± 5.49	53.14 ± 6.02	52.35 ± 5.92	51.76 ± 6.99	51.76 ± 7.23	51.37 ± 7.15	51.37 ± 7.15
	SGC+Z	$52.75 \pm 4.28_{1.18\downarrow}$	$52.38 \pm 7.31^{\sharp}_{0.76\downarrow}$	$\mathbf{52.66 \pm 6.93^{\sharp}_{0.31\uparrow}}$	$\mathbf{51.96 \pm 6.35_{0.20\uparrow}}$	$\mathbf{51.76 \pm 6.35_{0.00\uparrow}}$	$\mathbf{51.76 \pm 6.35_{0.39\uparrow}}$	$\mathbf{51.76 \pm 6.35_{0.39\uparrow}}$
	GCNII	54.12 ± 6.55	53.14 ± 6.94	51.37 ± 7.03	51.96 ± 8.53	49.02 ± 7.90	49.02 ± 7.51	48.63 ± 7.21
	GCNII+Z	$54.06 \pm 3.17^{\sharp}_{0.06\downarrow}$	$51.96 \pm 6.22_{1.18\downarrow}$	$\mathbf{51.76 \pm 6.99_{0.39\uparrow}}$	$\mathbf{52.35 \pm 6.73_{0.39\uparrow}}$	$\mathbf{53.33 \pm 6.97^{*}_{4.31\uparrow}}$	$\mathbf{50.78 \pm 8.39_{1.76\uparrow}}$	$\mathbf{52.55 \pm 6.52^{*}_{3.92\uparrow}}$

Table 10: Test accuracy (%) on WikipediaNetwork (chameleon/squirrel) with the zentropy readout (Z-GNN).

Dataset	Model	$L=2$	$L=4$	$L=8$	$L=16$	$L=32$	$L=64$	$L=128$
chameleon	GCN	46.43 ± 2.34	38.66 ± 3.20	29.65 ± 5.07	25.99 ± 1.44	25.02 ± 1.97	25.70 ± 1.48	25.90 ± 1.71
	GCN+Z	$46.14 \pm 2.44_{0.29\downarrow}$	$\mathbf{39.98 \pm 3.12^{*}_{1.32\uparrow}}$	$\mathbf{30.70 \pm 4.94_{1.05\uparrow}}$	$25.35 \pm 1.30_{0.64\downarrow}$	$\mathbf{25.46 \pm 1.57_{0.44\uparrow}}$	$24.69 \pm 0.99_{1.01\downarrow}$	$25.88 \pm 1.42_{0.02\downarrow}$
	RESGCN	45.33 ± 2.92	45.35 ± 1.45	43.93 ± 1.50	38.51 ± 3.76	39.34 ± 3.95	38.44 ± 3.35	38.20 ± 2.70
	RESGCN+Z	$\mathbf{46.05 \pm 1.93_{0.72\uparrow}}$	$44.93 \pm 1.16_{0.42\downarrow}$	$43.38 \pm 2.25_{0.55\downarrow}$	$\mathbf{41.93 \pm 3.34^{*}_{3.42\uparrow}}$	$\mathbf{40.42 \pm 3.75_{1.07\uparrow}}$	$\mathbf{38.93 \pm 2.13_{0.48\uparrow}}$	$\mathbf{38.38 \pm 2.40_{0.18\uparrow}}$
	GATV2	44.54 ± 1.38	37.50 ± 4.51	21.49 ± 2.71	22.24 ± 2.38	22.28 ± 1.93	22.85 ± 1.62	22.85 ± 1.62
	GATV2+Z	$\mathbf{46.14 \pm 2.05^{**}_{1.60\uparrow}}$	$\mathbf{42.15 \pm 3.93_{4.65\uparrow}}$	$\mathbf{24.65 \pm 1.15^{\sharp}_{3.16\uparrow}}$	$\mathbf{25.44 \pm 1.90^{\sharp}_{3.20\uparrow}}$	$\mathbf{24.74 \pm 1.20^{**}_{2.46\uparrow}}$	$\mathbf{24.67 \pm 1.14^{\sharp}_{1.82\uparrow}}$	$\mathbf{24.61 \pm 1.30^{\sharp}_{1.76\uparrow}}$
	SAGE	48.88 ± 2.32	44.69 ± 4.21	26.80 ± 1.39	26.64 ± 1.28	26.62 ± 1.36	26.78 ± 1.26	26.64 ± 1.41
	SAGE+Z	$48.68 \pm 2.17_{0.20\downarrow}$	$43.64 \pm 4.42_{1.05\downarrow}$	$\mathbf{27.17 \pm 1.44_{0.37\uparrow}}$	$\mathbf{27.19 \pm 1.49_{0.55\uparrow}}$	$26.62 \pm 1.36_{0.00\downarrow}$	$26.54 \pm 1.62_{0.24\downarrow}$	$\mathbf{27.21 \pm 1.39_{0.57\uparrow}}$
	APPNP	36.54 ± 1.56	23.66 ± 2.15	27.35 ± 2.01	22.59 ± 3.96	22.32 ± 2.25	21.84 ± 2.52	22.04 ± 2.53
	APPNP+Z	$\mathbf{40.33 \pm 2.09^{**}_{3.79\uparrow}}$	$23.60 \pm 1.11_{0.07\downarrow}$	$\mathbf{33.46 \pm 1.94^{**}_{6.12\uparrow}}$	$\mathbf{45.52 \pm 1.76^{**}_{22.93\uparrow}}$	$\mathbf{45.64 \pm 2.17^{**}_{23.31\uparrow}}$	$\mathbf{44.98 \pm 1.70^{**}_{23.14\uparrow}}$	$\mathbf{44.63 \pm 2.01^{**}_{22.59\uparrow}}$
	SGC	43.46 ± 2.25	43.42 ± 2.22	42.43 ± 1.80	43.25 ± 1.86	43.00 ± 1.92	43.05 ± 1.91	43.40 ± 2.24
	SGC+Z	$43.16 \pm 2.44_{0.31\downarrow}$	$42.50 \pm 1.97_{0.92\downarrow}$	$\mathbf{43.29 \pm 2.15_{0.86\uparrow}}$	$43.09 \pm 2.71_{0.15\downarrow}$	$42.61 \pm 2.44_{0.39\downarrow}$	$42.68 \pm 2.45_{0.37\downarrow}$	$42.70 \pm 2.33_{0.70\downarrow}$
	GCNII	37.87 ± 2.98	35.22 ± 1.79	35.83 ± 1.86	32.02 ± 2.22	32.04 ± 2.22	32.43 ± 2.21	32.65 ± 2.00
	GCNII+Z	$\mathbf{40.90 \pm 2.55_{3.03\uparrow}}$	$\mathbf{38.44 \pm 2.79^{**}_{3.22\uparrow}}$	$\mathbf{36.60 \pm 2.57_{0.77\uparrow}}$	$\mathbf{33.16 \pm 2.06^{**}_{1.14\uparrow}}$	$\mathbf{33.66 \pm 1.87^{**}_{1.62\uparrow}}$	$\mathbf{33.36 \pm 2.26_{0.92\uparrow}}$	$\mathbf{33.16 \pm 2.62_{0.50\uparrow}}$
squirrel	GCN	30.21 ± 1.68	31.80 ± 0.74	30.72 ± 0.65	30.66 ± 0.64	31.17 ± 1.06	31.08 ± 0.84	30.87 ± 0.56
	GCN+Z	$\mathbf{34.53 \pm 1.84^{**}_{4.32\uparrow}}$	$31.23 \pm 1.75_{0.57\downarrow}$	$\mathbf{32.02 \pm 1.10^{**}_{1.30\uparrow}}$	$\mathbf{31.51 \pm 0.91^{*}_{0.85\uparrow}}$	$\mathbf{31.30 \pm 1.32_{0.12\uparrow}}$	$\mathbf{31.42 \pm 1.27_{0.35\uparrow}}$	$\mathbf{31.43 \pm 1.26_{0.56\uparrow}}$
	RESGCN	31.41 ± 1.41	31.92 ± 1.52	30.31 ± 0.78	31.37 ± 1.46	31.46 ± 1.61	30.44 ± 2.28	29.57 ± 1.37
	RESGCN+Z	$\mathbf{35.40 \pm 1.50^{**}_{5.09\uparrow}}$	$\mathbf{34.48 \pm 1.49^{**}_{2.56\uparrow}}$	$\mathbf{30.47 \pm 2.65_{0.16\uparrow}}$	$30.52 \pm 1.58_{0.85\downarrow}$	$30.79 \pm 2.67_{0.67\downarrow}$	$\mathbf{31.77 \pm 1.25_{1.33\uparrow}}$	$29.26 \pm 1.75_{0.31\downarrow}$
	GATV2	25.24 ± 2.06	25.30 ± 4.19	20.83 ± 1.44	20.83 ± 1.44	19.36 ± 1.17	19.90 ± 1.57	20.00 ± 1.13
	GATV2+Z	$\mathbf{34.46 \pm 2.62^{**}_{9.21\uparrow}}$	$\mathbf{33.32 \pm 0.90^{**}_{8.02\uparrow}}$	$\mathbf{30.52 \pm 0.86^{**}_{9.69\uparrow}}$	$\mathbf{30.27 \pm 1.36^{**}_{9.44\uparrow}}$	$\mathbf{29.16 \pm 1.66^{**}_{9.81\uparrow}}$	$\mathbf{27.47 \pm 3.60^{**}_{7.57\uparrow}}$	$\mathbf{30.21 \pm 1.49^{**}_{10.21\uparrow}}$
	SAGE	31.93 ± 1.84	32.80 ± 1.33	30.37 ± 0.70	30.37 ± 0.70	30.37 ± 0.70	30.37 ± 0.70	30.37 ± 0.70
	SAGE+Z	$\mathbf{33.15 \pm 1.08_{1.22\uparrow}}$	$32.56 \pm 1.21_{0.23\downarrow}$	$\mathbf{30.56 \pm 0.93_{0.18\uparrow}}$	$\mathbf{30.46 \pm 0.78_{0.09\uparrow}}$	$\mathbf{30.52 \pm 0.77_{0.14\uparrow}}$	$\mathbf{30.50 \pm 0.77_{0.12\uparrow}}$	$\mathbf{30.50 \pm 0.77_{0.12\uparrow}}$
	APPNP	24.43 ± 1.49	21.39 ± 1.86	22.66 ± 3.00	23.59 ± 3.45	21.11 ± 1.68	21.04 ± 1.90	21.61 ± 2.82
	APPNP+Z	$\mathbf{32.91 \pm 0.72^{**}_{8.48\uparrow}}$	$\mathbf{30.15 \pm 1.21^{**}_{8.76\uparrow}}$	$\mathbf{28.93 \pm 2.30^{**}_{6.27\uparrow}}$	$\mathbf{30.75 \pm 1.70^{**}_{7.16\uparrow}}$	$\mathbf{31.74 \pm 1.86^{**}_{10.62\uparrow}}$	$\mathbf{31.59 \pm 2.65^{**}_{10.56\uparrow}}$	$\mathbf{31.57 \pm 1.71^{**}_{8.95\uparrow}}$
	SGC	29.72 ± 0.97	31.40 ± 0.93	31.74 ± 1.22	31.99 ± 1.38	31.82 ± 1.24	32.11 ± 0.94	31.93 ± 1.17
	SGC+Z	$\mathbf{30.92 \pm 1.35^{*}_{1.20\uparrow}}$	$\mathbf{31.49 \pm 1.08_{0.09\uparrow}}$	$\mathbf{31.74 \pm 0.85_{0.00\uparrow}}$	$31.88 \pm 1.28_{0.11\downarrow}$	$31.54 \pm 1.12_{0.28\downarrow}$	$31.97 \pm 0.97_{0.14\downarrow}$	$31.72 \pm 0.91_{0.21\downarrow}$
	GCNII	26.09 ± 1.78	25.97 ± 2.12	22.07 ± 2.82	23.65 ± 3.23	23.53 ± 3.01	23.71 ± 2.71	24.09 ± 3.45
	GCNII+Z	$\mathbf{34.74 \pm 1.39^{**}_{8.65\uparrow}}$	$\mathbf{33.24 \pm 1.30^{**}_{7.27\uparrow}}$	$\mathbf{31.07 \pm 0.53^{**}_{8.00\uparrow}}$	$\mathbf{30.66 \pm 0.63^{**}_{7.01\uparrow}}$	$\mathbf{30.23 \pm 1.29^{**}_{6.71\uparrow}}$	$\mathbf{30.81 \pm 0.80^{**}_{7.10\uparrow}}$	$\mathbf{30.49 \pm 1.36^{**}_{6.40\uparrow}}$