LOCAL–GLOBAL GEOMETRIC INSIGHTS FOR GRAPH NEURAL NETWORKS VIA ENTROPIC CURVATURE

Anonymous authors
Paper under double-blind review

ABSTRACT

Curvature notions on graphs, particularly Ollivier–Ricci and Forman curvature, have emerged as powerful tools for addressing fundamental issues in Graph Neural Networks (GNNs), such as oversmoothing and oversquashing. While these metrics capture the stability and efficiency of information propagation, they primarily rely on local neighborhood comparisons and may overlook broader global features. In contrast, we introduce *Entropic Curvature*, which extends the Lott–Sturm–Villani framework to discrete spaces by focusing on the global displacement of probability measures along W_1-geodesics in Wasserstein space. By relating entropy, optimal transport, and curvature lower bounds, this framework yields transport-entropy and Poincaré-type inequalities that establish thermodynamic regularity on information flow. We prove that positive entropic curvature provides quantitative guarantees against feature collapse and adversarial distribution shifts. Furthermore, we establish a fundamental expansion paradox, proving that sparsity, strong spectral expansion, and positive entropic curvature cannot coexist in large graphs, revealing that the fast communication required to prevent oversquashing inherently forces a move toward negative curvature and instability. Empirically, we introduce a tractable Weak Entropic Curvature proxy and an E-Gate mechanism that modulates message-passing based on this global prior, improving accuracy on node classification benchmarks over vanilla GNNs. Entropic Structural Encodings derived from κ_w outperform LCP, LAPE, RWPE, and EGO baselines.

1 INTRODUCTION

Graphs are increasingly analyzed through the lens of geometry, with curvature emerging as a key concept in understanding graph structure and dynamics (Bauer et al., 2015b;a; 2017; Bohar, 1989; Chung et al., 2014; Chung & Yau, 2017; Gong et al., 2019; Lin et al., 2011; Lin & Yau, 2010). Recently, there has been increasing attention to incorporating geometric insights into GNNs. In this context, *graph curvature* offers an efficient way to capture phenomena such as bottlenecks (Topping et al., 2022; Elhag et al., 2022), characterize community structures (Ni et al., 2019; Jost & Liu, 2014; Lubberts et al.), or to propose new architectures that leverage curvature to improve performance (Chen et al., 2025; Li et al., 2022; Bachmann et al., 2020).

In this context, discrete notions of curvature on graphs, e.g., Ollivier-Ricci Curvature (Ollivier, 2009) and Forman (Forman, 2003) curvature, have gained increasing attraction as analytical tools. Such curvature measures assign a value to each edge or local neighborhood, reflecting how tight the connection is. Previous works have used these notions to analyze fundamental GNN problems (Tori et al.), such as oversmoothing (Nguyen et al., 2023), where deep GNN layers produce indistinguishable node features, and oversquashing, where long-range information is compressed and lost when passed through graph bottlenecks (Topping et al., 2022). In particular, it has been found that positive Ollivier-Ricci Curvature along edges correlates with severe oversmoothing. This local geometric perspective yielded unified explanations for both issues and inspired curvature-based approaches.

In this work, we study another type of discrete curvature: *entropic curvature*, a theoretic notion of curvature grounded in optimal transport theory and the Lott–Sturm-Villani framework (Lott & Villani, 2009; Sturm, 2006). In contrast to edge-based curvatures, entropic curvature evaluates the curvature of a subgraph as a metric-measure space by examining how probability distributions evolve over the graph via diffusion or transport. Our main contributions are as follows,

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.

- **Rigorous Discrete Reformulation (Section 3):** We formally adapt the Lott–Sturm–Villani framework to graph-structured data, defining curvature through the convexity of entropy along W_1-Wasserstein geodesics. We introduce a tractable *Weak Entropic Curvature* proxy κ_w computable over 2-hop neighborhoods that provably lower-bounds the global entropic curvature (Theorem 3.3).
- **Anti-Oversmoothing Guarantee via Poincaré Inequality (Section 7):** We prove that positive κ_w implies a global Poincaré-type inequality that quantitatively bounds how fast feature variance decays under repeated aggregation (Theorem 7.1). This gives the first oversmoothing guarantee derived from a global, transport-based curvature notion.
- **The Expansion Paradox (Section 4):** We prove that sparsity, strong spectral expansion, and positive entropic curvature cannot coexist simultaneously in large graphs (Theorem 4.2). This reveals a fundamental trade-off: the fast long-range communication required to prevent oversquashing inherently forces negative curvature and propagation instability.
- **Curvature-Controlled Generalization (Section 5):** We derive a generalization bound showing the train-test risk gap is controlled by κ_w via a transport-entropy inequality (Theorem 5.1). Higher positive curvature tightens the bound, limiting the model's sensitivity to distributional shift between training and test node sets.
- **Curvature-Informed GNN Architectures (Sections 4 and 6):** We introduce two practical architectures. The *E-Gate* mechanism modulates message-passing aggregation using κ_w as a thermodynamic gate, improving accuracy on 5 of 6 node classification benchmarks over vanilla GCN. *Entropic Structural Encodings* (ENT) outperform LCP, LAPE, RWPE, and EGO baselines.

2 Background and Related Work

Ricci Curvature. Ricci curvature is a fundamental concept in differential geometry that describes how a manifold deviates from being flat. One way to detect curvature is through *holonomy* (Kobayashi & Nomizu, 1996), which measures how parallel transport around closed loops fails to preserve the geometrical data being carried, thereby revealing twists intrinsic to the manifold. Another approach involves *geodesic dispersion*; if nearby geodesics with identical initial velocities diverge, the space is positively curved; if they remain parallel, it is flat; and if they converge, it is negatively curved (Do Carmo & Flaherty Francis, 1992). A major advance in extending Ricci curvature to more general metric measure spaces beyond smooth manifolds was provided by Sturm (Sturm, 2006) and Lott and Villani (Lott & Villani, 2009) in two independent works. Building on McCann's pioneering work (McCann, 1997), they established a synthetic definition of Ricci curvature via optimal transport and entropy. Specifically, let (Ω, d) be a complete and separable metric space, and let $\mathcal{P}(\Omega)$ be the space of Borel probability measures on Ω. For $1 \leq p < \infty$ and $\mu, \nu \in \mathcal{P}(\Omega)$ we consider the Monge-Kantorovich (Wasserstein) distance W_p is defined by,

$$\mathrm{W}_p(\mu,\nu) := \inf_{\Gamma \in \Pi(\mu,\nu)} \left(\int_{\Omega\times\Omega} d(x,y)^p \mathrm{d}\Gamma(x,y) \right)^{1/p},$$

where $\Pi(\mu,\nu)$ denotes the set of all couplings of μ and ν. $\mathrm{W}_p(\mu,\nu)^p$ can be interpreted as the minimal cost required to transfer a distribution of mass from an initial configuration μ to another configuration ν, where the transportation cost per unit mass is given by $d(x,y)^p$. In this synthetic framework, relative entropy plays a key role in capturing curvature lower bounds. With a measure $\nu \in \mathcal{P}(\Omega)$ and a probability measure μ that is absolutely continuous with respect to ν, the relative entropy of μ with respect to ν is defined by $\mathrm{Ent}_\nu(\mu) := \int_\Omega \rho(x)\log\rho(x)\mathrm{d}\nu(x)$, where $\rho = \mathrm{d}\mu/\mathrm{d}\nu$.

The following result, proved in (Cordero-Erausquin et al., 2001; Otto & Villani, 2000; Renesse & Sturm, 2005), characterizes Ricci curvature lower bounds on Riemannian manifolds in terms of convexity properties of the relative entropy (with respect to the volume measure vol) and optimal transport.

Theorem 2.1 (Characterisation of Ricci lower bounds)**.** *Let $\kappa \in \mathbb{R}$. For a complete Riemannian manifold $\mathcal{M}$ with a Ricci Curvature* Ric*, the following assertions are equivalent,*

1. $\mathrm{Ric} \geq \kappa$ *on* $\mathcal{M}$.
2. *Each pair of probability measures $\mu_0, \mu_1 \in \mathcal{P}_2(\mathcal{M})$ can be connected by a constant speed W_2-geodesic $(\mu_t)_{t\in[0,1]}$, i.e., $W_2(\mu_s,\mu_t) = |t-s|\cdot W_2(\mu_0,\mu_1)$, along which the entropy satisfies the κ-convexity inequality,*

$$\mathrm{Ent}_{\mathrm{vol}}(\mu_t) \leq (1-t)\,\mathrm{Ent}_{\mathrm{vol}}(\mu_0) + t\,\mathrm{Ent}_{\mathrm{vol}}(\mu_1) - \frac{\kappa}{2}t(1-t)\mathrm{W}_2(\mu_0,\mu_1)^2. \tag{1}$$

This inequality establishes a connection between the Ricci curvature Ric of a Riemannian manifold and the convexity of entropy along optimal transport paths. This means positive curvature enhances the convexity of entropy, while negative curvature weakens it, reflecting how the manifold's geometry influences the behavior of entropy under optimal transport.

Graph Curvatures. Since graphs are discrete, and non-differentiable structures, the classical notion of Ricci curvature from differential geometry cannot be directly applied. Instead, discrete analogues of curvature have been introduced to mimic aspects of Ricci curvature. One such measure is Forman curvature (Forman, 2003) which assigns scalar weights to edges or nodes, providing a purely combinatorial curvature notion and offering insights primarily into local topology or geometry. By contrast, Ollivier-Ricci curvature (Ollivier, 2009) uses a probabilistic viewpoint by comparing single-step neighborhood mass transport between adjacent vertices, capturing local bottlenecks and concentration but ignoring how measures evolve over longer paths. Another Ricci-type notion is Bakry-Émery curvature (Bakry & Émery, 2006), which is evaluated at vertices rather than edges. It measures the local cohesion of a network around each vertex, so that higher cohesion corresponds to more positive curvature. Finally, the Balanced Forman curvature refines the original Forman curvature by incorporating symmetrical weighting factors (Topping et al., 2022), aiming to balance the contributions of vertices and edges in a way that can provide a more nuanced assessment of a graph's local and global curvature-like properties.

Graph Neural Networks (GNNs). A graph is denoted by $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$, where $\mathcal{V}$ is the node set, $\mathcal{E}$ is the edge set, and $\mathbf{X}$ contains node features. Each node $u \in \mathcal{V}$ is initialized with $h_u^{(0)} = x_u$. A broad class of message-passing GNNs updates node representations through two operations: an aggregation step $m_u^{(\ell)} = \mathrm{AGG}^{(\ell)}(\{h_v^{(\ell)} : v \in \mathcal{N}(u) \cup \{u\}\})$, and an update step $h_u^{(\ell+1)} = \mathrm{UPDATE}^{(\ell)}(h_u^{(\ell)}, m_u^{(\ell)})$. Here, $\mathrm{AGG}^{(\ell)}$ collects information from the local neighborhood, while $\mathrm{UPDATE}^{(\ell)}$ combines the aggregated message with the current node representation.

Generalization of GNNs. A fundamental GNN challenge in GNNs is characterizing their generalization capabilities, particularly when the training and testing data arise from distinct topological distributions on the graph (Tang & Liu, 2023; Abbahaddou et al., 2025). In the standard node classification setting, we treat the training and evaluation sets not merely as indices, but as discrete probability measures supported on the vertex set $\mathcal{V}$. Let $\mathcal{V}_{\text{train}}, \mathcal{V}_{\text{test}} \subset \mathcal{V}$ denote the labeled training and testing node sets, respectively. We define the empirical training distribution μ_{train} and the evaluation distribution μ_{test} as, $\mu_{\text{train}} := \frac{1}{|\mathcal{V}_{\text{train}}|}\sum_{u\in\mathcal{V}_{\text{train}}}\delta_u$, and $\mu_{\text{test}} := \frac{1}{|\mathcal{V}_{\text{test}}|}\sum_{u\in\mathcal{V}_{\text{test}}}\delta_u$, where δ_u represents the Dirac measure centered at node u. Consider a GNN $f_\theta : (\mathcal{V}, \mathcal{E}, X) \to \mathcal{Y}$ parameterized by θ, and a pointwise loss function $\mathcal{J}(\cdot,\cdot)$. The expected risk with respect to an arbitrary node distribution μ is given by

$$\mathcal{L}_\mu(f_\theta) = \mathbb{E}_{u\sim\mu}[\mathcal{J}(f_\theta(u), y_u)].$$

While the optimization objective is to minimize the empirical risk $\mathcal{L}_{\mu_{\text{train}}}(f_\theta)$, the ultimate goal is to minimize the risk on the test distribution. The generalization gap quantifies the discrepancy between these risks for the learned parameters $\hat{\theta} = \arg\min_\theta \mathcal{L}_{\mu_{\text{train}}}(f_\theta)$. Formally, this gap is defined as the absolute deviation between the test and training risks, $\Delta\mathcal{L}(f_{\hat{\theta}}) := \left|\mathcal{L}_{\mu_{\text{test}}}(f_{\hat{\theta}}) - \mathcal{L}_{\mu_{\text{train}}}(f_{\hat{\theta}})\right|$. As shown in Section 5, the entropic curvature of the underlying graph imposes strict bounds on this sensitivity via transport-entropy inequalities.

3 Entropic Curvature on Graphs

Notations and Assumptions. We consider a connected, undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathrm{X})$. We denote the combinatorial distance between vertices $u, v \in \mathcal{V}$ by $d(u, v)$, defined as the length of the shortest path connecting them. We assume $\mathcal{G}$ is *locally finite*, meaning every vertex has a finite degree. For any node $u \in \mathcal{V}$, we define the local 2-ball as the subgraph induced by nodes within a distance of at most two, denoted $\mathrm{B}_2(u) := \{y \in \mathcal{V} \mid d(u, y) \leq 2\}$. The 2-sphere is the set of nodes at a distance of exactly two from u, denoted $\mathrm{S}_2(u) := \{w \in \mathcal{V} \mid d(u, w) = 2\}$. Let $G(x, y)$ be the set of all geodesic paths joining x to y (i.e. all shortest paths from x to y). For any pair of vertices $u, v \in \mathcal{V}$, we denote by $[u, v] := \{z \in \mathcal{V} | z \in \gamma, \gamma \in G(x.y)\}$ and $]u, v[:= [x, y]\backslash\{x, y\}$ the set of *strict midpoints* lying on any shortest path between them. A *graph space* is defined as the tuple $(\mathcal{V}, d, \mathtt{m}, \mathtt{L})$, where $\mathtt{m}\colon \mathcal{V} \to \mathbb{R}_+$ is a positive reference measure and $\mathtt{L}\colon \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ is a generator

encoding the jump rates of a reversible Markov process on $\mathcal{V}$. The operator $\mathtt{L}$ is constructed to satisfy three key properties,

(i) **Conservation:** The diagonal terms ensure probability conservation, i.e., $\mathtt{L}(u,u) = -\sum_{v\neq u}\mathtt{L}(u,v)$.
(ii) **Sparsity:** We require $\mathtt{L}(u,v) > 0$ if and only if $u \sim v$. This ensures that $\mathtt{L}$ acts as a Graph Shift Operator (GSO) Dasoulas et al.; Gavili & Zhang (2017); Abbahaddou (2026), preserving the sparsity of $\mathcal{G}$.
(iii) **Reversibility:** The measure $\mathtt{m}$ satisfies the detailed balance condition, i.e., $\mathtt{m}(u)\mathtt{L}(u,v) = \mathtt{m}(v)\mathtt{L}(v,u) \quad \forall u,v \in \mathcal{V}$. This constraint implies that $\mathtt{L}$ is self-adjoint with respect to the inner product induced by m, guaranteeing a real spectrum.

The selection of the reference measure m inherently dictates the structure of the generator $\mathtt{L}$ through the reversibility condition. This relationship is illustrated in Example 1. While derivations for generators under general positive measures and concentrated approximations are provided in Appendices O and P, the present work focuses on the canonical pair $(\mathrm{m}_0, \mathrm{L}_0)$.

Example 1 (Uniform Measure)**.** Let $\mathcal{C} \subseteq \mathcal{V}$. In the standard case where $\mathcal{C} = \emptyset$ or $\mathcal{C} = \mathcal{V}$, we adopt the uniform measure $\mathtt{m}_0(u) = 1$. To satisfy detailed balance, the corresponding generator is simply the unnormalized Combinatorial Laplacian adjacency weights, i.e., $\mathrm{L}_0(u,v) := \mathbf{1}_{\{u,v\}\in\mathcal{E}} \quad$ for $u \neq v$.

Theoretical Definition of Entropic Curvature. The concept of entropic curvature provides a powerful way to extend Ricci-type geometric ideas to discrete spaces, particularly graphs. Recall from the Sturm, Lott and Villani framework, c.f. Section 2, that lower bounds on Ricci curvature can be captured by the convexity of the entropy functional along Wasserstein geodesics. This carries over to the discrete setting via the following definition,

Definition 3.1 (Entropic Curvature)**.** The *entropic curvature* $\kappa \in \mathbb{R}$ is the largest constant such that for every pair of probability measures μ_0, μ_1 on $\mathcal{V}$, there exists a constant-speed W_1-geodesic $(\mu_t)_{t\in[0,1]}$, i.e., $\mathrm{W}_1(\mu_s,\mu_t) = |t-s|\cdot\mathrm{W}_1(\mu_0,\mu_1)$, for which the entropy functional satisfies the κ-convexity inequality,

$$\mathrm{Ent}_{\mathtt{m}}(\mu_t) \leq (1-t)\,\mathrm{Ent}_{\mathtt{m}}(\mu_0) + t\,\mathrm{Ent}_{\mathtt{m}}(\mu_1) \quad - \frac{\kappa}{2}t(1-t)\mathrm{W}_1(\mu_0,\mu_1)^2. \tag{2}$$

This definition mirrors the role of Ricci curvature in smooth geometry; if $\kappa > 0$, then the entropy is uniformly displacement convex, which in turn imposes geometric rigidity on the space. To make the computation of curvature feasible in large graphs, we build on the work of Samson (2022), introducing a local proxy for κ, known as the *weak entropic curvature*. This approximation relies on evaluating the behavior of entropy in neighborhoods of radius two.

The classical Lott–Sturm–Villani framework characterises Ricci curvature lower bounds on smooth Riemannian manifolds via displacement convexity along constant-speed W_2-geodesics (Theorem 2.1). In the discrete graph setting, W_2-geodesics require interpolating measures that generally lack combinatorial interpretations and are computationally intractable. Following Rapaport & Samson (2024), we instead use W_1-geodesics, which are well-defined on finite graphs via linear programming and are computable in polynomial time. This substitution is theoretically grounded: on smooth manifolds, a Ricci lower bound is equivalent to displacement convexity along *both* W_1- and W_2-geodesics (Lott & Villani, 2009). In the discrete regime, the W_1 metric directly enables the transport-entropy inequality (Lemma F.2) which underpins all downstream theoretical results in Sections 7–5.

Formal definition of Weak Entropic Curvature. Let $z \in \mathcal{V}$ be a vertex and $\mathcal{W} \subseteq \mathcal{V}$ a target set, typically the 2-sphere $\mathrm{S}_2(z)$. For any $z'' \in \mathcal{W}$, we define the two-step transition weight $\mathtt{L}^2(z,z'') := \sum_v \mathtt{L}(z,v)\mathtt{L}(v,z'')$ and the relative path weight $\ell(z,z',z'') := \frac{\mathtt{L}(z,z')\mathtt{L}(z',z'')}{\mathtt{L}^2(z,z'')}$. The local functional $\mathrm{H}_{\mathtt{L}}(z,\mathrm{W})$ is defined as,

$$\mathrm{H}_{\mathtt{L}}(z,\mathrm{W}) := \sup_{\alpha\in\mathcal{P}} \left\{ \sum_{z''\in W} \mathtt{L}^2(z,z'') \quad \cdot \prod_{z'\in]z,z''[} \left(\frac{\alpha(z')}{\mathtt{L}(z,z')}\right)^{2\ell(z,z',z'')} \right\}, \tag{3}$$

Here, $]z,W[:= \bigcup_{z''\in W}]z,z''[$ denotes the union of all midpoints. The supremum is taken over all functions $\alpha :]z,W[\to \mathbb{R}_+$ satisfying the normalization condition $\sum_{v\in]z,W[}\alpha(v) = 1$. Beyond this formal definition, weak entropic curvature admits a useful geometric interpretation in terms of local

transport redundancy. In Appendix A, we show that the local functional $\mathrm{H_L}$ can be viewed through an entropy/KL-divergence lens: it measures how well the graph's two-hop geodesic structure supports the optimal redistribution of mass across intermediate nodes. This perspective reveals that high-curvature regions correspond to locally redundant, diffusion-friendly neighborhoods, whereas low- or negative-curvature regions indicate bottleneck-like structures where transport is forced through fewer fragile paths. We also relate this interpretation to a Motzkin–Straus-type characterization (Motzkin & Straus, 1965), showing how local clique density influences the curvature proxy.

Inspired by the functional form of *Boltzmann Entropy*, we define the *local weak entropic curvature*, a tractable local proxy for the entropic curvature defined, c.f., Definition 3.2.

Definition 3.2 (Weak Entropic Curvature)**.** Let $u \in \mathcal{V}$. We define the *local weak entropic curvature* at node z as, $\kappa_w(u) := -2\log \mathrm{H_L}(u, \mathrm{S}_2(u))$, and the *weak entropic curvature* of the graph as the infimum over all vertices, i.e., $\kappa_w(\mathcal{G}) := \inf_{u\in\mathcal{V}} \kappa_w(u)$.

While the global curvature examines the convexity of entropy along the entire Wasserstein space, the local weak entropic curvature focuses on the displacement of mass within a controlled two-hop neighborhood. This approximation allows for efficient computation while preserving the core geometric insights of the Lott-Sturm-Villani framework. The formal link between this local proxy and the global entropic curvature is established in Theorem 3.3.

Theorem 3.3 (Proof in Appendix E)**.** *The weak entropic curvature κ_w provides a rigorous lower bound for the global entropic curvature κ, i.e., $\kappa \geq \kappa_w$.*

Theorem 3.3 shows that the computable κ_w is not merely a heuristic local statistic: it is a certified lower bound on the global entropic curvature κ. Thus, although κ is defined through entropy convexity over the full Wasserstein space, its geometric effect can already be detected through two-hop transport structure. Intuitively, large positive values of κ_w arise in locally redundant regions where mass can be redistributed through many alternative midpoints, while low or negative values indicate bottleneck-like neighborhoods in which transport is forced through a small number of fragile paths. To illustrate this, Appendix B reports κ_w on canonical graph motifs, ranging from dense clustered structures to sparse trees. These examples show that weak entropic curvature separates diffusion-friendly local traps from negatively curved expansion and bottleneck regimes, providing the geometric intuition behind the oversmoothing and oversquashing results developed in the following sections.

Algorithmic Implementation. We evaluate the practical utility of entropic curvature through a node-wise implementation of the weak entropic curvature proxy. Algorithm 2 in Appendix M.1 operationalizes this by transforming the theoretical displacement convexity requirement into a constrained optimization problem. Specifically, for each node, we precompute the relative path weights $\ell(z, u, w)$, which represent the local transport efficiency across two-hop geodesics. The core of the algorithm involves solving a smooth concave maximization for the local functional $\mathrm{H_L}$ over the probability simplex $\Delta_{|\mathrm{S}_1(z)|}$. To achieve high-fidelity convergence at scale, we use the Sequential Least Squares Programming (SLSQP) solver (Nocedal & Wright, 2006). This gradient-based approach leverages the smoothness of the objective function $\mathrm{J}(\alpha)$, typically reaching a stable curvature estimate within a small number of iterations. A detailed analysis of the algorithm is provided in Appendices M and Q.

4 Expansion Paradox and Rewiring

A central question in GNN design is whether a graph can simultaneously achieve three desirable properties: *sparsity* (bounded degree, ensuring scalable message-passing), *spectral expansion* (fast mixing, preventing oversquashing), and *positive entropic curvature* (geometric regularity, preventing oversmoothing). Our main result of this section shows these three properties are fundamentally incompatible in the large-graph limit, a structural impossibility theorem with direct consequences for GNN design. To state it precisely, let $\mathbb{P}_{\mathcal{G}}$ denote the transition matrix of the lazy simple random walk on $\mathcal{G}$, $\mathbb{P}_{\mathcal{G}}(u, v) = \frac{1}{2}\delta_{uv} + \frac{\mathbf{1}_{\{u,v\}\in\mathcal{E}}}{2\deg(u)}$, and let $\lambda_2(\mathcal{G})$ be its second-largest eigenvalue. A small λ_2 indicates strong mixing; a large λ_2 (close to 1) signals slow convergence and bottlenecks.

Definition 4.1 (Perfect Graph)**.** Given constants $\Delta \geq 1$ and $\rho < 1$, a graph $\mathcal{G}$ is called *perfect* with respect to (Δ, ρ) if it simultaneously satisfies: *(A) Sparsity*, $\Delta(\mathcal{G}) \leq \Delta$; *(B) Spectral expansion*, $\lambda_2(\mathcal{G}) \leq \rho$; and *(C) Non-negative curvature*, $\kappa_w(\mathcal{G}) \geq 0$.

A perfect graph, as defined in 4.1, would be ideal for a GNN: cheap to run, free of bottlenecks, and geometrically stable. The following result, which extends (Salez, 2022) to our entropic curvature framework, proves that no sufficiently large graph can be perfect.

Theorem 4.2 (Expansion Paradox; Proof in Appendix I)**.** *Let $(\mathcal{G}_n)_{n\geq 1}$ be a sequence of finite graphs satisfying the sparsity condition* $\sup_{n\geq 1}\left\{\frac{1}{|\mathcal{V}_n|}\sum_{x\in\mathcal{V}_n}\deg_{\mathcal{G}_n}(x)\log\deg_{\mathcal{G}_n}(x)\right\}<\infty$*, and suppose that $\kappa_{\mathcal{G}_n}(x)\geq a>0$ for all but $o(|\mathcal{V}_n|)$ vertices. Then, for all $\rho<1$,* $\liminf_{n\to\infty}\frac{1}{|\mathcal{V}_n|}\#\{i:\lambda_i(\mathcal{G}_n)\geq\rho\}>0$.

Theorem 4.2 resolves a long-standing tension in GNN design. Graphs that are sparse and spectrally well-connected, exactly those that appear in standard benchmarks such as citation, social, and web networks, *must* exhibit negative entropic curvature. Since negative κ_w permits rapid diffusion of probability mass, it explains why long-range message-passing on these graphs is prone to oversquashing. Conversely, positive curvature controls this diffusion via the Poincaré inequality of Theorem 7.1, preventing oversmoothing. The paradox thus unifies oversmoothing and oversquashing as two manifestations of the same geometric trade-off: they are not independent failure modes, but opposite ends of a single curvature spectrum.

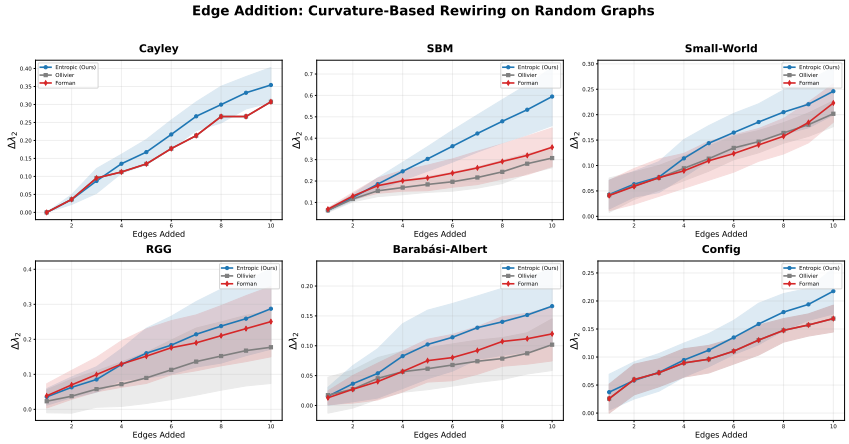

Figure 1: Spectral gap improvement $\Delta\lambda_2$ under curvature-guided edge addition (10 edges, 10 seeds, shaded ± 1 std). MCR (ours) consistently outperforms ORC and FRC across all six graph families.

Curvature-Based Rewiring. The expansion paradox shows that sparse graphs cannot simultaneously be strongly expanding and positively curved in the large-graph regime. This does not mean, however, that negative curvature is fixed by the input topology. Since κ_w is determined by the two-hop transport geometry around each node, one can locally modify the graph by adding edges that create additional geodesic midpoints. The purpose of curvature-based rewiring is therefore not to make the graph uniformly positively curved, but to reduce the severity of negative-curvature bottlenecks while preserving sparsity. The following result gives the local principle behind our rewiring strategy: adding alternative two-hop routes cannot decrease the weak entropic curvature at the rewired node.

Theorem 4.3 (Curvature Monotonicity Under Rewiring; Proof in Appendix L)**.** *Let $\mathcal{G}^+$ be obtained from $\mathcal{G}$ by adding edges that create new two-hop geodesics from $z\in\mathcal{V}$ to nodes in $\mathrm{S}_2(z)$, without removing any existing two-hop geodesic. Define the midpoint set $\mathrm{M}_{\mathcal{G}}(z,w):=\{m\in\mathcal{V}:z\sim m,\ m\sim w\}$. If $\mathrm{M}_{\mathcal{G}}(z,w)\subseteq\mathrm{M}_{\mathcal{G}^+}(z,w)$ for every $w\in\mathrm{S}_2(z)$, then $\kappa_w^{\mathcal{G}^+}(z)\geq\kappa_w^{\mathcal{G}}(z)$, with strict improvement whenever $|\mathrm{M}_{\mathcal{G}^+}(z,w)|>|\mathrm{M}_{\mathcal{G}}(z,w)|$ for some w.*

The theorem formalizes a simple geometric mechanism: increasing the number of admissible midpoints makes two-hop transport less concentrated, thereby improving the local entropic-curvature certificate. Based on this principle, our *Midpoint-Completion Rewiring* (MCR) procedure repeatedly targets the most negatively curved region and adds an edge that creates an additional midpoint for a poorly supported two-hop connection. We give the implementation details in Appendix C.

This monotonicity result gives a local curvature certificate for each accepted rewiring step. Its downstream consequences follow from the earlier theory: when rewiring increases the graph-level lower bound κ_w, Theorem 3.3 yields a tighter certificate for the global entropic curvature, and in

the positive-curvature regime Theorem 5.1 gives a sharper transport–entropy generalization bound. The effect on spectral expansion is evaluated empirically, since the expansion paradox identifies a structural trade-off rather than a stepwise monotonicity guarantee. Figure 1 shows that MCR consistently improves the spectral gap more than Ollivier–Ricci (ORC) and Forman–Ricci (FRC) rewiring across all six random graph families. The gains are especially pronounced on community-structured graphs, where adding midpoint-completion edges directly targets transport bottlenecks between regions rather than relying only on one-hop or degree-based curvature information.

5 Entropic Curvature and Generalization

A central question in GNN theory is how well a model trained on one part of a graph transfers to another, especially when the training and test nodes occupy different regions of the topology (Wang et al.). We formalize this discrepancy by viewing the training and test sets as probability measures μ_{train} and μ_{test} on $\mathcal{V}$. The generalization gap $\Delta\mathcal{L}(f_\theta) := |\mathcal{L}_{\mu_{\text{test}}}(f_\theta) - \mathcal{L}_{\mu_{\text{train}}}(f_\theta)|$ then measures the sensitivity of the learned predictor to a distributional shift over the graph. Entropic curvature provides a natural way to control this sensitivity: through transport–entropy inequalities, curvature bounds convert the cost of moving from μ_{train} to μ_{test} into an explicit bound on the change in risk.

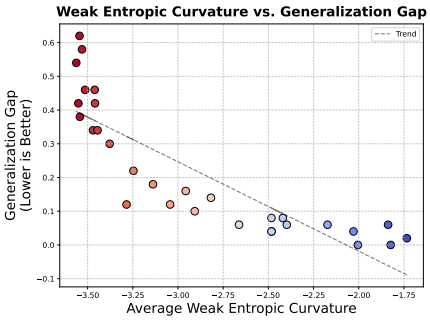

Figure 2: Empirical correlation between $\bar{\kappa}_w$ and the GNN generalization gap on SBM graphs.

Theorem 5.1 (Generalization Bound under Positive Entropic Curvature; Proof in Appendix F)**.** *Assume that $\mathcal{G}$ has weak entropic curvature bounded below by $\kappa_w > 0$. Suppose the loss $\mathcal{J}$ is σ-Lipschitz and the GNN f_θ is globally Λ-Lipschitz with respect to the graph distance d, i.e., $\|f_\theta(x) - f_\theta(y)\| \le \Lambda d(x, y) \qquad \forall x, y \in \mathcal{V}$. Then, for any test distribution μ_{test},*

$$|\mathcal{L}_{\mu_{\text{test}}}(f_\theta) - \mathcal{L}_{\mu_{\text{train}}}(f_\theta)| \le \sigma\Lambda\sqrt{\frac{2}{\kappa_w}\,\mathrm{Ent}_{\mu_{\text{train}}}(\mu_{\text{test}})}.$$

Theorem 5.1 separates the generalization gap into three interpretable factors. The product $\sigma\Lambda$ captures the sensitivity of the loss and the model, while $\mathrm{Ent}_{\mu_{\text{train}}}(\mu_{\text{test}})$ measures the informational cost of shifting from the training distribution to the test distribution. The curvature term $1/\sqrt{\kappa_w}$ acts as a geometric amplifier: larger positive curvature yields a tighter certificate, making the model less sensitive to the same train–test distributional shift. Positive curvature gives the cleanest transport–entropy bound, but many sparse real-world graphs contain negatively curved regions. The next result shows that a meaningful generalization certificate still holds when curvature is negative, at the cost of a diameter-dependent factor.

Theorem 5.2 (Generalization Bound under Negative Entropic Curvature; Proof in Appendix G)**.** *Suppose $\mathcal{G}$ is finite and connected with diameter $D < \infty$, and assume its weak entropic curvature satisfies $\kappa_w \ge -K$ for some $K > 0$. Let the loss be L-Lipschitz with respect to the predictions $f_\theta(x)$, and let f_θ be globally Λ-Lipschitz with respect to the graph distance d. Then,*

$$|\mathcal{L}_{\mu_{\text{test}}}(f_\theta) - \mathcal{L}_{\mu_{\text{train}}}(f_\theta)| \le L\Lambda\sqrt{\frac{4}{K}\left(e^{KD^2/8} - 1\right)\mathrm{Ent}_{\mu_{\text{train}}}(\mu_{\text{test}})}.$$

Together, Theorems 5.1 and 5.2 show that entropic curvature controls generalization in both curvature regimes. When $\kappa_w > 0$, the dependence scales as $1/\sqrt{\kappa_w}$. When $\kappa_w \ge -K$, the bound remains finite but deteriorates with both the negative-curvature scale K and the graph diameter D, reflecting the greater sensitivity of negatively curved, expansive regions to distributional shift.

Empirical Analysis of Generalization and Curvature. To empirically evaluate the relationship between graph geometry and learning performance, we analyze the correlation between the Average Weak Entropic Curvature (κ_w) and the Generalization Gap. Following the framework in Theorem 5.1, we define the generalization gap as the absolute difference between training and testing accuracies $\Delta\mathcal{A} = \text{Acc}_{\text{train}} - \text{Acc}_{\text{test}}$. This experiment tests the hypothesis that positive curvature lower bounds

Table 1: Transposed version of the E-Gate results table. Test accuracy (%, mean ± std) of baseline backbones and their E-Gate variants. Best results between baseline and E-Gate are shown in bold.

	Cora		CiteSeer		PubMed		Cornell		Wisconsin	
Backbone	Base	E-Gate	Base	E-Gate	Base	E-Gate	Base	E-Gate	Base	E-Gate
GCN	88.08 ± 1.04	**88.28 ± 1.86**	76.23 ± 0.76	**76.43 ± 0.94**	87.29 ± 0.47	**87.92 ± 0.58**	41.35 ± 6.05	**44.05 ± 9.75**	43.73 ± 4.81	**46.86 ± 7.41**
SAGE	87.56 ± 1.58	**87.64 ± 1.16**	75.11 ± 1.95	**75.47 ± 1.47**	88.45 ± 0.27	**88.69 ± 0.53**	61.89 ± 3.51	**67.30 ± 5.73**	75.10 ± 5.95	**78.82 ± 3.90**
GIN	**83.03 ± 2.53**	82.84 ± 2.18	69.55 ± 2.25	**69.97 ± 1.98**	84.45 ± 4.85	**86.26 ± 0.46**	42.16 ± 8.39	**45.95 ± 6.84**	**53.33 ± 5.67**	50.39 ± 7.75
GAT	86.77 ± 1.27	**87.01 ± 1.14**	74.79 ± 1.28	**75.02 ± 1.45**	**86.26 ± 0.37**	86.08 ± 0.39	40.00 ± 8.09	**42.70 ± 8.18**	47.84 ± 6.27	**51.76 ± 6.51**

effectively tighten the model's sensitivity to distributional shifts between the training and testing measures. Our pipeline utilizes a Stochastic Block Model (SBM) to generate 30 distinct graphs, each containing 100 nodes divided into two equal communities. We systematically modulate the graph's curvature by varying the intra-community edge probability p_{in} on a sliding scale from 5% to 10%, while maintaining a fixed inter-community probability $p_{out} = 5\%$. This transitions the topology from sparse, tree-like regimes to dense, clustered configurations. For each graph, we compute the node-wise weak entropic curvature by solving the exact optimization for the local functional $\mathrm{H_L}$. The global signature is defined as the mean curvature across all vertices: $\bar{\kappa}_w = \frac{1}{|\mathcal{V}|}\sum_{z\in\mathcal{V}} \kappa_w(z)$. We use a standard 2-layer GCN (Kipf & Welling, 2017), and the node features are initialized as an identity matrix. The model is trained for 200 epochs using Adam optimizer to predict community membership (Kingma, 2014).

As illustrated in Figure 2, the empirical results reveal a significant negative correlation: higher average weak entropic curvature consistently corresponds to a lower generalization gap. This trend provides concrete evidence for the theoretical prediction in Theorem 5.1. Specifically, as κ_w increases, the transport-entropy inequality tightens, limiting how much the testing distribution can stray from the training configuration without incurring a performance penalty.

6 Entropic Curvature-Based Architectures

Adapting GNN Dynamics to Entropic Geometry. We next evaluate whether κ_w can be used as a lightweight geometric prior for standard GNNs. We consider node classification on benchmark graphs spanning homophilic regimes, Cora, CiteSeer, and PubMed (Yang et al., 2016), and heterophilic regimes, Cornell, Texas, and Wisconsin (Pei et al., 2020). Motivated by Theorem 7.1, we introduce *E-Gate*, a node-wise modulation that can be applied to any aggregation-based GNN backbone. Given a backbone aggregation $\mathrm{AGG}_u^{(\ell)}$, E-Gate rescales it as $h_u^{(\ell+1)} = \mathrm{UPDATE}^{(\ell)}(h_u^{(\ell)}, \mathcal{R}(u)\,\mathrm{AGG}_u^{(\ell)}), \quad \mathcal{R}(u) = \exp(-\tau\kappa_w(u))$, where τ is a learnable temperature parameter. Thus, E-Gate is not tied to GCN: it can be inserted before the update step of GCN, GraphSAGE, GIN, GAT, or other message-passing backbones Feng et al. (2022); Hamilton et al. (2017); Veličković et al. (2018). Implementation details are given in Appendix H.

Table 1 shows that adding E-Gate generally improves performance across backbones and graph regimes. The gains are most pronounced on heterophilic datasets such as Cornell and Wisconsin, suggesting that curvature-aware modulation is especially useful when standard aggregation struggles with bottlenecks or non-homophilic neighborhoods. On homophilic citation graphs, the improvements are smaller but mostly positive, indicating that E-Gate does not substantially disrupt already effective local smoothing. Occasional drops, e.g., for specific backbone, dataset pairs, suggest that curvature gating is most beneficial when the graph structure contains meaningful transport heterogeneity, rather than when local neighborhood aggregation is already sufficient.

Entropic Curvature as a Structural Encoding While GNNs are powerful, their expressivity is bounded by the 1-Weisfeiler-Lehman test (Feng et al., 2022), limiting their ability to distinguish structurally distinct nodes that share identical local neighborhoods. To bridge this gap, recent works have proposed augmenting node features with structural Encodings derived from spectral properties or based on discrete geometry. Notably, Local Curvature Profiles (LCP) demonstrated that injecting Ollivier-Ricci curvature statistics can significantly enhances GNN performance (Fesser & Weber). However, discrete curvatures like Ollivier-Ricci primarily capture local clustering and may miss broader transport bottlenecks. We hypothesize that Weak Entropic Curvature, which is derived from the convexity of entropy along Wasserstein geodesics, provides a more globally-aware structural prior. By serving as a proxy for the stability of information flow, κ_w offers a richer topological signature than purely combinatorial metrics. We evaluate the utility of Entropic Curvature as a structural encoding (ENT) against strong baselines: Laplacian Eigenmaps (LAPE) (Belkin & Niyogi, 2003), Random

Walk PE (RWPE) (Eliasof et al., 2023), Ego-Network encodings (EGO) (Alvarez-Gonzalez et al., 2023), and the curvature-based LCP (Fesser & Weber). Following the LCP protocol, we augment the initial node features X with a 3-dimensional curvature signature. This new vector captures both the intrinsic geometry of the node and the geometric context of its immediate neighborhood. Additional details could be found in Appendix N.

Table 2: Structural encoding accuracy (%, mean±std, GCN backbone). **Bold**: best per row.

Dataset	ENT (ours)	LCP	LAPE	RWPE	EGO
Cornell	$\mathbf{42.60 \pm 3.54}$	41.96 ± 4.44	38.91 ± 2.42	38.47 ± 3.53	41.09 ± 3.81
Wisconsin	44.28 ± 3.22	$\mathbf{47.14 \pm 3.42}$	42.22 ± 2.95	43.65 ± 5.24	42.70 ± 2.67
PubMed	$\mathbf{85.98 \pm 0.23}$	85.76 ± 0.20	85.37 ± 0.18	85.77 ± 0.18	85.71 ± 0.26
Texas	$\mathbf{46.95 \pm 4.48}$	43.04 ± 1.92	43.48 ± 4.52	43.04 ± 3.52	45.43 ± 3.86
CiteSeer	$\mathbf{71.86 \pm 0.93}$	71.13 ± 0.78	71.34 ± 0.92	71.63 ± 1.08	70.97 ± 0.98

Table 2 presents the node classification accuracy across six benchmarks. ENT achieves the best performance on 4 out of 5 datasets, outperforming both spectral and discrete-curvature baselines. On heterophilic graphs like Texas (+3.9% vs. LCP) and Cornell (+0.6% vs. LCP), where bottlenecks are frequent, the entropic prior proves significantly more informative than local metrics. On other graphs, ENT maintains a consistent lead, suggesting that our curvature successfully capture relevant community structures.

7 Entropic Curvature and Oversmoothing

Oversmoothing refers to the progressive collapse of node representations in deep GNNs, where repeated aggregation drives features toward indistinguishable vectors (Li et al., 2018; Cai & Wang, 2020). Prior work has related this to local graph curvature notions such as Ollivier–Ricci and Forman curvature (Nguyen et al., 2023; Topping et al., 2022; Fesser & Weber, 2024). We show that the same geometric intuition extends to entropic curvature, but now through global functional inequalities rather than local edge comparisons. The key object is the one-sided Dirichlet energy $\mathcal{E}(f) := \int_{\mathcal{V}} \sup_{v \sim u} [f(u) - f(v)]_+^2 \, d\mu(u)$, which measures the amount of local variation remaining in a graph signal. Theorem 7.1 gives a global certificate for feature collapse: when local aggregation reduces the Dirichlet energy, the total feature variance is forced to decrease, with a rate controlled by the curvature lower bound. Larger positive κ_w tightens this control, meaning that high-curvature regions are more susceptible to rapid homogenization under repeated smoothing. Unlike spectral-gap results derived from Ollivier–Ricci curvature (Nguyen et al., 2023), this bound applies directly to arbitrary graph signals and therefore to intermediate GNN representations.

Theorem 7.1 (Poincaré-type Inequality under Positive Entropic Curvature; Proof in Appendix J)**.** *Let $(\mathcal{V}, d, \mathfrak{m}, \mathtt{L})$ be a finite connected graph space with $\kappa_w > 0$, and let $\mu := \mathfrak{m}/\mathfrak{m}(\mathcal{V})$. For any graph signal $f : \mathcal{V} \to \mathbb{R}$, we have* $\mathrm{Var}_\mu(f) \leq \frac{1}{\kappa_w} \int_{\mathcal{V}} \sup_{v \sim u} [f(u) - f(v)]_+^2 \, \mathrm{d}\mu(u)$.

The complementary question is what happens in negatively curved regions, which are typical in sparse graphs with strong expansion by Theorem 4.2. In this regime, the geometry has the opposite effect: expansion creates many directions for mass to spread, yielding a lower bound on feature variance whenever local variation remains.

Theorem 7.2 (Inequality under Negative Entropic Curvature; Proof in Appendix K)**.** *Let $(\mathcal{V}, d, \mathfrak{m}, \mathtt{L})$ be a finite connected graph space with weak entropic curvature bounded below by $\kappa_w \geq -K$ for some $K > 0$, and let $\mu := \mathfrak{m}/\mathfrak{m}(\mathcal{V})$. Let $D := \max_{u,v \in \mathcal{V}} d(u, v)$ denote the graph diameter. Then for any graph signal $f : \mathcal{V} \to \mathbb{R}$, we have* $\mathrm{Var}_\mu(f) \geq \frac{e^{-KD^2/2}}{K} \int_{\mathcal{V}} \sup_{v \sim u} [f(v) - f(u)]_+^2 \, d\mu(u)$.

8 Conclusion

This work introduces entropic curvature as a global and theoretical measure to analyze GNNs. In particular, a positive entropic-curvature lower bound yields log-Sobolev and transport–entropy inequalities that directly control information propagation in message-passing architectures. These inequalities translate into quantitative guarantees against oversmoothing, limits on distributional shift, and provable generalization. Entropic curvature offers a unifying geometric principle that enhance our theoretical understanding of GNN dynamics and yields concrete algorithmic benefits. We hope this framework will inspire further exploration of global geometric insights, not only for generalization and oversmoothing, but also for curriculum learning on evolving graphs, principled regularization of large-scale graph models, and the broader integration of optimal-transport ideas into machine learning.

AI USE STATEMENT

In this work, we used generative AI tools for writing assistance, specifically limited to paraphrasing and polishing the grammar, tone, and readability of author-written draft text. We have not used generative AI tools for research ideation, mathematical derivations, experimental design, software implementation, data generation, analysis, or the synthesis of primary literature, and other generative AI research pipeline tasks are not applicable to this work. Additionally, we used generative AI tools solely for the recommended disclosure of linguistic surface-level editing. We have reviewed all AI-assisted text in detail by manually verifying that every paraphrased sentence accurately preserved our intended technical meaning without introducing hallucinated claims or distorting factual accuracy. We take full responsibility for the final content of this work, including all text, claims, and artifacts produced with the aid of generative AI.

REFERENCES

Yassine Abbahaddou. Which graph shift operator? a spectral answer to an empirical question. *arXiv preprint arXiv:2602.06557*, 2026.

Yassine Abbahaddou, Sofiane ENNADIR, Johannes F Lutzeyer, Michalis Vazirgiannis, and Henrik Boström. Bounding the expected robustness of graph neural networks subject to node feature attacks. In *The Twelfth International Conference on Learning Representations*.

Yassine Abbahaddou, Fragkiskos D. Malliaros, Johannes F. Lutzeyer, Amine M. Aboussalah, and Michalis Vazirgiannis. Graph neural network generalization with gaussian mixture model based augmentation. In *Forty-second International Conference on Machine Learning*, 2025.

Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

Nurudin Alvarez-Gonzalez, Andreas Kaltenbrunner, and Vicenç Gómez. Beyond weisfeiler–lehman with local ego-network encodings. *Machine Learning and Knowledge Extraction*, 5(4):1234–1265, 2023.

László Babai. Spectra of cayley graphs. *Journal of Combinatorial Theory, Series B*, 27(2):180–189, 1979.

Gregor Bachmann, Gary Bécigneul, and Octavian Ganea. Constant curvature graph convolutional networks. In *International conference on machine learning*, pp. 486–496. PMLR, 2020.

Dominique Bakry and Michel Émery. Diffusions hypercontractives. In *Séminaire de Probabilités XIX 1983/84: Proceedings*, pp. 177–206. Springer, 2006.

F. Bauer, B. Hua, and S.-T. Yau. Davies-Gaffney-Grigor'yan lemma on graphs. *Comm. Anal. Geom.*, 23(5):1031–1068, 2015a.

F. Bauer, Y. Horn P. Lin, G. Lippner, D. Mangoubi, and S.-T. Yau. Li-Yau inequality on graphs. *J. Differential Geom.*, 99(3):359–405, 2015b.

F. Bauer, B. Hua, and S.-T. Yau. Sharp Davies-Gaffney-Grigor'yan lemma on graphs. *Math. Ann.*, 368(3-4):1429–1437, 2017.

Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.

Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

Itai Benjamini, Russell Lyons, and Oded Schramm. Unimodular random trees. *Ergodic Theory Dynam. Systems*, 35(2):359–373, 2015.

M. Bohar. Isoperimetric numbers of graphs. *Journal of Combinatorial Theory*, Series B 47(3): 274–291, 1989.

Lewis Bowen, Michael Chapman, Alexander Lubotzky, and Thomas Vidick. The aldous–lyons conjecture i: Subgroup tests. *arXiv preprint arXiv:2408.00110*, 2024.

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

Jialong Chen, Bowen Deng, Chuan Chen, Zibin Zheng, et al. Graph neural ricci flow: Evolving feature from a curvature perspective. In *The Thirteenth International Conference on Learning Representations*, 2025.

F. Chung and S.-T. Yau. A strong Harnack inequality for graphs. *Comm. Anal. Geom.*, 25(3):557–588, 2017.

F. Chung, Y. Lin, and S.-T. Yau. Harnack inequalities for graphs with nonnegative Ricci curvature. *J. Math. Anal. Appl.*, 415(1):25–32, 2014.

Fan RK Chung. *Spectral graph theory*, volume 92. American Mathematical Soc., 1997.

Tobias Holck Colding and William P Minicozzi II. Liouville properties. *arXiv preprint arXiv:1902.09366*, 2019.

Dario Cordero-Erausquin, Robert J McCann, and Michael Schmuckenschläger. A riemannian interpolation inequality à la borell, brascamp and lieb. *Inventiones mathematicae*, 146(2):219–257, 2001.

George Dasoulas, Johannes F Lutzeyer, and Michalis Vazirgiannis. Learning parametrised graph shift operators. In *International Conference on Learning Representations*.

Manfredo Perdigao Do Carmo and J Flaherty Francis. *Riemannian geometry*, volume 2. Springer, 1992.

Ahmed AA Elhag, Gabriele Corso, Hannes Stärk, and Michael M Bronstein. Graph anisotropic diffusion for molecules. In *ICLR2022 Machine Learning for Drug Discovery*, 2022.

Moshe Eliasof, Fabrizio Frasca, Beatrice Bevilacqua, Eran Treister, Gal Chechik, and Haggai Maron. Graph positional encoding via random feature propagation. In *International conference on machine learning*, pp. 9202–9223. PMLR, 2023.

Sofiane Ennadir, Yassine Abbahaddou, Johannes F Lutzeyer, Michalis Vazirgiannis, and Henrik Boström. A simple and yet fairly effective defense for graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 21063–21071, 2024.

Michael Fekete. Über die verteilung der wurzeln bei gewissen algebraischen gleichungen mit ganzzahligen koeffizienten. *Mathematische Zeitschrift*, 17(1):228–249, 1923.

Jiarui Feng, Yixin Chen, Fuhai Li, Anindya Sarkar, and Muhan Zhang. How powerful are k-hop message passing graph neural networks. *Advances in Neural Information Processing Systems*, 35: 4776–4790, 2022.

Lukas Fesser and Melanie Weber. Effective structural encodings via local curvature profiles. In *The Twelfth International Conference on Learning Representations*.

Lukas Fesser and Melanie Weber. Mitigating over-smoothing and over-squashing using augmentations of forman-ricci curvature. In *Learning on Graphs Conference*, pp. 19–1. PMLR, 2024.

Matthias Fey and Jan E. Lenssen. Fast graph representation learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.

Forman. Bochner's method for cell complexes and combinatorial ricci curvature. *Discrete & Computational Geometry*, 29:323–374, 2003.

Bailey K Fosdick, Daniel B Larremore, Joel Nishimura, and Johan Ugander. Configuring random graph models with fixed degree sequences. *Siam Review*, 60(2):315–355, 2018.

Shmuel Friedland and Hans Schneider. Spectra of expansion graphs. *The Electronic Journal of Linear Algebra*, 6:2–10, 1999.

Adnan Gavili and Xiao-Ping Zhang. On the shift operator, graph frequency, and optimal filtering in graph signal processing. *IEEE Transactions on Signal Processing*, 65(23):6303–6318, 2017.

C. Gong, Y. Lin, S. Liu, and S.-T. Yau. Li-Yau inequality for unbounded Laplacian on graphs. *Adv. Math.*, 357:33pp, 2019.

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

Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

Jürgen Jost and Shiping Liu. Ollivier's ricci curvature, local clustering and curvature-dimension inequalities on graphs. *Discrete & Computational Geometry*, 51(2):300–322, 2014.

Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

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

Shoshichi Kobayashi and Katsumi Nomizu. *Foundations of differential geometry, volume 2*, volume 2. John Wiley & Sons, 1996.

Haifeng Li, Jun Cao, Jiawei Zhu, Yu Liu, Qing Zhu, and Guohua Wu. Curvature graph neural network. *Information Sciences*, 592:50–66, 2022.

Qimai Li, Zhichao Han, and Xiao-Ming Wu. Deeper insights into graph convolutional networks for semi-supervised learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Y. Lin and S.-T. Yau. Discrete and computational geometry. 2003.

Y. Lin and S.-T. Yau. Ricci curvature and eigenvalue estimate on locally finite graphs. *Math. Res. Lett.*, 17(2):343–356, 2010.

Y. Lin, L. Lu, and S.-T. Yau. Ricci curvature of graphs. *Tohoku Math. J (2)*, 63(4):605–627, 2011.

J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math.*, 169(3):903–991, 2009.

Zachary Lubberts, Yu Tian, and Melanie Weber. Curvature-based clustering on graphs. In *2024 Joint Mathematics Meetings (JMM 2024)*. AMS.

R. J. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128(1):153–179, 1997.

Theodore S Motzkin and Ernst G Straus. Maxima for graphs and a new proof of a theorem of turán. *Canadian Journal of Mathematics*, 17:533–540, 1965.

Khang Nguyen, Nong Minh Hieu, Vinh Duc Nguyen, Nhat Ho, Stanley Osher, and Tan Minh Nguyen. Revisiting over-smoothing and over-squashing using ollivier-ricci curvature. In *International Conference on Machine Learning*, pp. 25956–25979. PMLR, 2023.

Chien-Chun Ni, Yu-Yao Lin, Feng Luo, and Jie Gao. Community detection on networks with ricci flow. *Scientific reports*, 9(1):9984, 2019.

Jorge Nocedal and Stephen J Wright. *Numerical optimization*. Springer, 2006.

Yann Ollivier. Ricci curvature of markov chains on metric spaces. *Journal of Functional Analysis*, 256(3):810–864, 2009.

F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173(2):361–400, 2000.

Louis M Pecora and Thomas L Carroll. Master stability functions for synchronized coupled systems. *Physical review letters*, 80(10):2109, 1998.

Hongbin Pei, Bingzhe Wei, Kevin Chen Chuan Chang, Yu Lei, and Bo Yang. Geom-gcn: Geometric graph convolutional networks. In *8th International Conference on Learning Representations, ICLR 2020*, 2020.

Mathew Penrose. *Random geometric graphs*, volume 5. OUP Oxford, 2003.

Martin Rapaport and Paul-Marie Samson. Criteria for entropic curvature on graph spaces. *ArXiv preprint: 2303.15874*, 2024.

M.-K. Von Renesse and K.-Th. Sturm. Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm. Pure Appl. Math.*, 58(7):923–940, 2005.

Justin Salez. Sparse expanders have negative curvature. *Geom. Funct. Anal*, 32:1486–1513, 2022.

Paul-Marie Samson. Entropic curvature on graphs along schrödinger bridges at zero temperature. *Probability Theory and Related Fields*, 184(3):859–937, 2022.

Daniel A Spielman. Algorithms, graph theory, and linear equations in laplacian matrices. In *Proceedings of the International Congress of Mathematicians 2010 (ICM 2010) (In 4 Volumes) Vol. I: Plenary Lectures and Ceremonies Vols. II–IV: Invited Lectures*, pp. 2698–2722. World Scientific, 2010.

K.-Th. Sturm. On the geometry of metric measure spaces. I and II. *Acta Math.*, 196(1):65–177, 2006.

Huayi Tang and Yong Liu. Towards understanding generalization of graph neural networks. In *International Conference on Machine Learning*, pp. 33674–33719. PMLR, 2023.

Jake Topping, Francesco Di Giovanni, Benjamin Paul Chamberlain, Xiaowen Dong, and Michael M. Bronstein. Understanding over-squashing and bottlenecks on graphs via curvature. In *International Conference on Learning Representations*, 2022.

Floriano Tori, Vincent Holst, and Vincent Ginis. The effectiveness of curvature-based rewiring and the role of hyperparameters in gnns revisited. In *The Thirteenth International Conference on Learning Representations*.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, Yoshua Bengio, et al. Graph attention networks. In *International conference on learning representations*, volume 6, 2018.

Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, 17(3):261–272, 2020.

Zhiyang Wang, Juan Cervino, and Alejandro Ribeiro. A manifold perspective on the statistical generalization of graph neural networks. In *Forty-second International Conference on Machine Learning*.

Ethan Wolfe. The effect of watts-strogatz and barabási-albert graphs on memory formation. Master's thesis, California Polytechnic State University, 2024.

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.

A Interpreting Weak Entropic Curvature via Local Transport

Interpretation of the Local Functional. To provide a geometric interpretation of $\mathrm{H}_{\mathtt{L}}$, it is instructive to specialize the functional to the case where the target set is a singleton, $\mathrm{W} = \{z''\}$, for some $z'' \in \mathrm{S}_2(z)$. The support of the optimization restricts to the set of strict midpoints $]z, z''[$. In this setting, the functional $\mathrm{H}_{\mathtt{L}}(z, \{z''\})$ simplifies to,

$$\mathrm{H}_{\mathtt{L}}(z, \{z''\}) = \mathtt{L}^2(z, z'') \sup_{\alpha} \prod_{z' \in]z, z''[} \left(\frac{\alpha(z')}{\mathtt{L}(z, z')} \right)^{2\ell(z, z', z'')}.$$

By applying the transformation $-2\log(\cdot)$. We convert the product maximization into a summation minimization. This yields the following energy functional,

$$-2\log \mathrm{H}_{\mathtt{L}}(z, z'') = \inf_{\alpha} \left[-2 \sum_{z' \in]z, z''[} \ell(z, z', z'') \quad \times \log \left(\frac{\alpha(z')}{\mathtt{L}(z, z')} \right) \right] + \mathrm{C}_{\mathtt{L}}, \tag{4}$$

where $\mathrm{C}_{\mathtt{L}} = -2\log \mathtt{L}^2(z, z'')$ is a constant independent of α. Recall that $\sum_{z'} \ell(z, z', z'') = 1$ and $\sum_{z'} \alpha(z') = 1$. By introducing the entropy of the geometric weights $\ell(z, \cdot, z'')$ into the sum, the equation equation 4 can be rewritten directly in terms of the Kullback–Leibler divergence $\mathcal{KL}(\cdot \parallel \cdot)$,

$$\inf_{\alpha} \left[2 \sum_{z' \in]z, z''[} \ell(z, z', z'') \log \left(\frac{\ell(z, z', z'')}{\alpha(z')} \right) \right] + \mathrm{C}'_L \quad = 2 \inf_{\alpha} \mathcal{KL}\Big(\ell(z, \cdot, z'') \parallel \alpha(\cdot) \Big) + \mathrm{C}'_L,$$

where C'_L is a constant independent from $\alpha(\cdot)$. Here, we use the property that both the path weights ℓ and the transport plan α sum to unity over the midpoints. This derivation reveals that the weak entropic curvature locally penalizes the divergence between the geometry of the graph (encoded by the relative path weights ℓ) and the optimal transport plan (α). When we return to the full definition where $\mathrm{W} = \mathrm{S}_2(z)$, the functional $\mathrm{H}_{\mathtt{L}}(z, \mathrm{S}_2(z))$ acts as a weighted regularization of these pairwise divergences over all possible targets z''.

Geometric Characterization via the Motzkin–Straus Theorem. To provide more geometric intuition for the local functional $\mathrm{H}_{\mathtt{L}}$, we establish a link between local graph density and its optimization landscape. The key tool for this characterization is the Motzkin–Straus Theorem (Motzkin & Straus, 1965), which relates the clique number of a graph to a quadratic optimization problem over the probability simplex. In their work, (Motzkin & Straus, 1965), established this elegant link between the size of the maximum clique, defined as the cardinality of the maximum complete subgraph $g(\mathcal{G})$, and the global maxima of a quadratic optimization problem on the standard simplex. By relating these global maxima to the functional $\mathrm{H}_{\mathtt{L}_0}(z, \mathrm{S}_2(z))$, we can interpret the graph's local geometry in a useful way that captures the branching of the space and the local expansion or contraction of the metric structure.

Consider a vertex $u \in \mathcal{V}$ where the local geometry satisfies a specific regularity condition: every node in the 2-sphere $\mathrm{S}_2(u)$ shares exactly two midpoints with the central node u, as visualized in Figure 5. Under the $\mathtt{L}_0$ generator and uniform measure, the weak entropic curvature directly reflects the density of the local neighborhood through the following structural characterization,

Theorem A.1 (Motzkin–Straus Type Identity; Proof in Appendix D)**.** *Let u be a node in $\mathcal{G}$ and assume that $\mathrm{B}_2(u)$ satisfies the condition that any $v \in \mathrm{S}_2(u)$ shares exactly two midpoints with u. Then,*

$$\mathrm{H}_{\mathtt{L}_0}(z, \mathrm{S}_2(u)) = 1 - \frac{1}{g(\mathrm{B}_2(u))}. \tag{5}$$

This identity reveals that the magnitude of the local functional $\mathrm{H}_{\mathtt{L}_0}$ serves as a signature of local redundancy and clustering. As the maximum clique size within the 2-hop neighborhood increases, the optimization value reflects a geometry that supports highly efficient, isotropic diffusion. Geometrically, these high-density regions act as *local traps* or dense communities where probability mass spreads uniformly. Conversely, as the clique number decreases, approaching tree-like regime, the functional value shifts, signaling the presence of structural bottlenecks where information flow is constrained to fewer, more fragile paths.

B EXPLICIT COMPUTATION OF WEAK ENTROPIC CURVATURE

B.1 BRIDGING LOCAL COMPUTATION AND GLOBAL TOPOLOGY

A key advantage of our framework is its ability to control global graph properties through local optimization. While Entropic Curvature is defined by the convexity of entropy along global Wasserstein geodesics, we derive a tractable local proxy, κ_w, computed strictly within 2-hop neighborhoods. This proxy acts as a lower bound for the global curvature. Consequently, by locally rewiring edges to maximize κ_w, we rigorously improve the global spectral gap and alleviate the bottleneck phenomena without incurring the prohibitive cost of global spectral analysis.

B.2 CANONICAL MOTIFS AND ENTROPIC CURVATURE REGIMES

Figure 3 visualizes the local weak entropic curvature κ_w across several canonical graph motifs. Dense or highly redundant neighborhoods exhibit larger curvature because two-hop mass transport can be spread across multiple intermediate nodes. In contrast, stars, trees, and bottleneck-like structures have lower or negative curvature, reflecting the limited number of available geodesic midpoints and the resulting concentration of transport mass. These examples provide concrete intuition for the role of κ_w as a proxy for local transport redundancy.

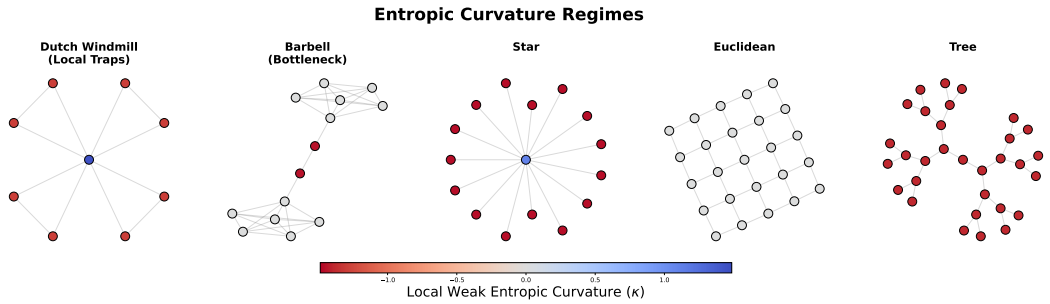

Figure 3: Local weak entropic curvature across various graph motifs.

B.3 THEORETICAL CALCULATION OF SOME EXAMPLES OF WEAK ENTROPIC CURVATURE

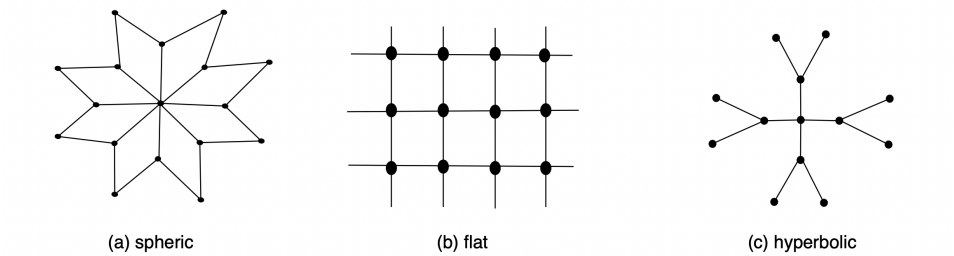

Figure 4: **Characterization of discrete geometric regimes via Weak Entropic Curvature (κ_w) under the uniform generator L_0.** (a) *Spherical ($\kappa_w > 0$):* High local clustering results in positive curvature, satisfying the lower bound $\kappa_w(z) \geq -2\log(1 - \frac{1}{\deg(z)})$. (b) *Flat ($\kappa_w = 0$):* The Euclidean grid exhibits zero curvature, corresponding to a local functional value $H_{L_0}(z, S_2(z)) = 1$. (c) *Hyperbolic ($\kappa_w < 0$):* Tree-like structures characteristic of exponential expansion yield strictly negative curvature, specifically $\kappa_w(z) = -2\log 2 \approx -1.386$.

In this appendix, we provide a theoretical computation of the weak entropic curvature in some specific configurations. Throughout, we work with the L_0 generator (see equation 1).

Spherical configuration: Case (a) in figure 4.

Let z denote the center. We consider

$$\mathrm{H}_{\mathrm{L}_0}(z, \mathrm{S}_2(z)) := \sup_{\alpha} \left\{ \sum_{z'' \in \mathrm{S}_2(z)} \#]z, z''[\prod_{z' \in]z, z''[} (\alpha(z'))^{\frac{2}{\#]z, z''[}} \right\} \tag{6}$$

such that $\sum_{z' \in \mathrm{S}_1(z)} \alpha(z') = 1$. In this geometry we have $\#]z, z''[= 2$ for every $z'' \in \mathrm{S}_2(z)$. Hence, $\sum_{z'' \in \mathrm{S}_2(z)} \#]z, z''[\prod_{z' \in]z, z''[} (\alpha(z'))^{\frac{2}{\#]z, z''[}} \leq (\sum_{z' \in \mathrm{S}_1(z)} \alpha(z'))^2 - \sum_{z' \in \mathrm{S}_1(z)} \alpha(z')^2 \leq \left(1 - \frac{1}{deg(z)}\right)$.

We therefore obtain in this case the following lower bound for the weak entropic curvature, $\kappa_w(z) \geq -2 \log \left(1 - \frac{1}{\deg(z)}\right)$.

Observe that this quantity becomes positive once $\deg(z)$ is sufficiently large ($deg(z) \geq 2$).

Flat configuration: Case (b) in figure 4.

Let z be an arbitrary node. In this configuration, the quantity $\#]z, z''[$ takes values either 1 or 2. Enumerate the vertices of $\mathrm{S}_1(z)$ as $z_1, z_2, \ldots, z_N$, and set $z_{N+1} = z_1$ for convenience.

Repeating the same computation as before, the expression inside the braces in equation 6 becomes

$$\sum_i \alpha(z_i)^2 + 2 \sum_i \alpha(z_i)\alpha(z_{i+1}) \leq \left(\sum_i \alpha(z_i) \right)^2 = 1.$$

Moreover, choosing $\alpha(z_i) = 1$ for some fixed i and $\alpha(z_j) = 0$ for every $j \neq i$, gives

$$\sum_i \alpha(z_i)^2 + 2 \sum_i \alpha(z_i)\alpha(z_{i+1}) = 1.$$

Hence, in this case we obtain $\mathrm{H}_{\mathrm{L}_0}(z, \mathrm{S}_2(z)) = 1$, which implies that $\kappa_w(z) = 0$.

Hyperbolic Configuration : Case (c) in figure 4.

Let z be the center of the figure. Repeating the same calculation we get the expression inside the braces in equation 6 becomes $2 \sum_{z' \in \mathrm{S}_1(z)} \alpha(z')^2$. The supremum of this expression under the condition $\sum_{z' \in \mathrm{S}_1(z)} \alpha(z') = 1$ is 2 which implies $\kappa_w(z) = -2 \log(2) < 0$.

C Spectral Gap Enhancement via Curvature-Based Rewiring

C.1 Motivation

The spectral gap of the graph Laplacian, denoted as λ_2 (also known as the algebraic connectivity, is a fundamental metric in spectral graph theory (Chung, 1997; Friedland & Schneider, 1999). It governs the rate of convergence of diffusion processes, the robustness of the network to bisection, and the expansion properties of the graph. In many real-world applications, such as synchronization in power grids or information flow in communication networks, maximizing λ_2 is desirable (Pecora & Carroll, 1998).

Traditional edge addition strategies often rely on global metrics or random heuristics. However, recent advances in discrete geometry suggest that graph curvature, specifically *bottleneck* regions characterized by negative curvature, dictates structural vulnerabilities (Topping et al., 2022). This experiment aims to demonstrate that a curvature-based rewiring strategy, i.e., targeting nodes with the lowest (most negative) curvature to be connected to the most distant nodes in the network, provides a more efficient path to spectral gap enhancement than traditional discrete curvatures across various synthetic graph topologies.

C.2 Experimental Setup

We evaluate the performance of our proposed Weak Entropic Curvature against two established discrete Ricci curvatures: Ollivier-Ricci Curvature (ORC) and Forman-Ricci Curvature (FRC). The experimental setup is as follows,

1. **Graph Generation:** We initialize six distinct random graph models to ensure topological diversity,
   - **Cayley:** Dihedral group-based graphs representing highly symmetric structures (Babai, 1979).
   - **SBM:** Stochastic Block Models representing community-structured networks (Holland et al., 1983).
   - **Small-World:** Watts-Strogatz graphs characterized by high clustering and short path lengths (Wolfe, 2024).
   - **RGG:** Random Geometric Graphs where connectivity is distance-dependent (Penrose, 2003).
   - **Barabási-Albert:** Scale-free networks with preferential attachment (Albert & Barabási, 2002).
   - **Config:** Configuration model graphs with fixed degree sequences (Fosdick et al., 2018).
2. **Iterative Rewiring:** For each model, we perform 10 iterative edge additions. In each step,
   - The curvature of all nodes is computed using the three respective metrics.
   - The node u with the minimum curvature value is selected.
   - An edge is added between u and the node v that is furthest from it in terms of shortest-path distance.
3. **Statistical Rigor:** Each experiment is repeated over 10 different random seeds. We report the mean change in spectral gap, $\Delta\lambda_2 = \lambda_2^{(t)} - \lambda_2^{(0)}$, where t is the number of added edges. Shaded regions in the figures represent the standard deviation across seeds.

C.3 Results and Discussion

The results of the edge addition experiment are illustrated in Figure 1 (main paper). Across all six graph topologies, the Entropic Curvature (Ours) strategy consistently outperforms both Ollivier and Forman Ricci curvatures in maximizing the spectral gap.

- **Symmetry and Structure (Cayley & SBM):** In the Cayley and SBM models, Entropic curvature shows a significantly steeper trajectory. In the SBM case specifically, Entropic curvature bridges communities more effectively, reaching a $\Delta\lambda_2 \approx 0.6$ after 10 edges, while Forman and Ollivier remain below 0.4.
- **Expansion in Scale-Free and Small-World Graphs:** For Barabási-Albert and Small-World graphs, our metric demonstrates superior sensitivity to the bottlenecks that limit algebraic connectivity. While Forman and Ollivier often track closely together, Entropic curvature identifies superior edges that lead to faster global integration.
- **Robustness (RGG & Config):** Even in Random Geometric Graphs, where spatial constraints often limit rewiring efficacy, our method maintains a lead. The Entropic curvature's ability to optimize a local distribution of weights allows it to identify critical bridge nodes that coarser metrics like Forman (based primarily on degrees) or Ollivier (based on 1-hop neighborhoods) may overlook.

C.4 Midpoint-Completion Rewiring (MCR) Algorithm

The Midpoint-Completion Rewiring strategy is motivated by the geometric interpretation of weak entropic curvature: $\kappa_w(z)$ measures how efficiently probability mass can be displaced from z to its 2-sphere $\mathrm{S}_2(z)$ via two-hop paths. A bottleneck node, one with highly negative $\kappa_w(z)$, has few midpoints connecting it to its 2-hop neighbours, forcing probability mass through a small number of fragile paths. Adding edges that create new two-hop paths enriches the local transport geometry, increases $|\mathrm{M}_{\mathcal{G}}(z, w)|$, and thereby raises curvature, as certified by Theorem 4.3.

Algorithm. The MCR procedure is summarised in Algorithm 1. At each step, we select the node with the most negative κ_w value, identify the two-hop target with the fewest existing midpoints, and connect it to the most geographically distant vertex in $\mathcal{G}$, a heuristic that maximises the spectral gap increase by bridging opposite ends of the bottleneck.

Algorithm 1 Midpoint-Completion Rewiring (MCR)

Require: Graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, generator $\mathtt{L}$, edge budget $T \in \mathbb{N}$
Ensure: Rewired graph $\mathcal{G}_T$
for $t = 1, \ldots, T$ **do**
1. Compute $\kappa_w(z)$ for all $z \in \mathcal{V}$ ▷ Algorithm 2
2. $z^\star \leftarrow \arg\min_{z \in \mathcal{V}} \kappa_w(z)$ ▷ Most negative bottleneck node
3. $w^\star \leftarrow \arg\min_{w \in \mathrm{S}_2(z^\star)} |\mathrm{M}_{\mathcal{G}}(z^\star, w)|$ ▷ Two-hop target with fewest midpoints
4. $v^\star \leftarrow \arg\max_{v \in \mathcal{V}} d(z^\star, v)$ ▷ Most distant vertex (BFS)
5. $\mathcal{E} \leftarrow \mathcal{E} \cup \{(z^\star, v^\star)\}$; update $\mathcal{G}$, recompute $\mathrm{S}_2(z^\star)$

end for
return $\mathcal{G}_T$

Connection to Theorem 4.3. Each edge addition in Algorithm 1 is guaranteed by Theorem 4.3 to be non-decreasing in $\kappa_w(z^\star)$: by connecting $z^\star$ to $v^\star$, we create at least one new shared midpoint between $z^\star$ and nodes in $\mathrm{S}_2(z^\star)$ that lie on the new two-hop path $z^\star \to v^\star \to w$, strictly increasing the pairwise curvature contribution along those channels. Moreover, each such step simultaneously improves spectral expansion (Theorem 4.2) and tightens the generalization bound (Theorem 5.1).

Why target the most distant vertex? The choice $v^\star = \arg\max_v d(z^\star, v)$ is a spectral heuristic: edges bridging topologically distant nodes tend to create shortcuts that most rapidly increase the algebraic connectivity λ_2. Intuitively, the spectral gap is governed by the graph's diameter and bottleneck structure; bridging the two furthest ends of a bottleneck node's neighbourhood maximises the reduction in effective resistance (Spielman, 2010), which is monotonically related to λ_2 for sparse graphs.

Complexity. Each iteration of Algorithm 1 requires: (i) computing κ_w for all nodes, $O(n \cdot k_{\mathrm{iter}} \cdot d_{\max}^2)$ (see Appendix Q); (ii) finding $z^\star$, $O(n)$; (iii) finding $w^\star$ by scanning $\mathrm{S}_2(z^\star)$, $O(d_{\max}^2)$; (iv) finding $v^\star$ by BFS, $O(n + |\mathcal{E}|)$. The dominant cost per

D Geometric Characterization and Proof of Theorem A.1

D.1 Geometric Interpretation via the Motzkin-Straus Theorem

We leverage the Motzkin-Straus Theorem to link local graph density to the optimization landscape of $\mathrm{H}_{\mathtt{L}}$. Consider a node u where every 2-hop neighbor shares exactly two midpoints with u, as illustrated in Figure 5.

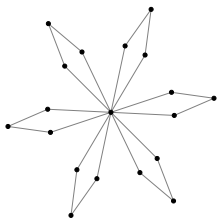

Figure 5: A positively curved motif where the entropic functional serves as a signature of local density and clustering.

D.2 Proof of Theorem A.1

In our set-up of theorem A.1, we have, by definition of $\mathrm{H}_{\mathtt{L}_0}(z, \mathrm{S}_2(u))$,

$$\mathrm{H}_{\mathrm{L}_0}(z, \mathrm{S}_2(u)) = 2 \max_{\alpha} \left\{ \sum_{z' \in \mathrm{S}_2(z))} \prod_{z'' \in]z, z'[} \alpha(z'') \right\} = 2 \max_{\alpha} \left\{ \sum_{\{i,j\} \in \mathcal{E}_z} \alpha_i \alpha_j \right\}$$

such that $\{z', w'\}$ is an edge of $\mathcal{E}_z$ if $\{z', w'\} =]z, z'[$ for some $z' \in \mathrm{S}_2(z)$.

Thus, Theorem A.1 is a direct consequence of the following theorem.

Theorem D.1 (Motzkin-Straus Theorem Motzkin & Straus (1965))**.** *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a simple undirected graph with clique number $\omega(\mathcal{G})$. Then the following relation holds:*

$$2 \max_{\alpha} \left\{ \sum_{\{i,j\} \in \mathcal{E}} \alpha_i \alpha_j \right\} = 1 - \frac{1}{\omega(\mathcal{G})}.$$

E Proof of Theorem 3.3

Proof of Theorem 3.3. From Theorem 2 in Rapaport & Samson (2024), we have $\kappa \geq \inf_S -2\log(\mathrm{K}(\mathrm{S}))$ where K is defined in (15) in Rapaport & Samson (2024) and $\mathrm{S} \subset \mathcal{X} \times \mathcal{X}$ is d-cyclically monotone i.e. For every finite collection of points $(x_1, y_1), \ldots, (x_N, y_N)$ in S, the following inequality holds:

$$\sum_{i=1}^{N} d(x_i, y_i) \leq \sum_{i=1}^{N} d(x_i, y_{i+1}),$$

where $y_{N+1} \equiv y_1$ by convention.

However, from (16) in Rapaport & Samson (2024), we have

$$\sup_S \mathrm{K}(\mathrm{S}) \leq \sup_{z \in \mathcal{V}} \mathrm{H}_{\mathrm{L}}(z, \mathrm{S}_2(z)),$$

which means that $\inf_{\mathrm{S}} -2\log(\mathrm{K}(\mathrm{S})) \geq \kappa_w$. This ends the proof. □

F Proof of Theorem 5.1

Remark F.1 (Global Lipschitz constant and non-vacuousness). The Lipschitz assumption is global with respect to graph distance, not merely local over adjacent nodes. For a K-layer GCN with weight matrices $\mathbf{W}^{(1)}, \ldots, \mathbf{W}^{(K)}$ and L_σ-Lipschitz activations, one obtains on bounded-degree graphs $\Lambda \leq L_\sigma^K \prod_{k=1}^{K} \|\mathbf{W}^{(k)}\|_{\mathrm{op}}$. Thus, Λ is finite for fixed architectures with bounded weights, and can be controlled in practice through spectral normalization or weight regularization (Abbahaddou et al.; Ennadir et al., 2024).

First, we will use the following lemma, necessarily for proving our generalization bound.

Lemma F.2. *The graph $(\mathcal{V}, d, \mathrm{m}, \mathtt{L})$ having weak entropic curvature $\kappa_w > 0$ implies a* transport-entropy inequality *of the form,*

$$\mathrm{W}_1(\nu, \mathtt{m})^2 \ \leq \ \frac{2}{\kappa_w} \mathrm{Ent}_{\mathtt{m}}(\nu). \tag{7}$$

Proof. Indeed from equation 1 by replacing μ_0 by $\mathtt{m}$ and μ_1 by ν we get

$$0 \leqslant \mathrm{Ent}_{\mathtt{m}}(\nu_t) \leq (1-t)\,\mathrm{Ent}_{\mathtt{m}}(\mathtt{m}) + t\,\mathrm{Ent}_{\mathtt{m}}(\nu) \tag{8}$$

$$- \frac{\kappa_w}{2} t(1-t)\mathrm{W}_1(\mathtt{m}, \nu)^2. \tag{9}$$

Since $\mathrm{Ent}_{\mathtt{m}}(\mathtt{m}) = 0$ and by simplifying by t we get $(1-t)\mathrm{W}_1(\nu, \mathtt{m})^2 \ \leq \ \frac{2}{\kappa_w} \mathrm{Ent}_{\mathtt{m}}(\nu)$. By taking $t = 0$ we get equation 7. □

This inequality constrains how far, in Wasserstein distance W_1, any distribution ν can stray without incurring large relative entropy.

Now, we will use Lemma F.2 and prove Theorem 5.1.

We first express

$$\Big|\mathcal{L}_{\mu_1}(f_\theta) \;-\; \mathcal{L}_{\mu_0}(f_\theta)\Big| \;=\; \Big|\mathbb{E}_{x\sim\mu_1}[\ell(f_\theta(x))] \;-\; \mathbb{E}_{y\sim\mu_0}[\ell(f_\theta(y))]\Big|.$$

By the definition of the 1-Wasserstein distance, let $\Gamma^* \in \Pi(\mu_1, \mu_0)$ be an optimal coupling between μ_1 and μ_0. Then

$$\big|\mathcal{L}_{\mu_1}(f_\theta) - \mathcal{L}_{\mu_0}(f_\theta)\big| \;\le\; \int_{\mathcal{V}\times\mathcal{V}} \big|\ell(f_\theta(x)) - \ell(f_\theta(y))\big|\,\mathrm{d}\Gamma^*(x,y).$$

By the Lipschitz property of ℓ, specifically σ - Lipschitz in its first argument, we get

$$\big|\ell(f_\theta(x)) - \ell(f_\theta(y))\big| \;\le\; \sigma\,\big\|\,f_\theta(x) - f_\theta(y)\big\|.$$

Lipschitz of the GNN's Embedding or Output

Assume the GNN output $\|f_\theta(x) - f_\theta(y)\|$ is at most $\Lambda \cdot d(x,y)$ for neighbors $x, y \in \mathcal{V}$, or more generally in the sense of a local Lipschitz constant. Then

$$\|f_\theta(x) - f_\theta(y)\| \;\le\; \Lambda\, d(x,y).$$

Hence,

$$\big|\ell(f_\theta(x)) - \ell(f_\theta(y))\big| \;\le\; \sigma\,\Lambda\, d(x,y).$$

Substituting into the integral and recognizing the infimum over all couplings Γ^*,

$$\big|\mathcal{L}_{\mu_1}(f_\theta) - \mathcal{L}_{\mu_0}(f_\theta)\big| \;\le\; \sigma\,\Lambda \int_{\mathcal{V}\times\mathcal{V}} d(x,y)\,\mathrm{d}\Gamma^*(x,y). \;=\; \sigma\,\Lambda\,\mathrm{W}_1(\mu_1,\mu_0).$$

Transport-Entropy Inequality from Positive Curvature

Because the graph satisfies κ-bounded entropic curvature, there is a classical *transport-entropy inequality* of the form

$$\mathrm{W}_1(\mu_1,\mu_0)^2 \;\;\le\;\; \frac{2}{\kappa}\,\mathrm{Ent}_{\mu_0}(\mu_1).$$

Therefore

$$\mathrm{W}_1(\mu_1,\mu_0) \;\;\le\;\; \sqrt{\frac{2}{\kappa}\,\mathrm{Ent}_{\mu_0}(\mu_1)}.$$

Combining with the Lipschitz bound above yields the final inequality:

$$\Big|\mathcal{L}_{\mu_1}(f_\theta) \;-\; \mathcal{L}_{\mu_0}(f_\theta)\Big| \;\;\le\;\; \sigma\,\Lambda\,\sqrt{\frac{2}{\kappa}\,\mathrm{Ent}_{\mu_0}(\mu_1)}.$$

Thus, the GNN's loss can only degrade significantly if the adversarially modified distribution μ_1 is far from μ_0 in Wasserstein distance, a distance that is itself controlled by the entropic curvature κ.

G Proof of Theorem 5.2

Proof. The proof proceeds in two main phases: first, bounding the generalization gap by the optimal transport distance $\mathrm{W}_1(\mu_{\text{test}}, \mu_{\text{train}})$, and second, establishing a Transport-Entropy inequality derived from the negative curvature and diameter constraints using the Benamou-Brenier formalism Benamou & Brenier (2000). We denote ehe expected generalization gap as,

$$\Delta\mathcal{L} = |\mathcal{L}_{\mu_{\text{test}}}(f_\theta) - \mathcal{L}_{\mu_{\text{train}}}(f_\theta)|.$$

Let $\Gamma^* \in \Pi(\mu_{\text{test}}, \mu_{\text{train}})$ be the optimal transport coupling between the two distributions. We rewrite the difference as an integral over the joint distribution:

$$\Delta\mathcal{L} \le \int_{\mathcal{V}\times\mathcal{V}} |\mathcal{J}(f_\theta(u), y_u) - \mathcal{J}(f_\theta(v), y_v)|\,\mathrm{d}\Gamma^*(u,v).$$

Given that the loss is L-Lipschitz and the GNN representation is Λ-Lipschitz with respect to the graph distance d, the composition is $(L \cdot \Lambda)$-Lipschitz. Applying this yields,

$$\Delta\mathcal{L} \leq \mathrm{L} \cdot \Lambda \int_{\mathcal{V}\times\mathcal{V}} d(u,v)\, \mathrm{d}\Gamma^*(u,v) = \mathrm{L} \cdot \Lambda \cdot \mathrm{W}_1(\mu_{\text{test}}, \mu_{\text{train}}). \tag{10}$$

To bound $\mathrm{W}_1(\mu_{\text{test}}, \mu_{\text{train}})$, we analyze the entropy along a geodesic. Let $(\mu_t)_{t\in[0,1]}$ be a constant-speed optimal transport geodesic connecting $\mu_0 = \mu_{\text{train}}$ to $\mu_1 = \mu_{\text{test}}$. Let $u(t) = Ent_{\mu_{\text{train}}}(\mu_t)$.

By the assumption of weak entropic curvature $\kappa_w \geq -\mathrm{K}$, the entropy functional satisfies the displacement convexity inequality,

$$u(t) \leq (1-t)u(0) + tu(1) + \frac{\mathrm{K}}{2}t(1-t)\mathrm{W}_1(\mu_{\text{train}}, \mu_{\text{test}})^2.$$

This integral inequality implies a lower bound on the second derivative (the curvature "deficit"),

$$u''(t) \geq -\mathrm{K} \cdot \mathrm{W}_1(\mu_{\text{train}}, \mu_{\text{test}})^2. \tag{11}$$

Next, we relate the entropy at the endpoints using a Taylor expansion with integral remainder centered at $t = 1$,

$$u(0) = u(1) - u'(1) + \int_0^1 s \cdot u''(s)\, ds.$$

Substituting the curvature bound equation 11 into the integral gives,

$$u(0) \geq u(1) - u'(1) - \mathrm{K} \cdot \mathrm{W}_1^2 \int_0^1 s\, ds = u(1) - u'(1) - \frac{K}{2}\mathrm{W}_1^2.$$

Since $u(0) = \mathrm{Ent}_{\mu_{\text{train}}}(\mu_{\text{train}}) = 0$ and $u(1) = \mathrm{Ent}_{\mu_{\text{train}}}(\mu_{\text{test}})$, rearranging yields,

$$\mathrm{Ent}_{\mu_{\text{train}}}(\mu_{\text{test}}) \leq u'(1) + \frac{\mathrm{K}}{2}\mathrm{W}_1^2. \tag{12}$$

To bound the derivative $u'(1)$, we employ the Benamou-Brenier continuity equation: $\frac{\partial \rho_t}{\partial t} + \nabla \cdot (\rho_t v_t) = 0$, where ρ_t is the density and v_t is the optimal velocity field. Differentiating the entropy:

$$\begin{aligned} u'(t) &= \int_{\mathcal{V}} (1 + \log \rho_t) \frac{\partial \rho_t}{\partial t}\, d\mu_{\text{train}} = -\int_{\mathcal{V}} (1 + \log \rho_t) \nabla \cdot (\rho_t v_t)\, d\mu_{\text{train}} \\ &= \int_{\mathcal{V}} \nabla(\log \rho_t) \cdot (\rho_t v_t)\, d\mu_{\text{train}} = \int_{\mathcal{V}} \langle \nabla \log \rho_t, v_t \rangle\, d\mu_t. \end{aligned}$$

Evaluating at $t = 1$ and applying the Cauchy-Schwarz inequality:

$$u'(1) \leq \sqrt{\int_{\mathcal{V}} |\nabla \log \rho_1|^2\, d\mu_{\text{test}}} \cdot \sqrt{\int_{\mathcal{V}} |v_1|^2\, d\mu_{\text{test}}} = \sqrt{I_{\mu_{\text{train}}}(\mu_{\text{test}})} \cdot W_2(\mu_{\text{train}}, \mu_{\text{test}}),$$

where I is the Fisher Information. Since $W_1 \leq W_2$, we substitute this back into equation 12 to obtain the HWI inequality:

$$\mathrm{Ent}_{\mu_{\text{train}}}(\mu_{\text{test}}) \leq \sqrt{I_{\mu_{\text{train}}}(\mu_{\text{test}})} \cdot \mathrm{W}_1 + \frac{\mathrm{K}}{2}\mathrm{W}_1^2.$$

Because negative curvature allows distances to grow indefinitely, we impose the graph diameter constraint $\mathrm{W}_1 \leq D$. Standard metric-measure theory establishes that a space with curvature bounded by $-\mathrm{K}$ and diameter D satisfies a Logarithmic Sobolev Inequality (LSI) with constant $\lambda \approx \frac{\mathrm{K}}{2(e^{\mathrm{K}D^2/8}-1)}$ such that $\mathrm{Ent} \leq \frac{1}{2\lambda} I$.

By the Otto-Villani theorem, satisfying LSI with constant λ implies the Transport-Entropy inequality $W_1^2 \leq \frac{2}{\lambda} Ent$. Substituting our explicit λ:

$$\mathrm{W}_1(\mu_{\text{test}}, \mu_{\text{train}})^2 \leq \frac{4}{\mathrm{K}} \left(e^{\frac{\mathrm{K}D^2}{8}} - 1 \right) \mathrm{Ent}_{\mu_{\text{train}}}(\mu_{\text{test}}). \tag{13}$$

Finally, substituting the Wasserstein bound equation 13 into our generalization gap inequality equation 10 yields the desired result:

$$|\mathcal{L}_{\mu_{\text{test}}}(f_\theta) - \mathcal{L}_{\mu_{\text{train}}}(f_\theta)| \leq \mathrm{L} \cdot \Lambda \sqrt{\frac{4}{\mathrm{K}} \left(e^{\frac{\mathrm{KD}^2}{8}} - 1 \right) \mathrm{Ent}_{\mu_{\text{train}}}(\mu_{\text{test}})}.$$

This establishes that under negative curvature, adversarial robustness degrades exponentially with the graph diameter, highlighting a fundamental trade-off between mitigating oversmoothing and maintaining model stability. □

H Datasets Statistics & Implementation Details

Our implementation is available in the supplementary materials (and will be publicly available afterwards). It is built using the open-source library *PyTorch Geometric* (PyG) under the MIT license (Fey & Lenssen, 2019). Note that we additionally utilized the PyTorch DeepRobust package[1] to implement the adversarial attacks used in this study. The experiments have been run on both a NVIDIA A100 GPU and a RTX A6000 GPU.

H.1 Benchmark Datasets

We evaluate our framework on six widely recognized graph benchmarks that represent distinct structural and homophilic regimes Yang et al. (2016); Pei et al. (2020). These are categorized as follows:

- **Homophilic Datasets:** *Cora*, *CiteSeer*, and *PubMed* are citation networks where nodes represent documents and edges represent citations. In these graphs, nodes with the same label (research topic) are highly likely to be connected, a property that generally corresponds to more regular local geometries.
- **Heterophilic Datasets:** *Cornell*, *Texas*, and *Wisconsin* are webpage networks from computer science departments. These graphs exhibit high heterophily, meaning connected nodes often belong to different classes. Such topologies typically contain more structural bottlenecks and negative curvature regions, providing a rigorous test for our E-Gate mechanism.

The statistics for these datasets are summarized in Table 3.

Table 3: Dataset Statistics

Dataset	Nodes	Edges	Features	Classes	Homophily
Cora	2,708	5,429	1,433	7	0.81
CiteSeer	3,327	4,732	3,703	6	0.74
PubMed	19,717	44,338	500	3	0.80
Cornell	183	295	1,703	5	0.30
Texas	183	309	1,703	5	0.11
Wisconsin	251	499	1,703	5	0.21

H.2 Experimental Protocol and Data Splits

For all datasets, we adopt a random splitting strategy to evaluate the model's generalization under distributional variance. For each of the 10 independent runs, we randomly sample 60% of the nodes for training, 20% for validation (to monitor convergence), and 20% for testing. This protocol ensures that the reported mean accuracy and standard deviation reflect the model's robustness to specific topological samplings.

[1] https://github.com/DSE-MSU/DeepRobust

H.3 Implementation Details

We utilize a two-layer GNN, e.g. GCN or GIN etc., backbone with 64 hidden units. The model is trained for 200 epochs using the Adam optimizer with a learning rate of 0.01 and weight decay of 5×10^{-4}. For the E-Gate mechanism, the temperature parameter τ is initialized at 1.0 and learned during training via backpropagation. To ensure statistical robustness, we report the mean and standard deviation across 10 independent runs using distinct random splits for each dataset. All experiments were implemented using PyTorch Geometric and executed on NVIDIA RTX 3090 GPUs.

I Proof of Theorem 4.2

The proof of this theorem adapts techniques from the work in Salez (2022).

I.1 Random Rooted Graphs

All graphs considered here are simple, undirected, countable, and locally finite. A *rooted graph* is a pair $(\mathcal{G}, o)$, where $\mathcal{G}$ is a graph and o is a distinguished vertex called the *root*. Two rooted graphs $(\mathcal{G}, o)$ and $(\mathcal{G}', o')$ are *isomorphic*, denoted $\mathcal{G} \simeq \mathcal{G}'$, if there exists a bijection $\phi\colon \mathcal{V}_{\mathcal{G}} \to \mathcal{V}_{\mathcal{G}'}$ that preserves the root ($\phi(o) = o'$) and the edges:

$$\forall x, y \in \mathcal{V}_{\mathcal{G}}, \quad \{x, y\} \in \mathcal{E}_{\mathcal{G}} \iff \{\phi(x), \phi(y)\} \in \mathcal{E}_{\mathcal{G}'}.$$

Let $\mathcal{G}_\bullet$ denote the set of connected rooted graphs, considered up to the isomorphism relation $\simeq$. For brevity, we use $(\mathcal{G}, o)$ to denote both the rooted graph and its equivalence class. We write $B_t(\mathcal{G}, o)$ for the *ball of radius t around the root* in $\mathcal{G}$, i.e., the finite rooted subgraph induced by the vertices $\{x \in \mathcal{V}_{\mathcal{G}} : d_{\mathcal{G}}(o, x) \le t\}$.

The space $\mathcal{G}_\bullet$ is equipped with the *local metric* $d_{\mathrm{LOG}}\colon \mathcal{G}_\bullet \times \mathcal{G}_\bullet \to [0, 1]$, defined as:

$$d_{\mathrm{LOG}}\big((\mathcal{G}, o), (\mathcal{G}', o')\big) := \frac{1}{1 + r}, \quad \text{where} \quad r = \sup\{t \ge 0 : B_t(\mathcal{G}, o) \simeq B_t(\mathcal{G}', o')\}.$$

Intuitively, two rooted graphs are *close* if their local structures agree up to a large radius r. The metric space $(\mathcal{G}_\bullet, d_{\mathrm{LOG}})$ is complete and separable. We equip it with its Borel σ-algebra, and any $\mathcal{G}_\bullet$-valued random variable is called a *random rooted graph*.

Let $\mathcal{P}(\mathcal{G}_\bullet)$ denote the space of Borel probability measures on $\mathcal{G}_\bullet$, equipped with the topology of weak convergence. For a finite graph $\mathcal{G}$, its *local profile* $\mathcal{S}_{\mathcal{G}} \in \mathcal{P}(\mathcal{G}_\bullet)$ is the empirical distribution of all possible rootings of $\mathcal{G}$:

$$\mathcal{S}_{\mathcal{G}} := \frac{1}{|\mathcal{V}_{\mathcal{G}}|} \sum_{x \in \mathcal{V}_{\mathcal{G}}} \delta_{(\mathcal{G}, x)},$$

where $\delta_{(\mathcal{G},x)}$ is the Dirac measure at $(\mathcal{G}, x)$.

I.2 Local Weak Limit of Graph Sequences

For a finite graph $\mathcal{G}$, the pair $(\mathcal{G}, x)$ is implicitly restricted to the connected component of x if $\mathcal{G}$ is not connected. Given a sequence of finite graphs $\mathcal{G}_n = (\mathcal{V}_n, \mathcal{E}_n)$ for $n \ge 1$, if their local profiles $(\mathcal{S}_{\mathcal{G}_n})_{n \ge 1}$ converge to a limit $\mathcal{S}$ in $\mathcal{P}(\mathcal{G}_\bullet)$, we call $\mathcal{S}$ the *local weak limit* of the sequence $(\mathcal{G}_n)_{n \ge 1}$ and denote this convergence by:

$$\mathcal{G}_n \xrightarrow[n \to \infty]{} \mathcal{S}.$$

Intuitively, $\mathcal{S}$ represents the law of a random rooted graph that captures the asymptotic local structure of $\mathcal{G}_n$ when viewed from a uniformly random root. Formally, for any continuous and bounded observable $f\colon \mathcal{G}_\bullet \to \mathbb{R}$, the following holds:

$$\frac{1}{|\mathcal{V}_n|} \sum_{x \in \mathcal{V}_n} f(\mathcal{G}_n, x) \xrightarrow[n \to \infty]{} \mathcal{S}[f(\mathcal{G}, o)].$$

The left-hand side is a spatial average of local contributions from the vertices of $\mathcal{G}_n$. Local weak convergence thus simplifies asymptotic analysis by replacing such averages with expectations over a limiting random rooted graph.

Local Observables

The class of continuous functions on $\mathcal{G}_\bullet$ includes, but is not limited to, all t-local observables ($t \geq 0$). Here, a function $f\colon \mathcal{G}_\bullet \to \mathbb{R}$ is called *t-local* if its value $f(\mathcal{G}, o)$ depends only on the isomorphism class of the finite ball $B_t(\mathcal{G}, o)$. Below is a list of key examples that will be used throughout this paper without further mention:

- The *root degree* $(\mathcal{G}, o) \mapsto \deg_{\mathcal{G}}(o)$ is 1-local.
- The *minimum curvature at o*, $(\mathcal{G}, o) \mapsto \kappa_w(\mathcal{G})(o)$, is 2-local.

I.3 Tightness, Unimodularity, and Stationarity

I.3.1 Tightness of Graph Sequences

One of the key advantages of the local weak convergence framework is that every "reasonable" sequence of sparse graphs admits a local weak limit. The following tightness criterion by Benjamini, Lyons, and Schramm makes this precise (note that passing to subsequences may be necessary):

Theorem I.1. *Let $\mathcal{G}_n = (\mathcal{V}_n, \mathcal{E}_n)$, $n \geq 1$, be finite graphs such that*

$$\sup_{n\geq 1} \left\{ \frac{1}{|\mathcal{V}_n|} \sum_{x\in\mathcal{V}_n} \phi(\deg_{\mathcal{G}_n}(x)) \right\} < \infty,$$

where $\phi\colon \mathbb{Z}_+ \to \mathbb{R}_+$ satisfies $\phi(d) \gg d$ as $d \to \infty$. Then, $(\mathcal{G}_n)_{n\geq 1}$ has a subsequence that admits a local weak limit.

Proof. See Theorem 3.1 in Benjamini et al. (2015). □

In particular, this criterion applies to the sequence $(\mathcal{G}_n)_{n\geq 1}$ in Theorem 3 with $\phi(d) = d \log d$. This ensures we can study the existence of non-negatively curved expanders directly at the level of local weak limits.

I.3.2 Unimodularity

Local weak limits of finite graphs inherit a powerful distributional invariance called *unimodularity*, stemming from the uniform randomness of the root in the local profile. Formally, a measure $\mathcal{S} \in \mathcal{P}(\mathcal{G}_\bullet)$ is *unimodular* if it satisfies the *Mass Transport Principle*:

$$\mathcal{S}\left[\sum_{x\in\mathcal{V}_\mathcal{G}} f(\mathcal{G}, o, x)\right] = \mathcal{S}\left[\sum_{x\in\mathcal{V}_\mathcal{G}} f(\mathcal{G}, x, o)\right],$$

for every Borel function $f\colon \mathcal{G}_{\bullet\bullet} \to [0, \infty]$, where $\mathcal{G}_{\bullet\bullet}$ is the space of graphs with two distinguished roots (instead of one).

- **Interpretation**: If $f(\mathcal{G}, o, x)$ represents "mass" sent from o to x, (6) states that the expected mass sent and received by the root o under $\mathcal{S}$ must balance.
- **Preservation**: This principle holds for the local profile of any finite graph and is preserved under weak convergence. Thus, all local weak limits of finite graphs are unimodular.

I.4 Properties of Local Weak Limits

I.4.1 Unimodularity of Limits

Theorem I.2 (Inherited Unimodularity)**.** *All local weak limits of finite graphs are unimodular.*

I.4.2 CONSEQUENCES OF UNIMODULARITY

Lemma I.3 (Everything Shows at the Root, Lemma 2.3 in Bowen et al. (2024)). *Let $\mathcal{S} \in \mathcal{P}(\mathcal{G}_\bullet)$ be unimodular, and $B \subseteq \mathcal{G}_\bullet$ a Borel set with $\mathcal{S}(B) = 1$. Then:*

$$\mathcal{S}\big(\{(\mathcal{G}, o) : \forall x \in \mathcal{V}_\mathcal{G}, (\mathcal{G}, x) \in B\}\big) = 1.$$

Proof. Apply the Mass Transport Principle with $f(\mathcal{G}, o, x) = \mathbf{1}_{(\mathcal{G},o)\notin B}$. □

I.4.3 STATIONARITY

Definition I.4 (Stationary Measure). A law $\mathcal{S} \in \mathcal{P}(\mathcal{G}_\bullet)$ is *stationary* if it's invariant under the Markov chain that moves the root according to $P_\mathcal{G}$:

$$\mathcal{S}\left[\sum_{x\in\mathcal{V}_\mathcal{G}} P^t_\mathcal{G}(o, x)h(\mathcal{G}, x)\right] = \mathcal{S}[h(\mathcal{G}, o)], \quad \forall t \geq 0,$$

for all Borel $h : \mathcal{G}_\bullet \to [0, \infty]$.

Lemma I.5 (Degree-Biasing Transformation). *For any unimodular $\mathcal{S} \in \mathcal{P}(\mathcal{G}_\bullet)$ with $\deg(\mathcal{S}) := \mathcal{S}[\deg_\mathcal{G}(o)] < \infty$, the measure:*

$$d\hat{\mathcal{S}}(\mathcal{G}, o) := \frac{\deg_\mathcal{G}(o)}{\deg(\mathcal{S})} d\mathcal{S}(\mathcal{G}, o)$$

is stationary.

Proof. Apply the Mass Transport Principle to $\mathcal{S}$ with $f(\mathcal{G}, o, x) = h(\mathcal{G}, o)\mathbf{1}_{\{x,o\}\in\mathcal{E}_\mathcal{G}}$. □

I.5 SPECTRAL RADIUS, ENTROPY, AND THE LIOUVILLE PROPERTY

I.5.1 MUTUAL ABSOLUTE CONTINUITY

Remark I.6 (Mutual Absolute Continuity). From the degree-biasing transformation (8), the original law $\mathcal{S}$ and its biased version $\widehat{\mathcal{S}}$ satisfy:

$$\mathcal{S}(B) = 1 \iff \widehat{\mathcal{S}}(B) = 1,$$

for any Borel set $B \subseteq \mathcal{G}_\bullet$. This equivalence allows transferring results between the two measures.

I.5.2 ASYMPTOTIC ENTROPY

Lemma I.7 (Random Walk Entropy). *Let $\mathcal{S} \in \mathcal{P}(\mathcal{G}_\bullet)$ be stationary with $\mathcal{S}[\log \deg_\mathcal{G}(o)] < \infty$. Then the limit*

$$\mathcal{H}(\mathcal{G}) := \lim_{t\to\infty} \frac{1}{t} \sum_{x\in\mathcal{V}_\mathcal{G}} P^t_\mathcal{G}(o, x) \log \frac{1}{P^t_\mathcal{G}(o, x)}$$

exists $\mathcal{S}$-almost surely and in $L^1(\mathcal{G}_\bullet, \mathcal{S})$, and is root-independent.

Proof. See Lemma 8 in Salez (2022). □

I.6 THE LIOUVILLE PROPERTY

Definition I.8. A function $f : \mathcal{V}_\mathcal{G} \to \mathbb{R}$ is *harmonic* if:

$$P_\mathcal{G} f = f \quad \text{where} \quad (P_\mathcal{G} f)(x) := \sum_{y\in\mathcal{V}_\mathcal{G}} P_\mathcal{G}(x, y)f(y)$$

Definition I.9 (Harmonic Functions). A function $f\colon \mathcal{V}_\mathcal{G} \to \mathbb{R}$ is *harmonic* if it satisfies $P_\mathcal{G} f = f$, where $(P_\mathcal{G} f)(x) := \sum_{y\in\mathcal{V}_\mathcal{G}} P_\mathcal{G}(x, y)f(y)$.

- **Trivial case**: All constant functions are automatically harmonic.
- **Non-trivial case**: The existence of non-constant bounded harmonic functions reflects rich geometric structure.

Definition I.10 (Liouville Property (Colding & Minicozzi II, 2019))**.** A graph G has the *Liouville property* if every bounded harmonic function on G is constant.

Remark I.11 (Entropic Characterization). For stationary random graphs, this functional-analytic property admits a profound probabilistic characterization through asymptotic entropy $\mathcal{H}(\mathcal{G})$. Specifically:

- The Liouville property corresponds to vanishing entropy ($\mathcal{H}(\mathcal{G}) = 0$)
- Non-zero entropy detects the existence of non-constant bounded harmonic functions

This bridges geometric, analytic, and probabilistic aspects of random graphs.

Theorem I.12 (Entropic Characterization of Liouville Property)**.** *For any stationary law $\mathcal{S} \in \mathcal{P}(\mathcal{G}_\bullet)$ with $\mathcal{S}[\log \deg_{\mathcal{G}}(o)] < \infty$, the following equivalence holds $\mathcal{S}$-almost surely:*

$$\mathcal{H}(\mathcal{G}) = 0 \quad \Longleftrightarrow \quad \mathcal{G} \text{ has the Liouville property}.$$

Proof. See theorem 9 in Salez (2022). □

I.7 Spectral Radius and Its Connections

I.7.1 Definition and Basic Properties

The entropy $\mathcal{H}(\mathcal{G})$ relates to several fundamental graph quantities including *speed*, *growth*, and *spectral radius*.

For any rooted graph $(\mathcal{G}, o) \in \mathcal{G}_\bullet$ and $t, s \geq 0$, the inequality

$$P_{\mathcal{G}}^{t+s}(o,o) \geq P_{\mathcal{G}}^{t}(o,o) P_{\mathcal{G}}^{s}(o,o)$$

holds. By Fekete's lemma (Fekete, 1923), the limit

$$\rho(\mathcal{G}) := \lim_{t\to\infty} \left(P_{\mathcal{G}}^{t}(o,o)\right)^{1/t} \tag{13}$$

exists in $(0, 1]$. Furthermore, graph connectivity and the inequality

$$P_{\mathcal{G}}^{t+2s}(o,o) \geq P_{\mathcal{G}}^{s}(o,x) P_{\mathcal{G}}^{t}(x,x) P_{\mathcal{G}}^{s}(x,o)$$

together imply that $\rho(\mathcal{G})$ is independent of the root choice o.

I.7.2 Relation to Entropy

Lemma I.13 (Spectral Radius vs Entropy)**.** *For any stationary law $\mathcal{S}$ with $\mathcal{S}[\log \deg_{\mathcal{G}}(o)] < \infty$, the inequality*

$$\mathcal{H}(\mathcal{G}) \geq 2 \log \frac{1}{\rho(\mathcal{G})}$$

holds $\mathcal{S}$-almost surely.

Proof. See lemma 10 in Salez (2022). □

I.8 Proof of Theorem I.

Let $a > 0$ and $\mathcal{G}_n := (\mathcal{V}_n, \mathcal{E}_n), n \geq 1$ be a sequence of finite graphs such that

$$\sup_{n\geq 1} \left\{ \frac{1}{|\mathcal{V}_n|} \sum_{x\in\mathcal{V}_n} \deg_{\mathcal{G}_n}(x) \log \deg_{\mathcal{G}_n}(x) \right\} < \infty \tag{14}$$

and

$$\forall \varepsilon > 0, \quad \frac{1}{|\mathcal{V}_n|} \# \left\{ x \in \mathcal{V}_n : \kappa_w(x) < a \right\} \xrightarrow[n\to\infty]{} 0. \tag{15}$$

The goal is to prove that

$$\forall \rho < 1, \quad \liminf_{n\to\infty} \left(\frac{1}{|\mathcal{V}_n|} \# \{ i : \lambda_i(\mathcal{G}_n) \geq \rho \} \right) > 0. \tag{16}$$

In view of equation 14 and Theorem I.1, we can, after extracting a subsequence if necessary, assume that

$$\mathcal{G}_n \xrightarrow{n\to\infty} \mathcal{S} \tag{17}$$

For some $\mathcal{S} \in P(\mathcal{G}_\bullet)$, note that $\mathcal{S}$ is automatically unimodular by Theorem I.2, and satisfies

$$\mathcal{S}\left[\deg_{\mathcal{G}}(o) \log \deg_{\mathcal{G}}(o)\right] < \infty. \tag{18}$$

Since curvature is a local property like the degree, it also "passes to the limit," meaning

$$\mathcal{S}\left(\kappa_w(\mathcal{G}) \geq a\right) = 1. \tag{19}$$

Proof. As previously noted, the observable $f\colon (\mathcal{G}, o) \mapsto \kappa_w(o)$ is 2-local, and therefore continuous on $\mathcal{G}_\bullet$. Applying the Portmanteau Theorem yields that

$$\begin{aligned} \mathcal{S}(f < a) &\leq \liminf_{n\to\infty} \mathcal{S}_{\mathcal{G}_n}(f < a) \\ &= \liminf_{n\to\infty} \left\{ \frac{1}{|\mathcal{V}_n|} \#\{ o \in \mathcal{V}_n : f(\mathcal{G}_n, o) < a \} \right\} = 0. \end{aligned}$$

This implies $\mathcal{S}(f < a) = 0$ by equation 15. By applying Lemma I.3 to the event $B = \{f \geq a\}$ we get The desired result. □

The first key step in our proof is to deduce from equation 19 that the entropy under $\mathcal{S}$ is zero. This is established in the following theorem.

Theorem I.14 (positive curvature implies zero-entropy)**.** *Let $a > 0$. Assume that $\mathcal{S}[\log \deg_{\mathcal{G}}(o)] < \infty$, the we have almost-surely under any stationary law $\mathcal{S} \in P(\mathcal{G}_\bullet)$*

$$\kappa_w(\mathcal{G}) \geq a \implies \mathcal{H}(\mathcal{G}) = 0.$$

Proof. Let $f : \mathcal{V} \to \mathbb{R}$ be an harmonic function i.e.,

$$\mathtt{L} f(x) = 0 \quad \forall x \in \mathcal{V}.$$

Assume in addition that f is bounded. We want that f must be constant.

By using Modified Logarithmic Sobolev Inequality (MLSI) (from Theorem 7.1): If $\kappa_w > 0$, then For any f,

$$\mathrm{Var}_\mu(f) \leq \frac{1}{\kappa_w} \int \sum_{y\in\mathcal{V}} \mathtt{L}(x,y)(f(y) - f(x))^2 d\mu(x).$$

We have,

$$\begin{aligned} \int \sum_{y\in\mathcal{V}} \mathtt{L}(x,y)(f(y) - f(x))^2 d\mu(x) &= \sum_{x,y\in\mathcal{V}} \mathtt{L}(x,y)(f(y) - f(x))^2 \mu(x) \\ &= \sum_{x,y\in\mathcal{V}} \mathtt{L}(x,y) f(y)^2 \mu(x) - 2 \sum_{x,y\in\mathcal{V}} \mathtt{L}(x,y) f(y) f(x) \mu(x) \\ &\quad + \sum_{x,y\in\mathcal{V}} \mathtt{L}(x,y) f(x)^2 \mu(x) \end{aligned}$$

Since $\sum_y \mathtt{L}(x,y) = 0$ (rows sum to zero, see 3), we have:

$$\sum_{x,y} f(x)^2 \mathtt{L}(x,y) \mu(x) = \sum_x f(x)^2 \left(\sum_y \mathtt{L}(x,y) \right) \mu(x) = 0.$$

For the last term, since $\mathtt{L}(x,y)\mu(x) = L(y,x)\mu(y)$ (see 3), we have

$$\sum_{x,y} f(y)^2 \mathtt{L}(x,y)\mu(x) = \sum_{x,y} f(y)^2 L(y,x)\mu(y) = \sum_y f(y)^2 \left(\sum_x L(y,x)\right)\mu(y) = 0.$$

Thus

$$\begin{aligned}\int \sum_{y\in\mathcal{V}} \mathtt{L}(x,y)(f(y)-f(x))^2 d\mu(x) &= -2\sum_{x,y\in\mathcal{V}} \mathtt{L}(x,y)f(y)f(x)\mu(x) \\ &= -2\sum_{x\in\mathcal{V}} f(x)\Big(\sum_{y\in\mathcal{V}} \mathtt{L}(x,y)(f(y)-f(x))\Big)\mu(x).\end{aligned}$$

Recall that f is harmonic function, thus $\sum_{y\in\mathcal{V}} \mathtt{L}(x,y)(f(y)-f(x)) = 0$. Then $\int \sum_{y\in\mathcal{V}} \mathtt{L}(x,y)(f(y)-f(x))^2 d\mu(x) = 0$. From this, we deduce

$$\mathrm{Var}_\mu(f) = 0,$$

This implies that f is a constant. Thus $\kappa_w(\mathcal{G})$ implies that G satisfies the Liouville properties and using Theorem I.12 we get the zero-entropy property: $\mathcal{H}(\mathcal{G}) = 0$. □

In view of Remark I.6, this result extends to any unimodular law $\mathcal{S} \in \mathcal{P}(\mathcal{G}_\bullet)$ satisfying

$$\mathcal{S}[\deg_{\mathcal{G}}(o)\log\deg_{\mathcal{G}}(o)] < \infty,$$

and in particular applies to the limit $\mathcal{S}$ in equation 17. Combining this observation with Lemma I.13, we immediately conclude that our local weak limit satisfies

$$\mathcal{S}(\rho(\mathcal{G}) = 1) = 1.$$

Remarkably, this simple condition suffices to establish equation 16. The following theorem formalizes this claim and completes the proof of our main result.

Theorem I.15 (Zero-entropy implies poor spectral expansion)**.** *Let $\{\mathcal{G}_n = (\mathcal{V}_n, \mathcal{E}_n)\}_{n\geq 1}$ be finite graphs with local weak limit $\mathcal{S}$, and suppose that $\mathcal{S}(\rho(\mathcal{G}) = 1) = 1$. Then, for any $\rho < 1$,*

$$\liminf_{n\to\infty} \left\{ \frac{1}{|\mathcal{V}_n|} \#\{i : \lambda_i(\mathcal{G}_n) > \rho\} \right\} > 0.$$

Proof. See Theorem 12 in Salez (2022). □

J Proof of Theorem 7.1

We first establish an auxiliary functional inequality that will be used to derive the desired Poincaré-type estimate. In particular, positive weak entropic curvature implies a modified logarithmic Sobolev inequality for exponential graph signals. The proof of Theorem 7.1 then follows by applying this inequality to a small perturbation $g = \varepsilon(f - \bar{f})$ and passing to the limit as $\varepsilon \to 0$.

Lemma J.1 (Modified logarithmic Sobolev inequality)**.** *Let (V, d, m, L) be a finite connected graph space with weak entropic curvature lower bound $\kappa_w > 0$, and let $\mu := m/m(V)$. Then, for every $g : V \to \mathbb{R}$,*

$$\mathrm{Ent}_\mu(e^g) \leq \frac{1}{2\kappa_w} \int_V |\nabla^+ g|^2(u) e^{g(u)}\, d\mu(u),$$

where

$$|\nabla^+ g|^2(u) := \sup_{v\sim u} [g(u) - g(v)]_+^2$$

and

$$\mathrm{Ent}_\mu(\phi) := \int_V \phi\log\phi\, d\mu - \left(\int_V \phi\, d\mu\right)\log\left(\int_V \phi\, d\mu\right).$$

Proof. It is enough to prove the inequality for functions of the form $\phi = e^g$, where $\phi : V \to (0, \infty)$. The general case follows by approximating ϕ by $\phi + \delta$ and letting $\delta \downarrow 0$.

Set

$$Z_\phi := \int_V \phi \, d\mu$$

and define the probability measure ν_0 by

$$d\nu_0 := \frac{\phi}{Z_\phi} \, d\mu.$$

Let $\nu_1 := \mu$. Since $\mu = m/m(V)$, the densities with respect to m and with respect to μ differ only by the constant factor $m(V)$. Hence entropy differences with respect to m and μ agree up to additive constants, which cancel in the argument below.

By the positive weak entropic curvature assumption, and since κ_w is a lower bound for the global entropic curvature, the entropy functional is κ_w-convex along admissible Wasserstein geodesics. Thus, for any admissible coupling $\pi \in \Pi(\nu_0, \nu_1)$, there exists an interpolating curve $(\nu_t)_{t\in[0,1]}$ from ν_0 to ν_1 such that

$$\mathrm{Ent}_m(\nu_t) \le (1-t)\,\mathrm{Ent}_m(\nu_0) + t\,\mathrm{Ent}_m(\nu_1) - \frac{\kappa_w}{2} t(1-t) W_1(\nu_0, \nu_1)^2.$$

Rearranging gives

$$\mathrm{Ent}_m(\nu_0) \le -\frac{\mathrm{Ent}_m(\nu_t) - \mathrm{Ent}_m(\nu_0)}{t} + \mathrm{Ent}_m(\nu_1) - \frac{\kappa_w}{2}(1-t) W_1(\nu_0, \nu_1)^2.$$

We now let $t \downarrow 0$.

Write $\pi(\cdot \mid u)$ for the conditional distribution of the endpoint given the starting point u. The first variation of the entropy at $t = 0$ along the corresponding graph interpolation satisfies

$$\frac{d}{dt}\,\mathrm{Ent}_m(\nu_t)\Big|_{t=0} \ge -\sum_{u\in V} |\nabla^+ \log\phi|(u) \left(\sum_{w\in V} d(u,w)\pi(w \mid u)\right) \nu_0(u).$$

Here we used that adding a multiplicative constant to ϕ does not change differences of $\log\phi$, and that along each shortest path the entropy variation is controlled by the maximal one-sided decrease of $\log\phi$ over neighbors of u.

By Young's inequality, for every $u \in V$,

$$|\nabla^+ \log\phi|(u) \left(\sum_{w\in V} d(u,w)\pi(w \mid u)\right) \le \frac{1}{2\kappa_w} |\nabla^+ \log\phi|^2(u) + \frac{\kappa_w}{2} \left(\sum_{w\in V} d(u,w)\pi(w \mid u)\right)^2.$$

Therefore,

$$\frac{d}{dt}\,\mathrm{Ent}_m(\nu_t)\Big|_{t=0} \ge -\frac{1}{2\kappa_w} \sum_{u\in V} |\nabla^+ \log\phi|^2(u)\nu_0(u) - \frac{\kappa_w}{2} \sum_{u\in V} \left(\sum_{w\in V} d(u,w)\pi(w \mid u)\right)^2 \nu_0(u).$$

The last term is bounded by the corresponding transport cost between ν_0 and ν_1. Consequently, when this estimate is inserted into the previous curvature inequality and $t \downarrow 0$, the transport-cost terms cancel. We obtain

$$\mathrm{Ent}_m(\nu_0) \le \mathrm{Ent}_m(\nu_1) + \frac{1}{2\kappa_w} \sum_{u\in V} |\nabla^+ \log\phi|^2(u)\nu_0(u).$$

Since $\nu_1 = \mu = m/m(V)$, the change from Ent_m to Ent_μ only contributes the same additive normalization constant to both sides. Hence

$$\mathrm{Ent}_\mu(\nu_0) \le \frac{1}{2\kappa_w} \sum_{u\in V} |\nabla^+ \log\phi|^2(u)\nu_0(u).$$

Using $d\nu_0 = (\phi/Z_\phi)d\mu$ gives

$$\mathrm{Ent}_\mu(\nu_0) \le \frac{1}{2\kappa_w Z_\phi} \int_V |\nabla^+ \log\phi|^2(u)\phi(u)\, d\mu(u).$$

Multiplying both sides by Z_ϕ and using

$$Z_\phi \operatorname{Ent}_\mu(\nu_0) = \operatorname{Ent}_\mu(\phi),$$

we get

$$\operatorname{Ent}_\mu(\phi) \le \frac{1}{2\kappa_w} \int_V |\nabla^+ \log \phi|^2(u)\phi(u)\, d\mu(u).$$

Finally, taking $\phi = e^g$ gives $\log \phi = g$, and therefore

$$\operatorname{Ent}_\mu(e^g) \le \frac{1}{2\kappa_w} \int_V |\nabla^+ g|^2(u) e^{g(u)}\, d\mu(u).$$

This proves the modified logarithmic Sobolev inequality. □

We now prove Theorem 7.1.

Proof. Let $\bar{f} := \int_{\mathcal{V}} f \, \mathrm{d}\mu$ and set $h := f - \bar{f}$. Then $\int_{\mathcal{V}} h\, d\mu = 0$ and

$$\operatorname{Var}_\mu(f) = \int_{\mathcal{V}} h^2\, d\mu.$$

For a graph function $g : \mathcal{V} \to \mathbb{R}$, define the one-sided local slope

$$|\nabla^+ g|^2(u) := \sup_{v \sim u} [g(u) - g(v)]_+^2,$$

where $[a]_+ := \max\{a, 0\}$.

Since the graph space has positive weak entropic curvature $\kappa_w > 0$, the entropy functional is displacement convex with curvature lower bound κ_w. Equivalently, for every $g : V \to \mathbb{R}$, the following modified logarithmic Sobolev inequality holds:

$$\operatorname{Ent}_\mu(e^g) \le \frac{1}{2\kappa_w} \int_{\mathcal{V}} |\nabla^+ g|^2(u) e^{g(u)}\, d\mu(u),$$

where

$$\operatorname{Ent}_\mu(\phi) := \int_{\mathcal{V}} \phi \log \phi \, d\mu - \left(\int_{\mathcal{V}} \phi \, d\mu \right) \log \left(\int_{\mathcal{V}} \phi \, d\mu \right).$$

Apply this inequality to $g = \varepsilon h$ with $\varepsilon > 0$. We obtain

$$\operatorname{Ent}_\mu(e^{\varepsilon h}) \le \frac{1}{2\kappa_w} \int_{\mathcal{V}} |\nabla^+(\varepsilon h)|^2(u) e^{\varepsilon h(u)}\, d\mu(u).$$

We now expand both sides as $\varepsilon \to 0$. Since

$$e^{\varepsilon h} = 1 + \varepsilon h + \frac{\varepsilon^2}{2} h^2 + O(\varepsilon^3),$$

and $\int_{\mathcal{V}} h\, d\mu = 0$, we have

$$\operatorname{Ent}_\mu(e^{\varepsilon h}) = \frac{\varepsilon^2}{2} \int_{\mathcal{V}} h^2\, d\mu + O(\varepsilon^3).$$

Therefore,

$$\operatorname{Ent}_\mu(e^{\varepsilon h}) = \frac{\varepsilon^2}{2} \operatorname{Var}_\mu(f) + \mathcal{O}(\varepsilon^3).$$

On the other hand,

$$|\nabla^+(\varepsilon h)|^2(u) = \varepsilon^2 |\nabla^+ h|^2(u).$$

Since $h = f - \bar{f}$, subtracting the constant $\bar{f}$ does not change edge differences, and hence

$$|\nabla^+ h|^2(u) = |\nabla^+ f|^2(u).$$

Thus,

$$|\nabla^+(\varepsilon h)|^2(u) = \varepsilon^2 |\nabla^+ f|^2(u).$$

Moreover,

$$e^{\varepsilon h(u)} = 1 + O(\varepsilon).$$

Hence

$$\frac{1}{2\kappa_w} \int_{\mathcal{V}} |\nabla^+(\varepsilon h)|^2(u) e^{\varepsilon h(u)}\, d\mu(u) = \frac{\varepsilon^2}{2\kappa_w} \int_{\mathcal{V}} |\nabla^+ f|^2(u)\, d\mu(u) + \mathcal{O}(\varepsilon^3).$$

Combining the two expansions gives

$$\frac{\varepsilon^2}{2} \operatorname{Var}_\mu(f) + O(\varepsilon^3) \le \frac{\varepsilon^2}{2\kappa_w} \int_{\mathcal{V}} |\nabla^+ f|^2(u)\, d\mu(u) + O(\varepsilon^3).$$

Dividing by $\varepsilon^2/2$ and letting $\varepsilon \to 0$, we get

$$\operatorname{Var}_\mu(f) \le \frac{1}{\kappa_w} \int_{\mathcal{V}} |\nabla^+ f|^2(u)\, d\mu(u).$$

Finally, substituting the definition of the one-sided local slope yields

$$\operatorname{Var}_\mu(f) \le \frac{1}{\kappa_w} \int_{\mathcal{V}} \sup_{v \sim u} [f(u) - f(v)]_+^2\, d\mu(u).$$

This proves the desired inequality. □

K Proof of Theorem 7.2

Proof. The proof proceeds by a parallel ε-perturbation argument to that of Theorem 7.1 (Appendix 7.1), now exploiting the *lower* curvature bound $\kappa_w \ge -K$ to derive a reversed functional inequality.

Let $h := f - \bar{f}$ where $\bar{f} := \int_V f\, d\mu$, so that $\int_V h\, d\mu = 0$ and $\operatorname{Var}_\mu(f) = \int_V h^2\, d\mu$. Since $\kappa_w \ge -K$, the entropy functional satisfies the $(-K)$-convexity inequality along admissible W_1-geodesics: for any $\nu_0, \nu_1 \in \mathcal{P}(V)$ and any constant-speed geodesic $(\nu_t)_{t\in[0,1]}$,

$$\operatorname{Ent}_m(\nu_t) \;\le\; (1-t) \operatorname{Ent}_m(\nu_0) + t \operatorname{Ent}_m(\nu_1) + \frac{K}{2}\, t(1-t)\, \mathrm{W}_1(\nu_0, \nu_1)^2. \tag{20}$$

The additional positive term $+\frac{K}{2} t(1-t)\mathrm{W}_1^2$ reflects the hyperbolic-like expansion: entropy along the geodesic may exceed the linear interpolation, unlike the positive curvature case.

Set $\nu_0 = \frac{e^{\varepsilon h}}{Z_\varepsilon}\mu$ with $Z_\varepsilon = \int_V e^{\varepsilon h}\, d\mu$, and $\nu_1 = \mu$. Rearranging equation 20 at $t \downarrow 0$ and using the same first-variation estimate as in Lemma I.1, we obtain the following *reversed* modified log-Sobolev inequality: for every $g : V \to \mathbb{R}$,

$$\operatorname{Ent}_\mu(e^g) \;\ge\; \frac{1}{2K} \int_V |\nabla^+ g|^2(u)\, e^{g(u)}\, d\mu(u) \;-\; \frac{K}{2}\, \mathrm{W}_1(\nu_0, \nu_1)^2 \operatorname{Ent}_\mu(e^g), \tag{21}$$

where $|\nabla^+ g|^2(u) := \sup_{v\sim u} [g(u) - g(v)]_+^2$.

To bound the W_1 term in equation 21, we use the graph diameter. For any two probability measures on V,

$$\mathrm{W}_1(\nu_0, \nu_1) \;\le\; D, \tag{22}$$

since $d(u, v) \le D$ for all $u, v \in V$ by definition of diameter. Substituting into equation 21 and rearranging,

$$\left(1 + \frac{KD^2}{2}\right) \operatorname{Ent}_\mu(e^g) \;\ge\; \frac{1}{2K} \int_V |\nabla^+ g|^2(u)\, e^{g(u)}\, d\mu(u). \tag{23}$$

Apply equation 23 to $g = \varepsilon h$ with $\varepsilon > 0$. Expanding both sides as $\varepsilon \to 0$, using $e^{\varepsilon h} = 1 + \varepsilon h + \frac{\varepsilon^2}{2} h^2 + O(\varepsilon^3)$ and $\int_V h\, d\mu = 0$:

Left-hand side. Since $\operatorname{Ent}_\mu(e^{\varepsilon h}) = \frac{\varepsilon^2}{2}\operatorname{Var}_\mu(f) + O(\varepsilon^3)$, we have

$$\left(1 + \frac{KD^2}{2}\right) \operatorname{Ent}_\mu(e^{\varepsilon h}) = \left(1 + \frac{KD^2}{2}\right) \frac{\varepsilon^2}{2} \operatorname{Var}_\mu(f) + O(\varepsilon^3). \tag{24}$$

Right-hand side. Since $|\nabla^+(\varepsilon h)|^2(u) = \varepsilon^2 |\nabla^+ h|^2(u)$ and $e^{\varepsilon h(u)} = 1 + O(\varepsilon)$,

$$\frac{1}{2K} \int_V |\nabla^+(\varepsilon h)|^2(u)\, e^{\varepsilon h(u)}\, d\mu(u) = \frac{\varepsilon^2}{2K} \int_V |\nabla^+ f|^2(u)\, d\mu(u) + O(\varepsilon^3), \tag{25}$$

where we used $|\nabla^+ h|^2 = |\nabla^+ f|^2$ since subtracting the constant $\bar{f}$ does not change edge differences. Combining, dividing by $\varepsilon^2/2$, and letting $\varepsilon \to 0$:

$$\left(1 + \frac{KD^2}{2}\right) \mathrm{Var}_\mu(f) \;\geq\; \frac{1}{K} \int_V \sup_{v \sim u} [f(u) - f(v)]_+^2 \, d\mu(u). \tag{26}$$

Finally, since $1 + \frac{KD^2}{2} \leq e^{KD^2/2}$ for all $K, D \geq 0$ (by the inequality $1 + x \leq e^x$), dividing both sides by $e^{KD^2/2}$ yields the stated bound:

$$\mathrm{Var}_\mu(f) \;\geq\; \frac{e^{-KD^2/2}}{K} \int_V \sup_{v \sim u} [f(u) - f(v)]_+^2 \, d\mu(u). \tag{27}$$

□

L Proof of Theorem 4.3

Proof. Under the uniform generator L_0, set $k_w := |\mathrm{M}_{\mathcal{G}}(z, w)|$ and $k_w^+ := |\mathrm{M}_{\mathcal{G}^+}(z, w)|$, so that $k_w^+ \geq k_w$ by assumption. Each two-hop path $z \to m \to w$ carries unit weight, hence $\mathrm{L}_0^2(z, w) = k_w$ and $(\mathrm{L}_0^+)^2(z, w) = k_w^+$. The path weights satisfy $\ell^{\mathcal{G}}(z, m, w) = 1/k_w$ and $\ell^{\mathcal{G}^+}(z, m, w) = 1/k_w^+$ for every midpoint m. By the arithmetic–geometric mean inequality, the supremum in the pairwise local functional is attained at the uniform distribution $\alpha(m) = 1/k_w$, yielding

$$\mathrm{H}_{\mathrm{L}_0}^{\mathcal{G}}(z, \{w\}) \;=\; k_w \cdot \left(\frac{1}{k_w}\right)^2 \;=\; \frac{1}{k_w}, \qquad \mathrm{H}_{\mathrm{L}_0}^{\mathcal{G}^+}(z, \{w\}) \;=\; \frac{1}{k_w^+}. \tag{28}$$

Since $k_w^+ \geq k_w$, we obtain $\mathrm{H}_{\mathrm{L}_0}^{\mathcal{G}^+}(z, \{w\}) \leq \mathrm{H}_{\mathrm{L}_0}^{\mathcal{G}}(z, \{w\})$ for every $w \in \mathrm{S}_2^{\mathcal{G}}(z)$. Aggregating over the full 2-sphere gives

$$\mathrm{H}_{\mathrm{L}_0}^{\mathcal{G}^+}(z, \mathrm{S}_2^{\mathcal{G}}(z)) \;\leq\; \mathrm{H}_{\mathrm{L}_0}^{\mathcal{G}}(z, \mathrm{S}_2^{\mathcal{G}}(z)). \tag{29}$$

Applying the decreasing transformation $-2\log(\cdot)$ (Definition 3.2) yields $\kappa_w^{\mathcal{G}^+}(z) \geq \kappa_w^{\mathcal{G}}(z)$. Strict improvement when $k_w^+ > k_w$ for some w follows from the strict monotonicity of $-2\log(\cdot)$. □

M Algorithmic Computation of Weak Entropic Curvature and Convergence Analysis

M.1 Algorithm Implementation of Weak Entropic Curvature

Algorithm 2 Exact Computation of $\kappa_w(\mathcal{G})$

Input: Graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, generator $\mathtt{L} : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$
Output: Curvatures: local $\boldsymbol{\kappa}$ and global $\kappa_w(\mathcal{G})$
foreach $z \in \mathcal{V}$ **do**
  Let $\mathrm{S}_1(z) \leftarrow \{v \in \mathcal{V} : (z, v) \in \mathcal{E}\}$
  Let $\mathrm{S}_2(z) \leftarrow \{w \in \mathcal{V} : d(z, w) = 2\}$
  if $\mathrm{S}_2(z) = \emptyset$ **then**
    $\kappa[z] \leftarrow \infty$
  else
    1. Precompute for each $w \in \mathrm{S}_2(z)$ and $u \in]z, w[$: $\ell(z, u, w) \leftarrow \frac{\mathtt{L}(z,u)\mathtt{L}(u,w)}{\mathtt{L}(z,w)^2}$
    2. Define objective function $\mathrm{J}(\alpha)$ for $\alpha \in \Delta_{|\mathrm{S}_1(z)|}$:
    $\mathrm{J}(\alpha) \leftarrow \sum_{w \in \mathrm{S}_2(z)} \mathtt{L}(z, w)^2 \prod_{u \in]z,w[} \left(\frac{\alpha(u)}{\mathtt{L}(z,u)}\right)^{2\ell(z,u,w)}$
    3. Solve constrained maximization:
    $\mathrm{H}^* \leftarrow \sup_\alpha \mathrm{J}(\alpha) \quad \text{s.t.} \sum_u \alpha(u) = 1, \alpha(u) \geq 0$
    $\kappa[z] \leftarrow -2 \log(\mathrm{H}^*)$
  end if
end foreach
$\kappa_w(\mathcal{G}) \leftarrow \inf_{z \in \mathcal{V}} \kappa[z]$
return $\kappa, \kappa_w(\mathcal{G})$

M.2 Fast Convergence of the Entropic Optimization

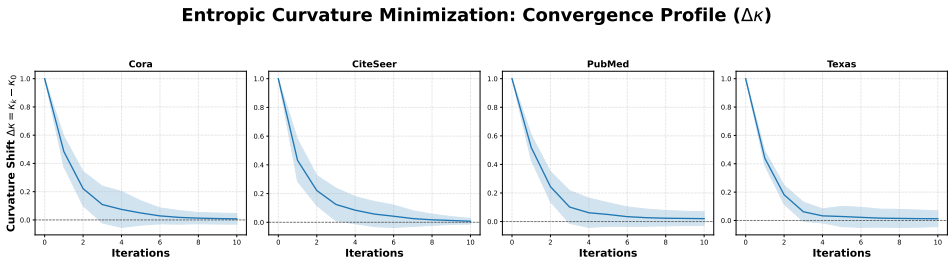

Figure 6: Empirical Convergence of Weak Entropic Curvature Optimization

To validate the computational tractability of the Weak Entropic Curvature, we empirically analyze the convergence behavior of the underlying optimization problem. As detailed in Algorithm 2, calculating the curvature at a specific node z requires maximizing the local functional $\mathrm{H}_{\mathrm{L}}(z, \mathrm{S}_2(z))$ over the probability simplex. We employ the `SLSQP` method from the SciPy library (Virtanen et al., 2020), which is well-suited for handling the linear constraints of the simplex $\Delta_{|\mathrm{S}_1(z)|}$.

Methodology. Since the absolute value of the weak entropic curvature $\kappa_w(z)$ varies significantly across different local topologies (e.g., varying from positive in cliques to negative in trees), aggregating raw optimization traces would yield uninterpretable variance. Instead, we track the *normalized convergence residual*. For a randomly sampled set of 50 nodes per dataset, we track the curvature estimate $\kappa^{(t)}$ at iteration t. We define the normalized residual $\Delta\tilde{\kappa}^{(t)}$ as,

$$\Delta\tilde{\kappa}^{(t)} = \frac{\kappa^{(t)} - \kappa^*}{\kappa^{(0)} - \kappa^*}, \tag{30}$$

where $\kappa^{(0)}$ is the initial curvature estimate (uniform distribution initialization) and κ^* is the converged value (minimum reached within the tolerance threshold). This metric effectively maps the optimization landscape to the unit interval $[0, 1]$, where 1 represents initialization and 0 represents convergence.

Results. Figure 6 illustrates the mean convergence profile ($\pm$ standard deviation) across four distinct benchmarks: Cora, CiteSeer, PubMed, and Texas.

The empirical results reveal a consistent "fast-decay" profile across all datasets, independent of the graph's homophily or size:

- **Rapid Descent:** The solver achieves the majority of the optimization gain within the first **3 to 5 iterations**. The steep initial drop indicates that the objective function $J(\alpha)$ is well-behaved and that the gradient signal effectively guides the transport plan α toward the optimal configuration.
- **Low Variance:** The narrow shaded regions (standard deviation) indicate that this convergence speed is robust. Regardless of whether a node resides in a dense community (Cora) or a sparse heterophilic bottleneck (Texas), the optimization complexity remains predictably low.
- **Stability:** Beyond 6 iterations, the updates become negligible ($< 10^{-4}$), confirming that setting a low maximum iteration cap (e.g., $k_{iter} \approx 10$) in the complexity term $O(n \cdot k_{iter} \cdot d_{max}^2)$ is sufficient for high-fidelity curvature estimation.

This empirical evidence supports our claim that the computational overhead of Entropic Curvature is dominated by neighborhood discovery rather than the optimization step itself, making it scalable to large graph datasets.

A critical observation from Figure 6 is that the convergence speed is effectively invariant to graph homophily. The solver converges just as rapidly for the disordered, heterophilic graphs as it does for structured, homophilic citation networks. This proves that the optimization landscape of the local entropic functional remains well-behaved even in the presence of noisy or fractured local topologies, distinguishing it from spectral approximations that often degrade on graphs with poor expansion properties.

N Implementation of Entropic Structural Encodings

In this section, we detail the implementation of the structural encodings used in our comparative analysis. All curvature computations were pre-processed and cached to ensure training efficiency.

N.1 Entropic Structural Encoding (ENT)

Our proposed method computes the Weak Entropic Curvature $\kappa_w(u)$ for every node u using the exact optimization procedure described in Section 6. To capture the local geometric context, we construct a 3-dimensional signature vector $\mathbf{p}_u^{\text{ENT}}$ consisting of the node's curvature and the extrema of its neighbors:

$$\mathbf{p}_u^{\text{ENT}} = \left[\kappa_w(u), \quad \min_{v \in \mathcal{N}(u)} \kappa_w(v), \quad \max_{v \in \mathcal{N}(u)} \kappa_w(v) \right]^\top \tag{31}$$

Our choice of a 3-dimensional signature $\mathbf{p}_u^{\text{ENT}}$, i.e., focusing on $\kappa_w(u)$ and the neighborhood extrema, is deliberate. By explicitly encoding the minimum ($\min \kappa_w$) and maximum ($\max \kappa_w$) curvature of the neighbors, the signature captures the geometric gradient, or anisotropy, of the local manifold. This differentiates nodes that sit on the boundary of a cluster (high max, low min) from those deep within a clique (high max, high min). Crucially, this 3D representation is more compact than the 5D statistical vector used in LCP (Fesser & Weber), reducing the risk of overfitting on small, data-scarce heterophilic datasets like Texas and Cornell while retaining the essential topological signal.

N.2 Baseline Implementations

We compare our method against the following structural and positional encodings.

1. Local Curvature Profiles (LCP). Following the standard implementation Fesser & Weber, LCP augments node features with statistics derived from the discrete Ollivier-Ricci Curvature (ORC).

- **Computation:** We compute the exact ORC for every edge (u, v) in the graph using the `GraphRicciCurvature` library. The transport cost is defined based on the shortest path distance between neighbors.
- **Encoding:** For each node u, we aggregate the curvature values of all incident edges $E_u = \{\kappa_{\text{ORC}}(u, v) \mid v \in \mathcal{N}(u)\}$. The resulting feature vector $\mathbf{p}_u^{\text{LCP}} \in \mathbb{R}^5$ consists of the following statistics:

$$\mathbf{p}_u^{\text{LCP}} = [\min(E_u), \max(E_u), \text{mean}(E_u), \text{std}(E_u), \text{median}(E_u)]^\top$$

  Nodes with no neighbors are assigned a zero vector.

2. Laplacian Positional Encodings (LAPE). We compute the graph Laplacian $\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$. The positional encoding is formed by the k smallest non-trivial eigenvectors of $\mathbf{L}$. These eigenvectors capture the low-frequency modes of the graph, preserving global distance information.

3. Random Walk Positional Encodings (RWPE). RWPE captures the structural role of a node by examining the probability of a random walker returning to itself. The encoding $\mathbf{p}_u^{\text{RWPE}} \in \mathbb{R}^k$ is defined by the diagonal entries of the random walk diffusion matrix at various steps:

$$\mathbf{p}_u^{\text{RWPE}} = \left[\mathbf{P}_{uu}^1, \mathbf{P}_{uu}^2, \ldots, \mathbf{P}_{uu}^k\right]^\top$$

where $\mathbf{P} = \mathbf{D}^{-1}\mathbf{A}$ is the transition matrix. This encoding effectively characterizes the local expansion properties around each node.

4. Ego-Network Encodings (EGO). This baseline relies on explicit structural feature extraction. For each node u, we extract its k-hop ego network and compute classic graph theoretic invariants, such as node degree, number of edges, and clustering coefficient, to serve as the structural signature.

O Alternative Generator Constructions

In this section, we detail the construction of reversible generators for general and concentrated measures, extending the standard uniform case presented in Section 3.

Example 2 (General Positive Measure)**.** For an arbitrary strictly positive measure $\mathtt{m}(\cdot) > 0$, we can satisfy reversibility by defining the off-diagonal terms as follows,

$$\mathtt{L}(u, v) = \mathbf{1}_{\{u,v\}\in\mathcal{E}} \left(\frac{\mathtt{m}(v)}{\mathtt{m}(u)}\right)^{1/2}. \tag{32}$$

Finally, we consider the case where we wish to concentrate the measure on a specific subset $\mathcal{C} \subset \mathcal{V}$. Defining m strictly as the indicator function $\mathtt{m}_{\mathcal{C}} = \mathbf{1}_{\mathcal{C}}$ is theoretically prohibitive, as m must be strictly positive to define a valid inner product space. Furthermore, if $\mathcal{C}$ is not connected (i.e., non-convex), a valid irreducible generator cannot exist on the support. We define a smoothed approximation $\mathtt{m}_\epsilon \simeq \mathtt{m}_{\mathcal{C}}$, which yields a Metropolis-Hastings style generator.

Example 3 (Concentrated Measure Approximation)**.** Let the target measure be approximated by $\mathtt{m}_\epsilon(u) = \max(\mathbf{1}_{\mathcal{C}}(u), \epsilon)$ where $0 < \epsilon \ll 1$. The generator $\mathtt{L}$ satisfying reversibility is given by,

$$\mathtt{L}(u, v) = \mathbf{1}_{\{u,v\}\in\mathcal{E}} \cdot \min\left(1, \frac{\mathtt{m}_\epsilon(v)}{\mathtt{m}_\epsilon(u)}\right). \tag{33}$$

P Interpretation of Graph Generators and Reference Measures

The choice of the reference measure $\mathtt{m}$ and the resulting generator $\mathtt{L}$ is a defining feature of the entropic curvature framework, as it dictates the *geometry of the mass* across the graph. Unlike standard curvature metrics that are fixed by the adjacency matrix, our framework allows the geometry to be modulated by a functional prior.

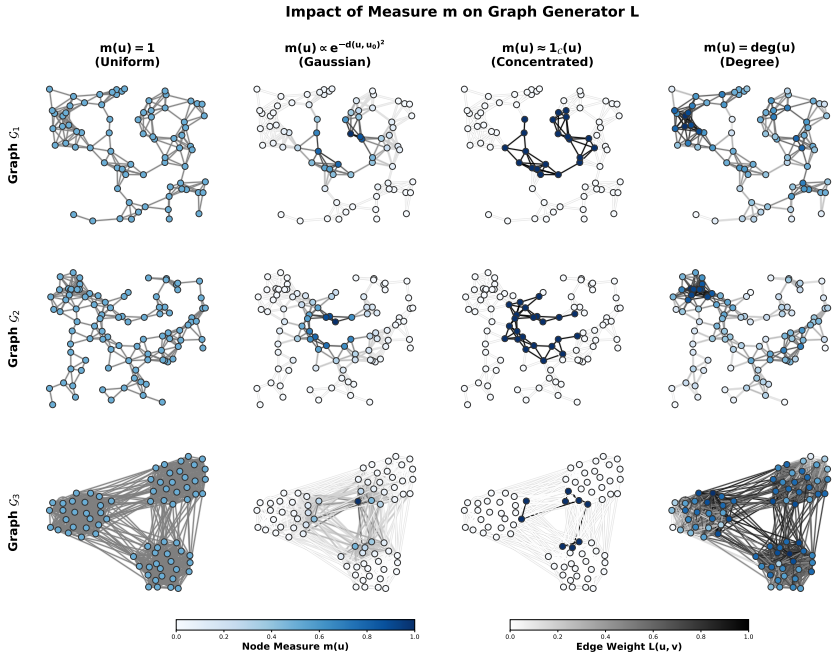

Figure 7: **Geometric reconfiguration via mass distribution.** Changing the reference measure $\mathfrak{m}$ fundamentally alters the graph's diffusion dynamics.

P.1 Thermodynamic Interpretation of Generators

The detailed balance condition $\mathfrak{m}(u)\mathtt{L}(u,v) = \mathfrak{m}(v)\mathtt{L}(v,u)$ ensures that the process is reversible with respect to $\mathfrak{m}$. From a thermodynamic perspective, $\mathfrak{m}$ represents the steady-state distribution or the capacity of nodes.

- **Uniform Measure ($m = 1$):** Corresponds to a *flat* thermodynamic prior where every node has equal capacity. The resulting generator $\mathtt{L}_0$ is the standard combinatorial Laplacian, where information flow is driven purely by the topological connectivity.
- **Degree-Based Measure ($m = deg$):** Aligns the geometry with the graph's natural connectivity. This measure typically smooths out the influence of high-degree hubs, as the cost of moving mass is normalized by the node's local popularity.
- **Concentrated/Gaussian Measure:** When $\mathfrak{m}$ is concentrated on a subset C (as seen in Figure 7), the generator $\mathtt{L}$ behaves like a *potential well*. Edges leading toward the concentration center are strengthened (higher jump rates), while edges leading away are *penalized*. This effectively creates a geometric *drift* in the Wasserstein geodesics.

P.2 Impact of Measure $\mathfrak{m}$ on Generator $\mathbf{L}$

Figure 7 visualizes how varying the reference measure $\mathfrak{m}$ fundamentally reconfigures the edge weights $\mathtt{L}(u,v)$ across three different graph topologies ($\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$).

- **Column 1 (Uniform):** The edge weights are binary (0 or 1), representing the raw topology where entropic curvature identifies structural bottlenecks like bridges in the Barbell graph.
- **Column 2 (Gaussian):** By centering a Gaussian measure $\mathfrak{m}(u) \propto e^{-d(u,u_0)^2}$ at a specific node u_0, the graph is transformed into a *localized metric space*. The curvature is now highest near the center u_0 and decays as one moves toward the periphery, effectively *cooling* the message-passing in the center to preserve local feature distinctness.
- **Column 3 (Concentrated):** Using the indicator approximation $\mathfrak{m}_\epsilon \simeq 1_C$, we observe a sharp transition in the generator weights. Edges within the community C maintain high

weights, while boundary edges are attenuated. This illustrates how entropic curvature can be used to *isolate* communities by creating artificial geometric boundaries that prevent feature oversmoothing across distinct clusters.

- **Column 4 (Degree):** Weighting by degree regularizes the influence of hubs. In dense graphs like $\mathcal{G}_3$, this choice spreads the entropic pressure more evenly, preventing the exponential collapse of variance typically seen in high-degree expanders.

Q Complexity Analysis

In this section, we analyze the computational complexity of the Entropic Curvature Algorithm, c.f., Algorithm 2. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph with $n = |\mathcal{V}|$ nodes and maximum degree $d_{\max}$. The computation of the Local Weak Entropic Curvature is a node-wise operation. We distinguish between the general setting involving arbitrary generators and the simplified case of standard unweighted graphs, i.e., $\mathtt{L} = \mathtt{L}_0$.

Q.1 The General Case of Arbitrary Generator $\mathbf{L}$ and Mass $\mathbf{m}$

For an arbitrary generator L and measure m, computing the curvature at a node z requires solving a constrained maximization problem over the probability simplex. This process involves three steps:

- **Step 1 - Neighborhood Discovery:** We need to identify the 1-hop neighborhood $\mathrm{S}_1(z)$ and the 2-hop neighborhood $\mathrm{S}_2(z)$. In the worst case, $|\mathrm{S}_1(z)| \leq d_{\max}$ and $|\mathrm{S}_2(z)| \leq d_{\max}^2$, yielding a complexity cost of $O(d_{\max}^2)$. For the optimization step, we must identify the set of midpoints $]z, w[= \mathrm{S}_1(z) \cap \mathrm{S}_1(w)$ for every $w \in \mathrm{S}_2(z)$. Finding the intersection of two sorted neighbor lists takes $O(d_{\max})$ as $max(|\mathrm{S}_1(z)|, |\mathrm{S}_1(w)|) \leq d_{\max}$. The overall complexity of this step is therefore $O(d_{\max}^2)$.
- **Step 2 - Calculation of the coefficients:** The algorithm computes the transport coefficients $l(z, u, w)$ for all the paths $z \to u \to w$, where $u \in \mathrm{S}_1(z)$ and $w \in \mathrm{S}_2(z)$. This step requires iterating through potential paths in the local neighborhood, scaling as $O(d_{\max}^3)$ in dense subgraphs.
- **Step 3 - Optimization:** The core computation is determining $\sup_\alpha J(\alpha)$ using Sequential Least Squares Programming (SLSQP) Nocedal & Wright (2006). The complexity of this iterative procedure is given by $O(k_{\text{iter}} \cdot C_{\text{step}})$ where k_{iter} is the number of iterations required for convergence, and C_{step} is the cost of computing the objective function and its gradient over the simplex. The cost per iteration, C_{step}, is directly determined by the output of Steps 1 and 2. Evaluating the objective function $J(\alpha)$ requires summing over the 2-hop neighborhood $\mathrm{S}_2(z)$, i.e, identified in Step 1, and accessing the precomputed sparse coefficients $l(z, u, w)$, i.e., calculated in Step 2.

Consequently, by aggregating the costs of neighborhood discovery, coefficient precomputation, and the iterative optimization, the overall complexity to compute the local entropic curvature for a single node z is $O(d_{\max}^3 + k_{\text{iter}} \cdot d_{\max}^2)$. To obtain the global weak entropic curvature $\kappa_w(\mathcal{G})$, the algorithm must execute this procedure for every node $z \in \mathcal{V}$, continuously tracking the minimum curvature value observed. Therefore, the total time complexity for the entire graph is $O(n \cdot (d_{\max}^3 + k_{\text{iter}} \cdot d_{\max}^2))$.

Q.2 The Simplified Case of $\mathbf{L} = \mathtt{L_0}$ and $\mathbf{m} = \mathbf{m_0}$

In the standard unweighted setting where the generator is $\mathtt{L}_0$, defined as $\mathtt{L}_0(z, z') = 1_{z \sim z'}$, and the measure m_0 is uniform, the computational complexity is significantly reduced.

Crucially, Step 2 (Calculation of coefficients) is unnecessary since the transition weights are binary, the transport coefficients $\ell(z, u, w)$ reduce to unity for every valid path $z \to u \to w$ in the neighborhood,

$$\ell(z, u, w) = \frac{\mathtt{L}_0(z, u)\mathtt{L}_0(u, w)}{\mathtt{L}_0^2(z, w)} = 1.$$

This simplification eliminates the cubic cost associated with precomputing path weights. Consequently, the algorithm proceeds directly from Neighborhood Discovery (Step 1) to Optimization

(Step 3). While the optimization must still be performed iteratively (e.g., using SLSQP), the objective function evaluation is streamlined as it no longer requires retrieving stored coefficient tensors. The global complexity to compute the local weak entropic curvature is $O(k_{\text{iter}} \cdot d_{\max}^2)$., and the complexity for weak entropic curvature on the entire graph is $O(n \cdot k_{\text{iter}} \cdot d_{\max}^2)$.

Q.3 Empirical Complexity

To complement our theoretical analysis, we empirically evaluate the runtime performance of the Weak Entropic Curvature algorithm, c.f., Algorithm 2. We utilize Stochastic Block Models (SBM) to generate synthetic graphs of varying sizes Holland et al. (1983), allowing us to systematically control the number of nodes $|\mathcal{V}|$. This setup enables us to benchmark the scalability of our approach across both the general case, i.e., arbitrary $\mathtt{L}, m$, and the simplified case, i.e., $\mathtt{L} = \mathtt{L}_0, m = m_0$. Given intra-community edge probability p and inter-community probability q, the expected number of edges is given by $E[|\mathcal{E}|] = \frac{n}{2}[(\frac{n}{k} - 1)p + (n - \frac{n}{k})q]$, while the expected degree is $E[d] = \frac{n}{k}p + (n - \frac{n}{k})q$, where $n = \mathcal{V}|$. By fixing the ratio of p and q, we benchmark the scalability of our approach across both the general case (arbitrary L, m) and the simplified case ($\mathtt{L} = \mathtt{L}_0, m = m_0$).

We measure the average time required to compute the global curvature $\kappa_w(\mathcal{G})$ as the graph size increases. The results, summarized in Table 4, highlight the practical computational efficiency gained in the simplified regime due to the elimination of the coefficient precomputation step.

Table 4: Empirical runtime comparison (in seconds) for computing global Weak Entropic Curvature on SBM ($k = 2$). We vary $|\mathcal{V}|$ and set p, q to target specific maximum degrees $d_{\max}$.

Graph Statistics			**SBM Parameters**		**Execution Time (s)**	
$\lvert\mathcal{V}\rvert$	$E[\lvert\mathcal{E}\rvert]$	$d_{\max}$	p	q	**Simplified** $(\mathtt{L_0}, \mathbf{m_0})$	**General** $(\mathbf{L}, \mathbf{m})$
1,000	5,000	≈ 10	0.015	0.005	35.770	406.997
1,000	15,000	≈ 30	0.040	0.020	1471.913	4903.886
2,000	20,000	≈ 20	0.015	0.005	485.642	5400.812
2,000	50,000	≈ 50	0.040	0.010	13142.537	38427.988
5,000	75,000	≈ 30	0.010	0.002	4080.36	39804.04
10,000	250,000	≈ 50	0.008	0.002	36029.70	335796.21

Q.4 Comparison with other Curvatures

We compare the complexity of Entropic Curvature against established discrete curvature notions in Table 5.

- **Forman Curvature** is computed by summing the weighted contributions of the edges immediately incident to a target link's endpoints, resulting in a linear complexity of $O(d_{\max})$ per edge.
- **Ollivier-Ricci Curvature** requires solving the Earth Mover's Distance (Optimal Transport) problem between neighborhood measures.
- **Balanced Forman Curvature** is computed by augmenting the standard Forman curvature formula with additional terms that account for triangles (3-cycles) and 4-cycles (squares) rooted at the edge, effectively capturing the local flatness of the graph geometry.

Our Entropic formulation shares the same complexity class as Ollivier-Ricci but solves a smooth concave maximization problem rather than a linear program, which often exhibits faster wall-clock convergence in practice due to efficient gradient-based solvers.

Moreover, unlike global spectral methods, e.g., Eigen-decomposition, which are difficult to parallelize effectively, our Entropic Curvature computation is node-wise independent, making it *embarrassingly parallelizable*. On large-scale graphs, this allows the runtime to scale linearly with the number of available CPU cores. Furthermore, unlike the Ollivier-Ricci curvature, which requires solving a Linear Programming, e.g.,Earth Mover's Distance (EMD), problem at every edge, our formulation relies on a smooth concave maximization. This structure allows the use of gradient-based solvers,

Table 5: Complexity comparison of discrete graph curvatures. $n = |\mathcal{V}|$: number of nodes, $m = |\mathcal{E}|$: number of edges, Δ: max degree.

Curvature Metric	Definition Basis	Per-Node Complexity	Per-Edge Complexity	Global Complexity
Forman Lin & Yau (2003)	Combinatorial	–	$O(d_{\text{max}})$	$O(m \cdot d_{\text{max}})$
Ollivier-Ricci Ollivier (2009)	Optimal Transport	–	$O(d_{\text{max}}^3)$	$\mathcal{O}(m \cdot d_{\text{max}}^3)$
Balanced Forman Topping et al. (2022)	Combinatorial	–	$O(d_{\text{max}}^2)$	$O(m \cdot d_{\text{max}}^2)$
Weak Entropic (Ours)	Optimization	$O(k_{\text{iter}} \cdot d_{\text{max}}^2)$	–	$O(n \cdot k_{\text{iter}} \cdot d_{\text{max}}^2)$

like SLSQP, which are often numerically more stable and faster in wall-clock time than the simplex or interior-point methods required for EMD.