WHAT DOES GRAPH CONSTRUCTION PRESERVE? A DISCRETE RICCI CURVATURE DIAGNOSTIC FOR POINT CLOUD GRAPHS

Anonymous authors
Paper under double-blind review

ABSTRACT

Graph representation learning is a powerful general method for learning encodings for complex data, but the impact of graph construction methods on learning tasks remains understudied. Understanding how a graph construction method extracts particular data properties is important both for our understanding of what properties of the problem contribute to model performance, and to further improve downstream model performance. In this work we propose a diagnostic *separation* score—the Wasserstein-1 distance between discrete Ricci curvature distributions of labeled populations—to evaluate graph constructions. This diagnostic is training-free, and exposes some properties preserved by graph construction methods. We evaluate this diagnostic in the context of two application domains: quantum chemistry (the QM9 dataset) and collider physics (the JetSet dataset). On QM9, where bonds are known, separation based on the Ollivier-Ricci variant of discrete curvature tracks bond-recovery F1 across 19 constructions (Spearman $\rho = 0.88$) without seeing the bonds and favors the 1.8 Å radius graph construction, whereas separation based on the Forman-Ricci variant does not. On JetSet, which have no ground truth, all construction methods separate the flavors, but once we control for the number of nodes in the graph, only the radius graph construction method keeps most of its separation over all nodes. In addition to the diagnostic, we also measure an *incremental predictive value* of curvature as a model input: on QM9, a model gains most on the 1.8Å graph teh diagnostic selects, whereas on jets curvature adds at most 0.02 AUROC to a mean aggregating GNN. However, this gain is mostly matched by using a simpler feature, namely, the node degree.

1 INTRODUCTION

In recent years, computational scientific discovery has been advanced by representation learning, where learned encodings of high-dimensional data have led to significant improvements in challenging classification tasks. A central concern, however, continues to be to better understand how these encodings work, both for improving how scientifically meaningful structure might be interpreted within learned encodings and for how we might additionally bias the encoding learning process to further improve discovery task performance (Bengio et al., 2013). Geometric and graph representation methods have been found to be particularly effective for a broad class of tasks, with each involving graph construction processes that impose different inductive biases that can lead to distinct learning behaviors (Bronstein et al., 2017). We aim to answer the following questions: what structure does a graph construction method preserve, whether that structure can be probed without training a model, and how the structure relates to what a downstream model extracts.

Graphs are a natural setting for these questions: they directly encode which entities are related and the shape those relations form, and as scientific data in chemistry, sociology, and physics are increasingly represented as graphs, the choice of representation shapes what structural information a model can capture (De Domenico et al., 2013). Yet that choice remains somewhat ad hoc: there is no practical way to determine which construction faithfully represents the geometry of the data, a gap that has motivated work on graph topology learning (Pu et al., 2021; Franceschi et al., 2019).

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.

Domain applications (e.g., classifying jets at particle colliders) often fix the construction by convention (Qu & Gouskos, 2020; Shlomi et al., 2021; Beuria, 2024; Gaertner & Reiten, 2023), or based on downstream task performance (Araz et al., 2026). However, task performance cannot distinguish a construction that discards the structure the label depends on from a model that cannot extract existing structure.

Two lines of work treat the construction in a less ad hoc way. Asymptotic theory evaluates constructions by how faithfully they recover the underlying manifold (Maier et al., 2008; Calder & García Trillos, 2022; van der Hoorn et al., 2021; 2023). However, these results ignore labels, assume uniform sampling, and hold only for graphs with large numbers of nodes. Graph structure learning treats the edge set as a parameter and fits it to the task loss (Franceschi et al., 2019; Kazi et al., 2023), so the learned graph is whatever structure lowers that loss for that model. Nothing in the objective ties it to the physical structure the label depends on, and it requires training.

Discrete Ricci curvature (Lin et al., 2011; Forman, 2003) is a graph descriptor that identifies structural bottlenecks (negative curvature) where information flow is constrained and topological clusters (positive curvature) where local correlations are higher (Ni et al., 2019). Curvature distributions have been used to classify networks (Sreejith et al., 2016; Weber et al., 2017) and to evaluate graph generative models, where there is no agreement on which graph descriptor to use (O'Bray et al., 2022; Thompson et al., 2022) and curvature has been proposed as a stable and expressive choice (Southern et al., 2023). However, these studies fix the graphs in advance, measure distances to a reference set of graphs, and do not control for graph size, which drives curvature through node degree. Curvature has also been used as input for classification tasks, e.g., at particle colliders (Beuria, 2024; Gaertner & Reiten, 2023). Both of these works compute curvature on one construction fixed in advance, and neither asks whether its result would survive the other's construction. Curvature-based rewiring is an approach to reduce the phenomenon of over-squashing in graph neural networks (GNNs) (Topping et al., 2022; Nguyen et al., 2023; Fesser & Weber, 2024). This line of work leverages curvature and does go beyond having a fixed graph; however, it also assumes the original edges are correct, which Tori et al. (2025) question.

Because curvature is a deterministic function of the graph, it carries no information a model given the graph does not already have. For this reason, probing studies measure how easily information can be extracted from a representation (Pimentel et al., 2020; Xu et al., 2020). For message passing networks, existing theory predicts such a gap: both discrete curvatures we consider in this paper—Forman-Ricci curvature (FRC) and Ollivier-Ricci curvature (ORC), defined in § 2.2—depend on short cycles, which these networks cannot count (Xu et al., 2019; Chen et al., 2020), and using substructure counts as inputs improves their performance (Bouritsas et al., 2023). Any improvement therefore measures usable information—information present in the graph but not extracted by the model.

To our knowledge, no existing diagnostic is structural, conditioned on labels, and applicable to small scientific graphs with tens of nodes. We propose a principled geometric diagnostic based on discrete Ricci curvature. We hold the data and labels fixed, vary the construction, and use curvature to measure what structure each construction preserves. Making the construction the variable lets us i) choose a construction and its parameters without training, ii) control for size, without which a curvature descriptor detects size rather than structure, and iii) test whether a structural feature's gain is only degree.

Testing the diagnostic requires domains whose physical structure is understood independently of any graph built from the data. We use two, ordered by how much ground truth they offer. In QM9 (Ramakrishnan et al., 2014), we build graphs from atomic coordinates, and the true bond graph is known, so each construction can be checked against it. In jets (ATLAS Collaboration, 2025)—collimated sprays of particles produced in high-energy collisions (Kogler et al., 2019)—we build graphs from the angular coordinates and momenta of the reconstructed particle tracks, and no correct graph exists. The task there is flavor tagging, inferring the type of quark that initiated the jet. Its key signature, a displaced secondary decay vertex (ATLAS Collaboration, 2016), never enters an angular construction; what such a construction can preserve is the tight angular cluster formed by the decay products, and this is the structure we ask curvature to detect. QM9 therefore validates the diagnostic against a known answer before we apply it to jets, where constructions can only be compared with one another. Neither domain is in the large-n regime that asymptotic theory assumes: a QM9 molecule has at most 29 atoms, and a jet in JetSet has at most 40 tracks.

We score each construction in two ways. *Separation* is the Wasserstein-1 distance (Villani, 2009) between the curvature distributions of labeled populations, computed within bins of point count and composition and compared against the complete graph, whose curvature depends on size alone (§ 3). The bins are needed because curvature depends on degree, and degree grows with the number of points. *Incremental predictive value* measures what a model extracts: the gain from adding curvature to a model's inputs.

The two scores answer different questions, and on jets they disagree: the construction the diagnostic selects is not the one that gains most from curvature features, because what the models extract from the curvature channel is mostly the node degree that curvature encodes: a plain per-node degree feature recovers most of the gain (§ 5). This is the diagnostic's limit, stated up front: it measures what a construction preserves, not what a model will do with it.

We make the following contributions:

- **A training-free, size-controlled diagnostic**: the curvature separation between labeled populations at fixed point count and composition, read against the complete graph.
- **A positive control**: On QM9, without accessing the bonds, ORC separation tracks bond-recovery F1 across 19 constructions on five of six population pairs ($\rho = 0.76$–0.92) and selects the 1.8 Å ϵ construction, while FRC does not.
- **A size detector result**: on jets at fixed (n, n_{sv}), only the ϵ-based construction retains most of its separation, and every other construction falls to the FC construction.
- **Gain as extraction**: on jets a per-node degree feature recovers most of the gain from curvature features, and on QM9 the gain follows separation within a sweep but not across families.

The paper is organized as follows. We describe the graph construction families and discrete curvature types we use in § 2, the distributional and predictive measures we use in § 3, and the datasets, edge-recovery control used in QM9, and the model protocol in § 4. We present results in § 5 and conclude in § 6.

2 Background

2.1 Graph construction families

Every example, a molecule or a jet, is a set of N points that we represent as an undirected graph with one node per point. All constructions use the pairwise Euclidean distance $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$, combined with a per-point w_i (p_T for jets, atomic mass for QM9) in the generalized k_T metric (Cacciari et al., 2008)

$$d_{ij}^{(p)} = \min\left(w_i^{2p},\, w_j^{2p}\right) d_{ij}^2, \tag{1}$$

where $p = 0$ is purely geometric, $p = 1$ enhances low-weight (soft) points, and $p = -1$ enhances high-weight (hard) points. We consider six families:

k-nearest neighbors (kNN) (Steinbach & Tan, 2009): each node is connected to its k nearest neighbors by d_{ij}, producing sparse graphs with $O(kN)$ edges.

ϵ-neighborhood: An edge is added between nodes i and j if $d_{ij} < \epsilon$. The resulting graph density depends on the local spatial distribution of its local density.

k_T quantile threshold: We define a per-jet threshold based on the quantile of the pairwise distance distribution, $d_{\text{cut}} = \text{Quantile}_q\left(\{d_{ij}^{(p)}\}_{i\neq j}\right)$, and connect nodes (i, j) if $d_{ij}^{(p)} < d_{\text{cut}}$. The parameter $q \in (0, 1)$ controls the fraction of connected pairs and lets the connectivity adapt to each graph's density. Node degrees are kept within $[k_{\min}, k_{\max}]$, where nodes below $k_{\min}$ are linked to their $k_{\min}$ nearest neighbors by d_{ij} and $k_{\max}$ is imposed to bound computational cost.

Laman (Araz et al., 2026): We build locally rigid sparse graphs by ordering points in decreasing importance, initializing a triangle on the first three points, and each subsequent point is connected to its two nearest previously inserted neighbors according to $d_{ij}^{(p)}$.

Unique-k (Araz et al., 2026): A globally rigid generalization of the Laman approach where we fully connect the first $k+1$ points and then connect each subsequent point to its k nearest (as measured by $d_{ij}^{(p)}$) previously inserted neighbors.

Fully connected (FC): We connect every pair of points. Its curvature depends only on the number of nodes, so it serves as the null construction against which the scores of § 3.1 are reported. Parameters are swept per dataset (Table 1, Appendix A): 36 constructions on JetSet (12 default, 24 swept) and 22 on QM9 (13, 9), reducing to 33 and 19 distinct graphs after deduplication § A.1.

2.2 Discrete Ricci curvatures

Ricci curvature provides a local measure for how a manifold deviates from its Euclidean space (Bauer et al., 2013). We use its two most common discretizations on graphs and they respond to construction differently: FRC is determined by node degrees and triangle counts alone and therefore purely captures the topological structure, whereas ORC compares the neighborhoods of adjacent nodes through optimal transport, and the two are highly correlated on many networks but capture different aspects of structure (Samal et al., 2018).

Forman-Ricci Curvature. The augmented Forman-Ricci curvature (FRC) of an edge (i, j) is (Forman, 2003; Sreejith et al., 2016; Samal et al., 2018)

$$\kappa_{\mathrm{FR}}(i, j) = 4 - \deg(i) - \deg(j) + 3\,|\triangle(i, j)| \tag{2}$$

where $\deg(\cdot)$ denotes the degree of a node and $|\triangle(i, j)|$ the number of triangles supported on (i, j) (common neighbors of i and j).

Ollivier-Ricci Curvature. For an edge (i, j), the ORC is defined as follows (Lin et al., 2011):

$$\kappa_{\mathrm{OR}}(i, j) = 1 - \frac{W_1(\mu_i, \mu_j)}{d(i, j)}, \tag{3}$$

where μ_i keeps mass $\alpha = 0.5$ on i and spreads the rest uniformly over its neighbors, and d is the hop distance on the unweighted graph. For a node i, the ORC is defined as the mean of the ORC values of its incident edges, $\kappa(i) = \frac{1}{\deg(i)} \sum_{j \in \mathcal{N}(i)} \kappa(i, j)$. Implementation details are in § B.1.

3 Curvature Diagnostic for Graph Construction

The framework answers three questions.

(Q1) Separation: do a construction's curvature distributions separate labeled populations of point clouds, and where in the graph? **(Q2) Size and composition:** how much of that separation survives at fixed point count and composition, and for which constructions? **(Q3) Predictive value:** does adding curvature improve a model that already sees the points, does the gain follow the conditioned separation, and is what the model extracts structure or degree?

Q1 and Q2 need no training (§ 3.1, § 3.2; thresholds, bins, and bootstrap in Appendix C), and Q3 uses downstream models (§ 3.3). Q2 is the diagnostic assessment: Q1 alone cannot distinguish structure from size, and Q3 measures the model as much as the construction.

Setup. A point cloud X has n points. A construction $g \in \mathcal{G}$ builds a graph from X (duplicates removed, § A.1), and $\kappa \in \{\mathrm{FRC}, \mathrm{ORC}\}$ assigns per-node and per-edge values (§ 2.2, § B.1). A *population* f is a labeled set of point clouds, nodes, or edges; a *structural subset* s is the part of the graph on which curvature is read. For jets, s ranges over all nodes, all edges, PV and SV nodes, and PV-PV, SV-SV, PV-SV edges, with flavors $\{\mathrm{light}, c, b, \tau\}$ as populations and a PV vs SV comparison within each flavor; for QM9 the populations are the six pairs(§ 4.1). $\hat{\mu}_{s,\kappa}^{g,f}$ is the empirical curvature distribution of f within s under g, pooled over all values (*union*) or with one mean per graph (*per-graph mean*). We compare distributions with the Wasserstein-1 distance W_1, which, outside the ORC definition, acts on scalar curvature samples, where it is the area between the two empirical CDFs (Peyré & Cuturi, 2019, Sec. 2.6), in curvature units (§ C.1).

3.1 Population Separability

For a construction g, structural subset s, curvature κ, and a set $\mathcal{P}$ of population pairs (f_A, f_B), the *separation* score is the average W_1 between paired populations,

$$S_{\text{sep}}(g; s, \kappa) = \frac{1}{|\mathcal{P}|} \sum_{(f_A, f_B) \in \mathcal{P}} W_1(\hat{\mu}^{g,f_A}_{s,\kappa}, \hat{\mu}^{g,f_B}_{s,\kappa}). \tag{4}$$

For jets, $\mathcal{P} = \{(b, \text{light}), (c, \text{light}), (\tau, \text{light})\}$, since tagging is a signal vs background task and heavy-heavy pairs would dilute the score. For the within-flavor subsets $\mathcal{P}$ is the single pair (PV, SV). For QM9 each pair is scored on its own. We also report every pair term, since pairs separate by very different amounts. Raw FRC scales with degree and is not comparable across constructions, whereas ORC is dimensionless, so we compare constructions by ORC separation and by the dimensionless ratio of Equation 6. Two caveats apply. Empirical W_1 is biased upward at small sample sizes, so every distribution must hold at least $N_{\text{val}} = 50$ values (Bobkov & Ledoux, 2019) (§ C.1). And $S_{\text{sep}} > 0$ shows only statistical dependence: on the FC, curvature is a function of n alone, so FC separates populations of different sizes while carrying no structure.

3.2 Conditioning on size and composition

Curvature depends on how many points a cloud has and how they are split among its labeled parts, so two populations that differ only in size have different curvature distributions. To measure how much of a separation is due to a variable c, we hold c fixed and bin point clouds on c, compare populations only within a bin, and recombine within-bin results (Cochran, 1968). For jets $c \in \{n, n_{\text{sv}}, r = n_{\text{sv}}/n, (n, n_{\text{sv}})\}$. For QM9 $c \in \{n, n_{\text{H}}, n_{\text{heavy}}, r^{\text{H}}, (n_{\text{heavy}}, n_{\text{H}})\}$. Because c is computed from the points, a point cloud falls in the same bin under every construction. Let $\mathcal{B}_{g,p}$ be the bins in which pair $p = (f_A, f_B)$ passes the occupancy thresholds of § C.2, $\mathcal{X}_c$ the point clouds that fall in one (for $c = n$, 99% of the sample), $\hat{\mu}^{g,f}(b)$ the distribution restricted to bin b, and $w_{p,b}$ the number of point clouds of f_A or f_B in b. The binned separation of p is the occupancy-weighted mean of its within-bin separations,

$$S^{\text{binned}}_{\text{sep}}(g, p; c) = \frac{\sum_{b \in \mathcal{B}_{g,p}} w_{p,b}\, W_1\big(\hat{\mu}^{g,f_A}(b), \hat{\mu}^{g,f_B}(b)\big)}{\sum_{b \in \mathcal{B}_{g,p}} w_{p,b}}, \tag{5}$$

averaged over pairs as in Equation 4. Weights count point clouds, so a bin gains nothing from its point clouds having more nodes. The pooled separation $S^{\text{pooled}}_{\text{sep}}(g; c)$ is Equation 4 on the same point clouds $\mathcal{X}_c$ with the bins ignored, and their ratio

$$\text{preserved-sep}_c(g) = S^{\text{binned}}_{\text{sep}}(g; c) \,/\, S^{\text{pooled}}_{\text{sep}}(g; c) \tag{6}$$

where preserved-sep$_c$ is the fraction of a construction's separation preserved at fixed c. Near one, the separation comes from something other than c, whereas, near zero, it was c in disguise. Finite bin width and the small-sample bias of W_1 keep the ratio of a structure free construction above zero, so every value is read against the FC. Values above one and further interpretation are discussed in § C.2. Coverage, the fraction of point clouds in retained bins, is reported with every value, and a worked example is in § C.4. Every score carries a 95% bootstrap interval from $B = 500$ (§ C.3), which is unreliable when conditioning removes the separation (Sommerfeld & Munk, 2018).

3.3 Incremental predictive value

For a model class $\mathcal{M}$ trained on input features X, and curvature κ_g computed under construction g, the incremental predictive value of curvature is

$$\Delta\text{Performance}(g; \mathcal{M}) = \text{Performance}\big(\mathcal{M};\, X, \kappa_g\big) - \text{Performance}\big(\mathcal{M};\, X\big), \tag{7}$$

that is, the change in performance from concatenating curvature to the inputs. Performance is AUROC for classification and $-$MAE for regression, so positive always means curvature helped. We use six model classes with node ORC and FRC as two extra channels. Results are averaged over five seeds with bootstrap uncertainties (§ 4.2).

ΔPerformance **measures extraction, not information** Since $\kappa_g = h(g(X))$ and no step uses the label Y, $Y \to X \to \kappa_g$ is a Markov chain with κ_g deterministic in X, so $I(\kappa_g; Y \mid X) = 0$, equivalently $P(Y \mid X, \kappa_g) = P(Y \mid X)$ (Cover & Thomas, 2006, Thm. 2.8.1). Curvature therefore cannot improve the optimal predictor from X: a positive ΔPerformance measures structure already present in X that the model failed to extract, as in the probing literature (Hewitt & Liang, 2019; Pimentel et al., 2020), which reconciles $S_{\text{sep}} > 0$ with ΔPerformance ≈ 0. With finite training data curvature can still help by supplying structure the model would need far more samples to find, and such a gain is still extraction. The identity applies to the transformer and the GNN, which receive X; the summary-statistic models see statistics of X from which curvature cannot be computed, so for them a gain may also reflect missing information. For the GNN a gain is expected because message passing cannot count the short cycles curvature depends on (§ 1) and because mean aggregation makes it nearly invariant to degree, which the degree control of § 5.3 tests.

4 Experimental Setup

4.1 Datasets

QM9 comes with a ground-truth graph, the chemical bond graph, which gives a positive control on the constructions (§ F.1); JetSet does not. No label described below is ever used as an input feature.

QM9. Each of the 133, 885 QM9 molecules has up to nine heavy (non-hydrogen) atoms drawn from C, N, O, and F, together with hydrogen atoms bonded to them. For each molecule, quantum chemistry calculations provide relaxed 3D coordinates, a Mulliken partial charge per atom, and molecular properties, of which we predict two: the HOMO–LUMO gap in Hartree and the dipole moment μ in Debye, as in the QM9 release; every MAE carries its target's unit (Table 6). No bond list is stored so connectivity must be inferred from atomic positions alone, which is exactly what the constructions of § 2.1 do. For the edge recovery control we take bonds from the MoleculeNet release (Wu et al., 2018) and verify them against the coordinate-free SMILES strings, which removes 4, 451 molecules (3.3%) and leaves 129, 434 (§ E.1). 103, 547 are used for training and 12, 943 for testing. Every construction of Table 1 is built for every molecule, and none is left without edges. Molecules with fewer than 29 atoms are zero-padded for the transformer and GNN, with padded atoms masked out. Six population pairs are read off the bond graph and the elements, independent of any construction: three atom partitions (hydrogen vs heavy, ring vs chain, terminal vs branch) and three edge partitions (H–H vs heavy–heavy, H–heavy vs heavy–heavy, bond vs non-bond; § D.2). We lean hardest on hydrogen vs heavy: a hydrogen has exactly one bond, so any construction imposing a minimum degree must fail on it (§ F.1), and it is the closest analog of JetSet's PV vs SV split.

JetSet. Each jet has up to 40 reconstructed tracks, each described by its pseudorapidity and azimuthal distance to the jet axis ($\Delta\eta$, $\Delta\phi$) and its transverse momentum p_T (Table 7), zero-padded for the per-track models. We discard jets with fewer than three tracks, keep only jets on which every construction yields at least one edge, and subsample with a fixed seed to 300,000 training jets (148, 853 light, 117, 043 b, 26,759 c, 7, 345 τ) and 50,000 held-out jets (24, 929 light, 19, 499 b, 4, 416 c, 1, 156 τ). All jet results use this set; separation uses the training jets. Every track carries a PV/SV origin label and every jet a flavor label, used only to group curvature distributions.

4.2 Models and training

Every model class $\mathcal{M}$ of § 3.3 is trained with and without the curvature features and scored on the same held-out point clouds. The summary-statistic and transformer baselines do not depend on the graph and are trained once per seed; the GNN's baseline is retrained per construction.

The four summary-statistic models (logistic or ridge regression, one-layer and three-layer MLPs, gradient-boosted trees) get a single fixed-length vector per point cloud, ten summary statistics of each per-point variable plus the point count (31 features for jets), with the same statistics of per-point ORC and FRC as 20 curvature features.(§ D.3). The transformer is a two-layer encoder with four heads over the padded points, mean-pooled over the real points, and it never sees the graph. The GNN is a three-layer message-passing network over the constructed graph that averages incom-

ing messages and mean-pools node states; this mean aggregation matters for § 5.3. Every model is trained with five seeds, evaluated once on the held-out set with 50 bootstrap resamples, and reported as the mean over seeds, with seed and bootstrap uncertainties combined in quadrature The architectures, hyperparameters, and early stopping are in Table 8 of § D.3.

5 RESULTS

Further diagnostic scores are in Appendix F and full predictive value tables in Appendix G.

5.1 POSITIVE CONTROL ON QM9

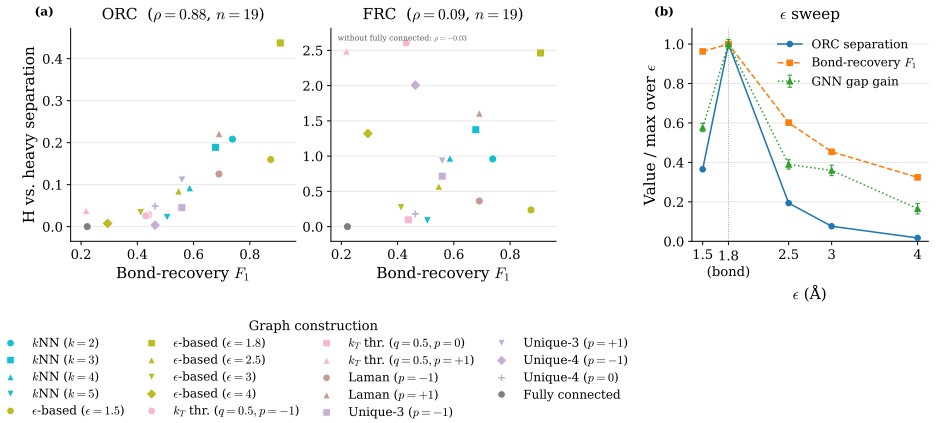


Figure 1: Curvature separation tracks bond recovery on QM9 without seeing the bond graph. (a) Hydrogen vs heavy separation S_{sep} (per molecule mean), against bond recovery F1 for 19 constructions under ORC (left) and FRC (right), with Spearman ρ in each title. Color and marker shape are the construction family. The FC sits at zero separation under both. (b) The ϵ-based sweep: ORC separation, F1, and the GNN gain on the HOMO-LUMO gap (Equation 7), each normalized; the dotted line marks $\epsilon = 1.8$ Å the bond-like radius; error bars are the combined seed and bootstrap σ of § 4.2.

We first test curvature where the answer is known, scoring each construction against the bond graph (Table 9). The 1.8 Å ϵ-based recovers every bond at a precision of 0.836, and every error joins two hydrogens on the same heavy atom. kNN falls short (precision 0.530 at $k = 3$) for a structural reason: a hydrogen has exactly one bond, so a minimum degree of k forces $k - 1$ non-bonds on every hydrogen and bounds precision by $\bar{d}/k \approx 2/3$ (§ F.1). The rigidity constructions reach exactly $1/k$ on hydrogen edges, so their $p = 0, +1$ variants are indistinguishable on edge recovery.

Without seeing the bonds, ORC hydrogen vs heavy separation orders the constructions as F1 does (Spearman $\rho = 0.88$, $p < 10^{-6}$, with one mean per molecule; 0.87 union; Figure 1a). Within a family it decreases as non-bond edges are added (0.437 at 1.8 Å; 0.085, 0.033, 0.008 at 2.5, 3.0, 4.0 Å; 0.21, 0.19, 0.09, 0.02 for kNN at $k = 2$–5) and as bonds go missing (0.160 at 1.5 Å). It also resolves what edge recovery cannot: Laman's $p = 0, +1$ variants have identical precision and recall but ORC separation 0.22 ($p = 1$) against 0.13 ($p = -1$), because $p = -1$ attaches a hydrogen's non-bond edges to heavy atoms and $p = 0, +1$ to other hydrogens. ORC tracks F1 on five of the six pairs of Appendix D ($\rho = 0.76$–0.92). The exception is bond vs non-bond ($\rho = 0.28$; Table 14). FRC tracks recovery on no node pair ($|\rho| \leq 0.13$), and its hydrogen vs heavy separation peaks on constructions that recover bonds poorly (Table 12). Both populations of every QM9 pair share the same molecules, so conditioning on size moves the per-molecule-mean scores by under 3% except at the W_1 bias floor; the FC keeps 0.15 (Figure 5). The size confound arises only when populations are different point clouds, as on jets.

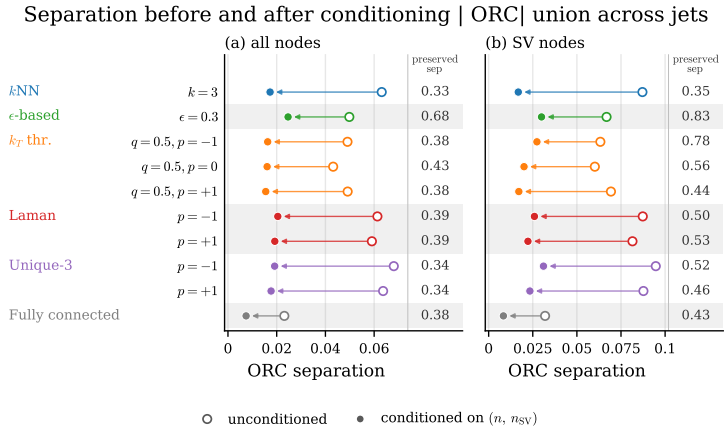


Figure 2: Flavor separation before and after conditioning. Each arrow runs from the pooled separation (open circle) to the binned separation at fixed (n, n_{sv}) (Equation 5), averaged over the three flavor pairs. The right column is their ratio, preserved-sep$_{(n,n_{\mathrm{sv}})}(g)$ (Equation 6). (a) ORC over all nodes, (b) ORC over SV nodes. Rows: default constructions by family. The FC row marks the level left when no separation survives. Intervals are 95% bootstrap; bins and coverage are in § C.2.

5.2 Separation on jets, before and after conditioning

Before conditioning, every construction separates the flavors. Across all 33 constructions, mean ORC separation over all nodes ranges from 0.031 (ϵ-based, $\epsilon = 0.5$) to 0.068 (unique-3, $p = -1$), against 0.023 for the FC (Table 15). The mean is dominated by τ vs light, the smallest pair and the one the complete graph separates most (for kNN: 0.030 b, 0.006 c, 0.153 τ; Figure 7). Pair by pair, the ϵ-based construction ties the best on b vs light (0.034, while unique-3 0.036, kNN 0.030) and stands alone on c vs light (0.028 against at most 0.006). The rigid and kNN constructions lead only on τ vs light.

After conditioning, only the ϵ-based construction keeps its separation. At fixed (n, n_{sv}) (63% of jets, Figure 2a), the FC preserves 0.38 of its all-node ORC separation, and kNN (0.33), the k_T thresholds (0.38–0.43), Laman (0.39), and unique-3 (0.34) sit at that control level. The ϵ-based construction preserves 0.68. SV nodes agree (0.83 against 0.35–0.56, control 0.43; panel b), except that k_T at $p = -1$ preserves 0.78. The effect belongs to the family, not a tuned radius: all four ϵ values preserve 0.64–0.68, and none of the other 28 constructions exceeds 0.47 at any k, q, or p (Figure 16). The lead holds within every multiplicity bin, by at least 2× on b vs light and 3× on c vs light (Figure 15). On τ vs light, Laman and k_T lead only at $n \leq 5$, where the multiplicity gap is largest.

Conditioning on n alone gives the same ordering (Table 19): the FC preserves 0.18–0.21 on all nodes and node subsets, kNN, unique-3, and Laman ($p = -1$) sit at that floor (0.15–0.30), k_T and Laman ($p = 0, 1$) slightly above it (up to 0.40), and the ϵ-based preserves 0.64 on all nodes and 0.87 on PV nodes. Edge subsets are noisier (FC: 0.33 on PV-PV edges), so we read the diagnostic on nodes. Conditioning on n_{sv} or n_{sv}/n alone pushes preserved-sep above one for almost every construction (Figure 25), because flavors matched on n_{sv} differ more in n, so these variables are informative only jointly with n.

FRC (Figure 9) depends only on degrees and triangle counts, so at fixed size it survives only where degree follows point positions rather than a rule: the ϵ-based construction (0.66 all nodes, 1.07 SV nodes) and the k_T thresholds (0.46–0.70, 0.87–0.94). Laman, unique-3, and kNN stay at the control level (0.29–0.39 against 0.35). Per-jet means leave every ordering unchanged (§ F.8). The unconditioned ranking, with unique-3 and kNN first, is a ranking by track count; conditioned, only the ϵ-based construction stays above the control.

5.3 Q3: Predictive value of curvature

QM9. Models gain most on the construction the diagnostic selects: the 1.8 Å graph gives the largest gain for all six model classes and both targets, and the complete graph gains less than 0.001 Ha on the gap (Table 21). With that curvature, the transformer, which never sees a graph, beats the FC GNN on both targets (0.0095 vs. 0.0159 Ha for the gap, 0.448 vs. 0.497 D for μ), and the GNN on that graph halves its gap error although it already has the edges, so the gain is extraction. For the gap no gain is significantly negative, and along the ϵ sweep gain and ORC separation both fall away from 1.8 Å (Figure 1b). Across families the match is looser (Laman at $p = 1$: second highest separation, modest gain), so we read the diagnostic within a sweep. For μ, the GNN gains only at 1.5 and 1.8 Å the two rows with F1$\geq$ 0.87, and is significantly worse with curvature on 14 of the other 20 rows: separation is necessary for a gain but not sufficient. We attribute this to locality: curvature on a bond-like graph encodes the bonding the gap depends on, not the long-range geometry of μ.

JetSet. Gains are small (Table 20). On b and τ vs light, both per-track models always gain: on b vs light by +0.009 to +0.014 (transformer) and +0.005 to +0.017 (GNN), on τ vs light by up to +0.026 and +0.017, and on c vs light by at most +0.003. Among the defaults, the GNN gain runs against the diagnostic, smallest on the FC (+0.005) and the ϵ-based (+0.007) and largest on Laman and kNN (+0.014 to +0.017), while the baselines run with it: the ϵ-based is the best sparse construction (0.715 against 0.704–0.713) and the FC the best overall (0.720). The transformer gains +0.013 even from FC curvature, a function of n alone: the models extract mostly degree and multiplicity.

Degree control. Both curvatures depend on degree, FRC exactly (Equation 2) and ORC through its neighborhood measures, but the mean-aggregating GNN is nearly blind to degree and to n. We therefore add each node's degree under g as the only extra channel (eight constructions; Table 37) and compare with curvature at the same learning rate, 3×10^{-4} (Table 33). On the four constructions run both ways, degree alone recovers all of the GNN's curvature gain on the ϵ-based, 92% on k_T, 86% on Laman, and 69% on kNN, whose minimum degree is fixed at k. The transformer agrees gaining +0.0165 even from FC curvature at this rate, a track count that mean pooling cannot form.

Does separation predict the gain? No. Across the nine sparse defaults and three pairs, separation at fixed n is uncorrelated with the GNN's gain (ρ_s = +0.04 for ORC, −0.28 for FRC) and the transformer's (+0.14, −0.12), and likewise at fixed (n, n_{sv}) (§ G.8), consistent with the degree control. Separation does line up with the curvature-free GNN baseline (ρ_s = +0.57, FRC, fixed (n, n_{sv}); +0.55 without the ϵ-based construction) (§ G.8).

5.4 Limitations

Conditioning on n_{sv} uses a truth-level label, so conditioning on n alone is what a model faces. Empirical W_1 is biased upward in thin bins, most under joint conditioning, and the entropic solver biases ORC upward on high degree nodes. Jet effects are small ($\leq$ 0.03 AUROC, five seeds). The GNN uses mean aggregation and no distance features, and the degree control covers eight constructions at one learning rate, so a sum aggregating or degree-aware GNN would be a stronger probe. We do not score degree distributions directly; FRC, set by degrees and triangles, fails to track bond recovery where ORC succeeds, suggesting ORC's agreement is not a degree effect. Without impact parameters, our tagging is far below the state of the art.

6 Conclusion

We draw three conclusions. i) The unconditioned curvature score is mostly a size detector — entirely on the FC and nearly on kNN, k_T, Laman, and unique-k — and conditioning reverses the comparison. ii) Without training, conditioned ORC (not FRC) separation selects one construction per domain: the 1.8 Å radius graph on QM9, which also maximizes bond recovery, and the ϵ-based graph on jets, which also gives the best sparse GNN baseline. iii) Curvature feature gains are extraction, on jets mostly degree and multiplicity; separation is necessary for a gain but does not predict it, so we compare constructions by conditioned separation, not ΔPerformance. The diagnostic costs no training, needs no ground truth, and runs on graphs of tens of nodes, outside the regime where asymptotic theory applies. In practice, degree aware models (sum aggregation or explicit degree features) should be tried before curvature features, and no curvature descriptor should be read as structure without size control.

AI USE STATEMENT

In this work, we used generative AI tools for tasks with required disclosure. To implement methods and reformat data, an LLM coding assistant adapted our jet graph-construction code to the QM9 data format, wrote the parallel implementation of the W_1 separation, and changed the GNN to pass messages in both directions along each edge. To design experiments, an LLM helped choose the QM9 structural subsets that correspond to the jet subsets, and AI-generated pre-submission reviews of drafts gave feedback on our methodology and experiments. We have not used generative AI tools to generate synthetic data, develop theoretical models or conceptual frameworks, formulate mathematical claims, propose or refine hypotheses, or interpret results, and proofs, translation, and qualitative or thematic data analysis are not applicable to this work.

Additionally, we used generative AI tools for tasks with recommended disclosure: to identify literature on graph curvature, explain the QM9 dataset, restructure analysis code for memory efficiency and minimize code duplication for the two datasets, debug and document code, write and restyle plotting scripts, generate LaTeX tables from results files, edit the paper for readability and suggest changes to the structure of the introduction, format references, and draft this statement. We also used them for project logistics: syncing our git repository with our online LaTeX editor, resolving merge conflicts, and fixing build errors. The tools were Claude (Anthropic), Qwen 3.8 served locally, and Google Scholar Labs.

We have reviewed all AI-assisted work. We logged our AI use as the work progressed, and this statement is built from those logs. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work uses two public scientific datasets, QM9 (Ramakrishnan et al., 2014) and the ATLAS JetSet open simulation (ATLAS Collaboration, 2025), and does not involve personal data.

REPRODUCIBILITY STATEMENT

Every construction and parameter setting is listed in Appendix A, and the curvature implementation in Appendix B. The diagnostic is defined in § 3, with thresholds, bins, bootstrap procedure, and a worked example in Appendix C. Data selection is in § 4, bond graph verification in Appendix E, model architectures, hyperparameters, seeds, and training protocol in § D.3, and compute in § D.4. All results are over five fixed seeds with the uncertainties described there. Code to build the constructions, compute curvatures and separation, and produce every table and figure is at the anonymized link: https://anonymous.4open.science/r/ricci-graph-construction-diagnostic-20EB/README.md

REFERENCES

Jack Y. Araz, Dimitrios Athanasakos, Mateusz Ploskon, and Felix Ringer. Graph theory inspired anomaly detection at the LHC. *Journal of High Energy Physics*, 2026(2):254, 2026. doi: 10.1007/JHEP02(2026)254.

ATLAS Collaboration. Performance of b-jet identification in the ATLAS experiment. *Journal of Instrumentation*, 11(04):P04008, 2016. doi: 10.1088/1748-0221/11/04/P04008.

ATLAS Collaboration. ATLAS $t\bar{t}$ simulation for ML-based jet flavour tagging (JetSet). CERN Open Data Portal, 2025. URL https://doi.org/10.7483/OPENDATA.ATLAS.QG8W.TO8P.

Frank Bauer, Bobo Hua, Jürgen Jost, and Shiping Liu. Generalized Ricci curvature and the geometry of graphs. *Actes des rencontres du CIRM*, 3(1):69–78, 2013. doi: 10.5802/acirm.56.

Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013. doi: 10.1109/TPAMI.2013.50.

Jyotiranjan Beuria. Intrinsic geometry of collider observations and Forman Ricci curvature. *Physical Review D*, 110(3):035023, 2024. doi: 10.1103/PhysRevD.110.035023.

Sergey Bobkov and Michel Ledoux. One-dimensional empirical measures, order statistics, and Kantorovich transport distances. *Memoirs of the American Mathematical Society*, 261(1259), 2019. doi: 10.1090/memo/1259.

Giorgos Bouritsas, Fabrizio Frasca, Stefanos Zafeiriou, and Michael M. Bronstein. Improving graph neural network expressivity via subgraph isomorphism counting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):657–668, 2023. doi: 10.1109/TPAMI.2022.3154319.

Michael M. Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4): 18–42, 2017.

Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063, 2008. doi: 10.1088/1126-6708/2008/04/063.

Jeff Calder and Nicolás García Trillos. Improved spectral convergence rates for graph Laplacians on ε-graphs and k-NN graphs. *Applied and Computational Harmonic Analysis*, 60:123–175, 2022. doi: 10.1016/j.acha.2022.02.004.

Zhengdao Chen, Lei Chen, Soledad Villar, and Joan Bruna. Can graph neural networks count substructures? In *Advances in Neural Information Processing Systems*, volume 33, pp. 10383–10395, 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/hash/75877cb75154206c4e65e76b88a12712-Abstract.html.

William G. Cochran. The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics*, 24(2):295–313, 1968. doi: 10.2307/2528036.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley and Sons, second edition, 2006. doi: 10.1002/047174882X.

Manlio De Domenico, Albert Solé-Ribalta, Emanuele Cozzo, Mikko Kivelä, Yamir Moreno, Mason A. Porter, Sergio Gómez, and Alex Arenas. Mathematical formulation of multilayer networks. *Physical Review X*, 3(4):041022, 2013. doi: 10.1103/PhysRevX.3.041022.

Lukas Fesser and Melanie Weber. Mitigating over-smoothing and over-squashing using augmentations of Forman-Ricci curvature. In *Proceedings of the Second Learning on Graphs Conference*, volume 231 of *Proceedings of Machine Learning Research*, pp. 19:1–19:28. PMLR, 2024. URL https://proceedings.mlr.press/v231/fesser24a.html.

Matthias Fey and Jan Eric Lenssen. Fast Graph Representation Learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019. URL https://arxiv.org/abs/1903.02428.

Christopher A. Field and Alan H. Welsh. Bootstrapping clustered data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 69(3):369–390, 2007. doi: 10.1111/j.1467-9868.2007.00593.x.

Robin Forman. Bochner's method for cell complexes and combinatorial Ricci curvature. *Discrete & Computational Geometry*, 29(3):323–374, 2003. doi: 10.1007/s00454-002-0743-x.

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

Tejes Gaertner and Jared Reiten. The simplicial substructure of jets. *arXiv preprint arXiv:2312.06596*, 2023. URL https://arxiv.org/abs/2312.06596.

John Hewitt and Percy Liang. Designing and interpreting probes with control tasks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 2733–2743, 2019. doi: 10.18653/v1/D19-1275.

Anees Kazi, Luca Cosmo, Seyed-Ahmad Ahmadi, Nassir Navab, and Michael M. Bronstein. Differentiable graph module (DGM) for graph convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2):1606–1617, 2023. doi: 10.1109/TPAMI.2022.3170249.

Roman Kogler, Benjamin Nachman, Alexander Schmidt, Lily Asquith, Emma Winkels, Mario Campanelli, Chris Delitzsch, Philip Harris, Andreas Hinzmann, Deepak Kar, Christine McLean, Justin Pilot, Yuta Takahashi, Nhan Tran, Caterina Vernieri, and Marcel Vos. Jet substructure at the Large Hadron Collider. *Rev. Mod. Phys.*, 91:045003, Dec 2019. doi: 10.1103/RevModPhys.91.045003.

Yong Lin, Linyuan Lu, and Shing-Tung Yau. Ricci curvature of graphs. *Tohoku Mathematical Journal, Second Series*, 63(4):605–627, 2011. doi: 10.2748/tmj/1325886283.

Markus Maier, Ulrike von Luxburg, and Matthias Hein. Influence of graph construction on graph-based clustering measures. In *Advances in Neural Information Processing Systems*, volume 21, 2008. URL https://proceedings.neurips.cc/paper_files/paper/2008/hash/0a348ede8ac3768875037baca5de6e26-Abstract.html.

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 *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 25956–25979. PMLR, 2023. URL https://proceedings.mlr.press/v202/nguyen23c.html.

Chien-Chun Ni, Yu-Yao Lin, Feng Luo, and Jie Gao. Community detection on networks with Ricci flow. *Scientific Reports*, 9(1):9984, 2019. doi: 10.1038/s41598-019-46380-9.

Leslie O'Bray, Max Horn, Bastian Rieck, and Karsten M. Borgwardt. Evaluation metrics for graph generative models: Problems, pitfalls, and practical solutions. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=tBtoZYKd9n.

Gabriel Peyré and Marco Cuturi. Computational Optimal Transport. *Foundations and Trends in Machine Learning*, 11(5–6):355–607, 2019. doi: 10.1561/2200000073.

Tiago Pimentel, Josef Valvoda, Rowan Hall Maudslay, Ran Zmigrod, Adina Williams, and Ryan Cotterell. Information-theoretic probing for linguistic structure. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 4609–4622, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.420.

Xingyue Pu, Tianyue Cao, Xiaoyun Zhang, Xiaowen Dong, and Siheng Chen. Learning to learn graph topologies. In *Advances in Neural Information Processing Systems*, volume 34, pp. 4249–4262, 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/hash/21e4ef94f2a6b23597efabaec584b504-Abstract.html.

Huilin Qu and Loukas Gouskos. Jet tagging via particle clouds. *Physical Review D*, 101(5):056019, 2020. doi: 10.1103/PhysRevD.101.056019.

Raghunathan Ramakrishnan, Pavlo O. Dral, Matthias Rupp, and O. Anatole von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data*, 1(1):140022, 2014. doi: 10.1038/sdata.2014.22.

Areejit Samal, R. P. Sreejith, Jiao Gu, Shiping Liu, Emil Saucan, and Jürgen Jost. Comparative analysis of two discretizations of Ricci curvature for complex networks. *Scientific Reports*, 8(1):8650, 2018. doi: 10.1038/s41598-018-27001-3.

Jonathan Shlomi, Peter Battaglia, and Jean-Roch Vlimant. Graph neural networks in particle physics. *Machine Learning: Science and Technology*, 2(2):021001, 2021. doi: 10.1088/2632-2153/abbf9a.

Max Sommerfeld and Axel Munk. Inference for Empirical Wasserstein Distances on Finite Spaces. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 80(1):219–238, 2018. doi: 10.1111/rssb.12236.

Joshua Southern, Jeremy Wayland, Michael Bronstein, and Bastian Rieck. Curvature filtrations for graph generative model evaluation. In *Advances in Neural Information Processing Systems*, volume 36, pp. 63036–63061, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/c710d6b4507e70c6332bee871b8d1ca5-Abstract-Conference.html.

Remanan Pushpa Sreejith, Karthikeyan Mohanraj, Jürgen Jost, Emil Saucan, and Areejit Samal. Forman curvature for complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2016(6):063206, 2016. doi: 10.1088/1742-5468/2016/06/063206.

Michael Steinbach and Pang-Ning Tan. kNN: k-nearest neighbors. In Xindong Wu and Vipin Kumar (eds.), *The Top Ten Algorithms in Data Mining*, pp. 165–176. Chapman and Hall/CRC, 2009. doi: 10.1201/9781420089653-15.

Rylee Thompson, Boris Knyazev, Elahe Ghalebi, Jungtaek Kim, and Graham W. Taylor. On evaluation metrics for graph generative models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=EnwCZixjSh.

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. URL https://openreview.net/forum?id=7UmjRGzp-A.

Floriano Tori, Vincent Holst, and Vincent Ginis. The effectiveness of curvature-based rewiring and the role of hyperparameters in GNNs revisited. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=EcrdmRT99M.

Pim van der Hoorn, William J. Cunningham, Gabor Lippner, Carlo Trugenberger, and Dmitri Krioukov. Ollivier-Ricci curvature convergence in random geometric graphs. *Physical Review Research*, 3(1):013211, 2021. doi: 10.1103/PhysRevResearch.3.013211.

Pim van der Hoorn, Gabor Lippner, Carlo Trugenberger, and Dmitri Krioukov. Ollivier curvature of random geometric graphs converges to Ricci curvature of their Riemannian manifolds. *Discrete & Computational Geometry*, 70(3):671–712, 2023. doi: 10.1007/s00454-023-00507-y.

Cédric Villani. The Wasserstein Distances. In *Optimal Transport: Old and New*, pp. 93–111. Springer, 2009. doi: 10.1007/978-3-540-71050-9_6.

Melanie Weber, Emil Saucan, and Jürgen Jost. Characterizing complex networks with Forman-Ricci curvature and associated geometric flows. *Journal of Complex Networks*, 5(4):527–550, 2017. doi: 10.1093/comnet/cnw030.

Zhenqin Wu, Bharath Ramsundar, Evan N. Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S. Pappu, Karl Leswing, and Vijay Pande. MoleculeNet: A Benchmark for Molecular Machine Learning. *Chemical Science*, 9(2):513–530, 2018. doi: 10.1039/C7SC02664A.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=ryGs6iA5Km.

Yilun Xu, Shengjia Zhao, Jiaming Song, Russell Stewart, and Stefano Ermon. A theory of usable information under computational constraints. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=r1eBeyHFDH.

APPENDIX: TABLE OF CONTENTS

A Constructions Built on Each Dataset . . . 15
A.1 Duplicate constructions . . . 15

B Curvature . . . 16
B.1 Implementation . . . 16

C Details of the Diagnostic . . . 16
C.1 Finite-sample bias and occupancy thresholds . . . 16
C.2 Conditioning on size and composition . . . 18
C.3 Uncertainty estimates . . . 21
C.4 A worked example of the distributional measures . . . 23

D Data and Model Details . . . 25
D.1 Input variables and targets . . . 25
D.2 Structural subsets of QM9 . . . 25
D.3 Model architectures and training protocol . . . 26
D.4 Compute . . . 27

E QM9 Reference Bond Graph . . . 27
E.1 Verification procedure . . . 27

F Additional Results: Diagnostic Scores . . . 28
F.1 Positive control: edge recovery on QM9 . . . 28
F.2 Edge recovery for all constructions . . . 30
F.3 Separation for QM9 population pairs . . . 32
F.4 Separation figures for JetSet . . . 35
F.5 Separation: union aggregation across jets per bin . . . 39
F.6 Separation: mean per jet aggregation across jets per bin . . . 39
F.7 Separation: union aggregation across jets . . . 40
F.8 Separation: mean per jet aggregation . . . 45
F.9 Separation tables for JetSet . . . 50
F.10 Separation preserved by conditioning on n . . . 54

G Predictive Value: Full Results . . . 55
G.1 QM9 . . . 56
G.2 JetSet: b vs light . . . 61
G.3 JetSet: c vs light . . . 63
G.4 JetSet: τ vs light . . . 65
G.5 b vs light and c vs light with learning rate of 3×10^{-4} . . . 67
G.6 Degree control . . . 68
G.7 Track count control . . . 70
G.8 Does separation predict the gain? . . . 71

A Constructions Built on Each Dataset

Table 1 lists every construction that was built, with the default settings in bold. The two grids differ, which is why the passes cover 36 constructions on JetSet and the 22 of Table 10 on QM9. Both counts are taken as built. The duplicate removal of § A.1 collapses three pairs on each dataset, leaving 33 distinct graphs on JetSet and 19 on QM9. The QM9 dataset (Ramakrishnan et al., 2014)

Table 1: Constructions built on each dataset. Default settings are in bold, the remainder are the parameter sweep. All k_T threshold settings use $k_{\min} = 2$, and $k_{\max} = 7$ except where stated.

Family	Settings
JetSet	
kNN	$k = \mathbf{3}, 4, 5, 7$
ϵ-based	$\epsilon = 0.2, \mathbf{0.3}, 0.4, 0.5$
k_T threshold, $p = -1$	$q = 0.4, \mathbf{0.5}, 0.6, 0.7, 0.8, 0.9$; $q = 0.8$ with $k_{\max} = 5, 10$
k_T threshold, $p = +1$	$q = 0.4, \mathbf{0.5}, 0.8, 0.9$
k_T threshold, $p = 0$	$q = 0.4, \mathbf{0.5}, 0.8, 0.9$
Laman	$p = \mathbf{+1}, \mathbf{0}, \mathbf{-1}$
unique-k	$k = \mathbf{3}$ with $p = \mathbf{+1}, \mathbf{0}, \mathbf{-1}$; $k = 4$ with $p = +1, 0, -1$; $k = 5$ with $p = 0, -1$
FC	–
QM9	
kNN	$k = 2, \mathbf{3}, 4, 5$
ϵ-based	$\epsilon = 1.5, \mathbf{1.8}, 2.5, \mathbf{3.0}, 4.0$
k_T threshold	$q = \mathbf{0.5}$ with $p = \mathbf{+1}, \mathbf{0}, \mathbf{-1}$
Laman	$p = \mathbf{+1}, \mathbf{0}, \mathbf{-1}$
unique-k	$k = 2$ with $p = -1$; $k = \mathbf{3}$ with $p = \mathbf{+1}, \mathbf{0}, \mathbf{-1}$; $k = 4$ with $p = 0, -1$
FC	–

contains 133,885 small organic molecules. Each has up to nine heavy (non hydrogen) atoms drawn from C, N, O, and F, together with hydrogen atoms bonded to them. For each molecule, quantum chemistry calculations provide relaxed 3D coordinates, a Mulliken partial charge per atom, and molecular properties, of which we predict two: the HOMO–LUMO gap in Hartree units and the dipole moment μ in Debye units throughout, as in the QM9 release, and every MAE we report carries the unit of its target (Table 6). No bond list is stored so connectivity must be inferred from atomic positions alone, which is exactly what the constructions of § 2.1 do. For the edge reconvery control we take bonds from the MoleculeNet (Wu et al., 2018) release (`gdb9.sdf`, via PyTorch Geometric (Fey & Lenssen, 2019)) and verify them against the SMILES strings in the original `xyz` files, which do not depend on 3D coordinates and so are independent of any distance rule, including the ϵ-based construction under evaluation. That check and the QM9 authors' own consistency flag remove 4,451 molecules (3.3%), leaving 129,434 (§ E.1). We shuffle once with a fixed seed and use 103,547 training and 12,943 test molecules, holding out the remainder. Every construction of Table 1 is built for every molecule, and none is left without edges. Molecules with fewer than 29 atoms are zero-padded for the transformer and GNN, with padded atoms masked out.

A.1 Duplicate constructions

Constructions were treated as duplicates if their unconditioned node-level curvature distributions had $W_1 = 0$ across every combination of flavor, node subset, curvature, and aggregation. One member of each pair was dropped from all analyses so that no graph appears twice. Among the jet rigidity constructions, $p = 0$ and $p = +1$ coincided under decreasing p_T insertion, and $p = 0$ was removed.

B CURVATURE

B.1 IMPLEMENTATION

In the ORC definition, Equation 3, μ_i and μ_j are probability measures over the neighbors of i and of j, $d(i, j)$ is the graph distance between the two nodes, and $W_1(\mu_i, \mu_j)$ is the Wasserstein-1 distance between the two measures.

Both curvatures are computed with the open-source `GraphRicciCurvature` package (Ni et al., 2019). The W_1 term of Equation 3 is evaluated with the package's entropic (Sinkhorn) solver at regularization $\varepsilon = 0.1$, via POT. Entropic optimal transport is biased relative to exact transport, and the bias grows with the size of the support, here $\deg(i) + \deg(j) + 2$. ORC uses idleness parameter $\alpha = 0.5$, so μ_i places mass α on i itself and distributes the remainder over its neighbors by the package's exponential kernel, which reduces to the uniform neighbor measure on our unit-weight graphs. All graphs are unweighted (every edge length is 1), so $d(i, j)$ is the shortest path distance in hop count. The package normalizes Equation 3 by the edge weight $w(i, j)$, which equals the distance between adjacent nodes. On unit-weight graphs this factor is 1, so the implementation and Equation 3 coincide here.

For FRC, Equation 2, we use the package's augmented form, which is the default and which reproduces Equation 2 exactly: the edges of K_n give n, and the middle edge of a path on four nodes gives 0.

C DETAILS OF THE DIAGNOSTIC

C.1 FINITE-SAMPLE BIAS AND OCCUPANCY THRESHOLDS

What a nonzero score does and does not mean. A nonzero S_{sep} tells us one thing: the curvature distributions of the two populations differ. In information-theoretic terms, curvature and the population label are statistically dependent, $I(K; Y) > 0$, where Y is the population label, K is the curvature value of a randomly drawn node or edge, and $I(\cdot\,;\cdot)$ is mutual information. Two caveats limit what that is worth, and they motivate the thresholds below and the conditioned scores of § C.2.

The empirical W_1 is biased upward. The empirical W_1 between two finite samples is positive even when the populations behind them are identical. Let μ be a law on $\mathbb{R}$ with distribution function F, and let $\hat{\mu}_m$ be the empirical distribution of m independent draws from it, with distribution function F_m. Bobkov & Ledoux (2019) show that $\mathbb{E}\, W_1(\hat{\mu}_m, \mu)$ decays at the standard rate $m^{-1/2}$ if and only if $J_1(F) = \int_{\mathbb{R}} \sqrt{F(t)\,(1 - F(t))}\, dt$ is finite, and they bound it from both sides in terms of F. The upper bound $\mathbb{E}\, W_1(\hat{\mu}_m, \mu) \le J_1(F)/\sqrt{m}$ follows from Jensen's inequality applied to $W_1(\hat{\mu}_m, \mu) = \int_{\mathbb{R}} |F_m(t) - F(t)|\, dt$. Two samples of sizes m_A and m_B drawn from μ inherit this rate: the triangle inequality and the convexity of W_1 in each argument give

$$\max_{k \in \{A,B\}} \mathbb{E}\, W_1(\hat{\mu}_{m_k}, \mu) \;\le\; \mathbb{E}\, W_1(\hat{\mu}_{m_A}, \hat{\mu}_{m_B}) \;\le\; \mathbb{E}\, W_1(\hat{\mu}_{m_A}, \mu) + \mathbb{E}\, W_1(\hat{\mu}_{m_B}, \mu), \tag{8}$$

so the bias is of order $J_1(F)\sqrt{1/m_A + 1/m_B}$, and the smaller sample controls it. If the values lie in an interval of length D, then $J_1(F) \le D/2$, because $F(1 - F) \le 1/4$ and the integrand vanishes outside the interval. Both curvatures are bounded here, so J_1 is finite.

The two thresholds. Occupancy and sample size count different things, so two thresholds apply. $N_{\min} = 200$ counts point clouds: population f enters bin b only if at least $N_{\min}$ of its point clouds fall in b (jets of flavor f for JetSet. Molecules in the bin for QM9, whose pair subsets share molecules). $N_{\text{val}} = 50$ counts values: a W_1 term is computed only if both distributions hold at least N_{val} values, one per node or edge under the union and one per graph containing the subset under the per-graph mean. $N_{\min}$ is the same for every construction, subset, and aggregation; N_{val} varies with subset and aggregation, and on edge subsets with g, so passing one does not imply passing the other. The counts are not proportional: a bin of 200 jets may hold roughly a thousand PV nodes but a dozen SV-SV edge values, and $N_{\min}$ light jets can yield fewer than N_{val} per-jet SV means. Without the thresholds, the sparsely populated tails at high n and high n_{sv} (Table 2) would show separation inflated by small samples rather than by graph structure.

Retained sets. Two sets record what survives in each bin: $\mathcal{F}_b$, the populations with at least $N_{\min}$ point clouds in b, and $\mathcal{P}_b$, the pairs in $\mathcal{F}_b \times \mathcal{F}_b$ whose two distributions each hold at least N_{val} values. These sets fix the terms of Equation 12 and reappear in the aggregation: $N_{\min}$ fixes $\mathcal{F}_b$ and with it the weight $w_{p,b}$, and the two thresholds together fix $\mathcal{B}_{g,p}$, the bins over which each pair is averaged. Table 2 reports the point clouds each population retains in every bin. The unconditioned score of Equation 4 is subject to N_{val}, since every distribution has a value count, but not to $N_{\min}$, which counts within a bin and so has no unconditioned counterpart. It runs over the full sample and the full pair set $\mathcal{P}$.

Table 2: Jet counts $N_{f,b}$ per flavor f and bin b for the three single variable conditioning passes. The counts are independent of the graph construction. Cells shaded gray and marked † fall below $N_{\min} = 200$, so that flavor is dropped from that bin. "Outside all bins" counts jets not assigned to any bin of that variable. "Retained fraction" is the fraction of each flavor's jets that lie in a bin where the flavor is kept. The joint conditioning is in Table 3

Bin	light	c	b	τ
Conditioning on n				
3–5	40 602	7 172	19 068	5 165
6–9	73 129	13 667	58 645	1 919
10–14	30 706	5 276	34 362	248
15–19	3 903	577	4 614	13†
20–30	502	65†	351	0†
31–40	11†	2†	3†	0†
Outside all bins	0	0	0	0
Retained fraction	0.9999	0.9975	>0.9999	0.9982
Conditioning on n_{sv}				
0	112 280	1 689	685	232
1–2	34 594	14 865	16 318	4 243
3–4	1 892	8 851	47 348	2 816
5–7	87†	1 317	44 361	53†
8–10	0†	36†	7 754	1†
11–15	0†	1†	570	0†
16–20	0†	0†	7†	0†
21–40	0†	0†	0†	0†
Outside all bins	0	0	0	0
Retained fraction	0.9994	0.9986	0.9999	0.9926

Table 3: This is continuation of Table 2 for joint conditioning.

Bin	light	c	b	τ
Conditioning jointly on (n, n_{sv})				
n = 3–5				
n_{sv} = 0	34 370	653	237	146†
n_{sv} = 1–2	6 147	5 091	6 813	3 019
n_{sv} = 3–4	85†	1 411	10 996	1 985
n_{sv} = 5–7	0†	17†	1 022	15†
n = 6–9				
n_{sv} = 0	55 532	808	330	78†
n_{sv} = 1–2	16 837	7 397	7 281	1 076
n_{sv} = 3–4	746	4 917	26 733	738
n_{sv} = 5–7	14†	541	23 316	27†
n_{sv} = 8–10	0†	4†	985	0†
n = 10–14				
n_{sv} = 0	20 072	217	105†	8†
n_{sv} = 1–2	9 797	2 189	2 003	143†
n_{sv} = 3–4	781	2 252	8 750	88†
n_{sv} = 5–7	56†	602	17 886	9†
n_{sv} = 8–10	0†	16†	5 366	0†
n_{sv} = 11–15	0†	0†	252	0†
n = 15–19				
n_{sv} = 0	2 105	10†	12†	0†
n_{sv} = 1–2	1 561	166†	206	5†
n_{sv} = 3–4	224	251	829	5†
n_{sv} = 5–7	13†	135†	1 994	2†
n_{sv} = 8–10	0†	14†	1 299	1†
n_{sv} = 11–15	0†	1†	272	0†
n_{sv} = 16–20	0†	0†	2†	0†
n = 20–30				
n_{sv} = 0	198†	1†	1†	0†
n_{sv} = 1–2	246	21†	15†	0†
n_{sv} = 3–4	54†	20†	39†	0†
n_{sv} = 5–7	4†	21†	143†	0†
n_{sv} = 8–10	0†	2†	102†	0†
n_{sv} = 11–15	0†	0†	46†	0†
n_{sv} = 16–20	0†	0†	5†	0†
n = 31–40				
n_{sv} = 0	3†	0†	0†	0†
n_{sv} = 1–2	6†	1†	0†	0†
n_{sv} = 3–4	2†	0†	1†	0†
n_{sv} = 5–7	0†	1†	0†	0†
n_{sv} = 8–10	0†	0†	2†	0†
Outside all bins	0	0	0	0
Retained fraction	0.9971	0.9839	0.9960	0.9282

C.2 Conditioning on size and composition

Throughout this subsection the structural subset s and the curvature type κ are fixed and dropped from the notation, so $\hat{\mu}^{g,f}$ is the curvature distribution of population f under construction g. A *graph* is one point cloud under one construction. Because every construction is built on the same tracks or atoms, a point cloud has the same multiplicity, the same composition, and the same bin under every g.

In JetSet, a population is a jet flavor, $\mathcal{F}$ is the set of flavors, and $\mathcal{P}$ is the pair set of Equation 4. In QM9, a population is one of the truth-labeled subsets of § D.2, $\mathcal{P}$ is a set of subset pairs, and $\mathcal{F}$ is the set of subsets appearing in $\mathcal{P}$. The flavors of JetSet are disjoint sets of point clouds, whereas the

two subsets of a QM9 pair are drawn from the same molecules. This difference sets the bin weights of Equation 14.

Why condition. Populations that differ only in n or in how points split among labeled parts have different curvature distributions, so $S_{\text{sep}} > 0$ even with identical local geometry; the complete graph, whose FRC is n on every edge (§ C.4), separates jet flavors for this reason alone.

Conditioning variables and bins. For a jet with track set T, let $v(t)$ be the truth vertex index of track t: $v(t) = 0$ for a PV track, $v(t) > 0$ for an SV track, and $v(t) < 0$ for a track matched to neither. For a molecule with atom set A, let $h(a) \in \{0, 1\}$ indicate whether atom a is a hydrogen. The conditioning variables are

$$\text{JetSet:} \quad n = |T|, \qquad n_{\text{sv}} = \big|\{t \in T : v(t) > 0\}\big|, \qquad r = n_{\text{sv}}/n, \tag{9}$$

$$\text{QM9:} \quad n = |A|, \qquad n_{\text{H}} = \big|\{a \in A : h(a) = 1\}\big|, \qquad n_{\text{heavy}} = n - n_{\text{H}}, \qquad r^{\text{H}} = n_{\text{H}}/n. \tag{10}$$

Unmatched tracks count toward n but not toward n_{sv}. For jets we condition on n, n_{sv}, and r separately, and on (n, n_{sv}) jointly. For QM9 we condition on n, n_{H}, n_{heavy}, and r^{H} separately, and on $(n_{\text{heavy}}, n_{\text{H}})$ jointly. The joint QM9 variable is equivalent to (n, n_{H}), but since most QM9 molecules have exactly nine heavy atoms, it gives a regular grid rather than a diagonal one.

The hydrogen fraction r^{H} tests the prediction of § F.1 that edge recovery falls as r^{H} rises, since each hydrogen carries $k - 1$ forced non-bond edges under minimum degree k.

A bin is a range of one conditioning variable or, under joint conditioning, a pair of ranges. Bins are listed in Table 4. They are disjoint but need not cover the full range of a variable, so a point cloud falls into at most one bin, and point clouds falling into none are discarded.

Table 4: Bins for the conditioned scores. Integer bins are closed on both ends. The fraction bins are right-closed, except the first, which is closed on both ends. The occupancy thresholds are $N_{\text{min}} = 200$ and $N_{\text{val}} = 50$ (§ C.1).

Variable	Bins
JetSet	
n	$[3, 5]$, $[6, 9]$, $[10, 14]$, $[15, 19]$, $[20, 30]$, $[31, 40]$
n_{sv}	$\{0\}$, $[1, 2]$, $[3, 4]$, $[5, 7]$, $[8, 10]$, $[11, 15]$, $[16, 20]$, $[21, 40]$
r	$\{0\}$, $(0, 0.2]$, $(0.2, 0.4]$, $(0.4, 0.6]$, $(0.6, 0.8]$, $(0.8, 1]$
(n, n_{sv})	product of the n and n_{sv} bins (6×8 joint bins)
QM9	
n	$[3, 6]$, $[7, 10]$, $[11, 14]$, $[15, 18]$, $[19, 22]$, $[23, 29]$
n_{H}	$[0, 2]$, $[3, 5]$, $[6, 8]$, $[9, 11]$, $[12, 14]$, $[15, 20]$
n_{heavy}	$[1, 4]$, $[5, 6]$, $\{7\}$, $\{8\}$, $\{9\}$
r^{H}	$[0, 0.3]$, $(0.3, 0.4]$, $(0.4, 0.5]$, $(0.5, 0.6]$, $(0.6, 0.7]$, $(0.7, 1]$
$(n_{\text{heavy}}, n_{\text{H}})$	product of the n_{heavy} and n_{H} bins (5×6 joint bins)

Separation within a bin. Let $\hat{\mu}^{g,f}(b)$ be the curvature distribution of population f under construction g, restricted to the point clouds in bin b and built with the same aggregation as the unconditioned distribution. Nodes and edges are assigned to subsets exactly as in the unconditioned analysis, so conditioning changes which point clouds enter a distribution and nothing else. The separation of pair $p = (f_A, f_B)$ within b is the single W_1 term of Equation 4 on that bin,

$$S_{\text{sep}}(g, p \mid b) = W_1\big(\hat{\mu}^{g,f_A}(b),\ \hat{\mu}^{g,f_B}(b)\big), \tag{11}$$

and the separation of g within b is its mean over the pairs that survive in that bin,

$$S_{\text{sep}}(g \mid b) = \frac{1}{|\mathcal{P}_b|} \sum_{p \in \mathcal{P}_b} S_{\text{sep}}(g, p \mid b), \tag{12}$$

where $\mathcal{P}_b \subseteq \mathcal{P}$ is the set of pairs for which both populations satisfy the N_{min} and N_{val} thresholds in b. We report the pair terms as well as the mean, because pairs need not respond to a bin in the same way, and the average over pairs can hide opposite trends.

Separation at fixed c. To obtain one number per construction, we average over bins, weighting each bin by the number of point clouds it contains, so that every point cloud contributes equally regardless of its bin. For a construction g and pair $p = (f_A, f_B)$, let $\mathcal{B}_{g,p}$ be the set of bins in which Equation 11 is defined, and let $\mathcal{P}_g \subseteq \mathcal{P}$ be the pairs with non-empty $\mathcal{B}_{g,p}$. The separation of g at fixed c is the two-level average

$$S_{\text{sep}}(g \mid c) = \frac{1}{|\mathcal{P}_g|} \sum_{p \in \mathcal{P}_g} S_{\text{sep}}(g, p \mid c), \qquad S_{\text{sep}}(g, p \mid c) = \frac{\sum_{b \in \mathcal{B}_{g,p}} w_{p,b}\, S_{\text{sep}}(g, p \mid b)}{\sum_{b \in \mathcal{B}_{g,p}} w_{p,b}}, \tag{13}$$

which is Equation 5 together with the averaging over pairs. The pair score is a weighted mean over the bins retained for that pair, and the overall score is an unweighted mean over pairs. The averages are taken in this order because $\mathcal{B}_{g,p}$ can differ between pairs: pooling all (pair, bin) terms would let a bin retained for one pair contribute that pair's distance to a score meant to describe all pairs. We report both quantities, along with $|\mathcal{P}_g|$, which equaled $|\mathcal{P}|$ for every construction in our runs. For QM9, $|\mathcal{P}| = 1$ and the two coincide.

The weight $w_{p,b}$ counts the distinct point clouds in bin b that belong to either population of p, not curvature values, so a bin gains no weight from its point clouds having more nodes or edges. In JetSet, each jet carries one flavor, so the counts add. In QM9, both populations come from the same molecules, so the weight is the number of molecules in b:

$$w_{p,b} = \begin{cases} N_{f_A,b} + N_{f_B,b} & \text{JetSet,} \\ N_b & \text{QM9,} \end{cases} \tag{14}$$

where $N_{f,b}$ is the number of point clouds of population f in b and N_b the number of molecules in b. The weight does not depend on the construction.

Preserved fraction. Binning, comparing within bins, and recombining with the weights above is adjustment by subclassification (Cochran, 1968), with populations, curvature, and the conditioning variable in place of Cochran's groups, outcome, and covariate; the bias he removes is the quantity Q2 asks about. Since W_1 units differ across curvatures and constructions, we report the dimensionless Equation 6: conditioning on n gives the *size-preserved* fraction preserved-sep$_n$, and on n_{sv}, r, n_{H}, or r^{H} the *composition-preserved* fractions.

The matched denominator. Both terms of the fraction must use the same sample. $S_{\text{sep}}(g \mid c)$ uses only point clouds in retained bins, and the dropped ones are not random: at fixed n, those outside the bin range; at fixed n_{sv} on SV subsets, every jet with $n_{\text{sv}} = 0$, which is most light jets. We therefore compute the pooled separation on exactly the point clouds the bins retain,

$$S_{\text{sep}}(g \mid \mathcal{B}) = \frac{1}{|\mathcal{P}_g|} \sum_{p \in \mathcal{P}_g} S_{\text{sep}}(g, p \mid \mathcal{B}), \qquad S_{\text{sep}}(g, p \mid \mathcal{B}) = W_1\big(\hat{\mu}^{g,f_A}(\mathcal{B}_{g,p}),\ \hat{\mu}^{g,f_B}(\mathcal{B}_{g,p})\big), \tag{15}$$

where $\hat{\mu}^{g,f}(\mathcal{B}_{g,p})$ is $\hat{\mu}^{g,f}$ restricted to the point clouds in the retained bins, exactly as $\hat{\mu}^{g,f}(b)$ is its restriction to one bin. and the per-pair fraction $\text{pres}_c(g, p) = S_{\text{sep}}(g, p \mid c)/S_{\text{sep}}(g, p \mid \mathcal{B})$ uses the single pair term in numerator and denominator. Matching matters even at near-full coverage, since a few high-multiplicity point clouds can carry a large share of the edge values.

How to read preserved-sep$_c$**.** Two effects keep even a pure size effect above zero: bins have finite width (subclassification removes about 64% of a covariate's contribution with two equal subclasses and 92% with six; Cochran 1968), and every within-bin distance carries the bias of Equation 8. FC, whose preserved-sep$_n$ would be zero under exact conditioning on an infinite sample, shows what pure size dependence looks like under our bins, and we read every other construction against it. The bias grows as cells get thinner, so it is largest under joint conditioning, where a value higher than under a single variable can reflect thinner cells rather than weaker dependence.

preserved-sep$_c$ can also exceed one. For a quantity measured per point cloud, variance splits exactly into within-bin and between-bin parts, $\text{Var}(y) = \mathbb{E}_b[\text{Var}(y \mid b)] + \text{Var}_b[\mathbb{E}(y \mid b)]$, both non-negative, so conditioning can only remove variance. W_1 has no such decomposition: the pooled distance compares two mixtures over bins, and within-bin differences can cancel in the mixture,

for instance when one population is more curved than the other at low n and less curved at high n. The within-bin distance then exceeds the pooled one. preserved-sep$_c$ is non-negative because $S_{\text{sep}}(g \mid c) \geq 0$, but it has no upper bound, and a value above one says that pooling was hiding a separation rather than that the variable produced one.

Coverage. A value of preserved-sep$_c$ describes only the point clouds it was computed on, so each is reported with its *coverage*, the fraction of the sample the retained bins hold,

$$c_p(g) = \frac{\sum_{b \in \mathcal{B}_{g,p}} w_{p,b}}{W_p}, \tag{16}$$

where W_p is the same count taken over the whole sample: $N_{f_A} + N_{f_B}$ for jets, with N_f the number of jets of flavor f in the analyzed sample, and the number of molecules in the sample for QM9. A point cloud is uncovered however it was lost — to N_{min}, to N_{val}, or by lying outside every bin of Table 4. For the pair-averaged fraction we report the pooled coverage $\sum_p \sum_{b \in \mathcal{B}_{g,p}} w_{p,b} / \sum_p W_p$, in which a light jet counts once per pair, so it is a pair-weighted fraction rather than a fraction of jets. The counts $N_{f,b}$ behind these weights, and the populations that fall below N_{min} in each bin, are given in Table 2.

C.3 Uncertainty estimates

Resampling scheme. Every separation score and every share is computed from a finite sample of point clouds. We attach to each a nonparametric bootstrap interval whose resampling unit is the point cloud (a jet or a molecule), not the curvature value. Values from one graph are dependent, since nodes share edges and edges share endpoints, so resampling them individually would count a point cloud with forty points as forty independent observations and give intervals that are too narrow. We therefore draw point clouds with replacement (Field & Welsh, 2007).

Let $j = 1, \ldots, N$ index the point clouds, divided into groups $\mathcal{N}_\ell$ of sizes N_ℓ that are resampled separately. Replicate $r = 1, \ldots, B$ is a vector of multiplicities $m_r = (m_{r,1}, \ldots, m_{r,N})$, drawn independently in each group,

$$(m_{r,j})_{j \in \mathcal{N}_\ell} \sim \text{Multinomial}\big(N_\ell;\, 1/N_\ell, \ldots, 1/N_\ell\big), \tag{17}$$

so that $\sum_{j \in \mathcal{N}_\ell} m_{r,j} = N_\ell$ in every replicate. In JetSet, the two populations of a pair are different flavors and therefore disjoint sets of jets, so the groups are the flavors and every replicate contains the observed number of jets of each flavor. In QM9, the two populations of a pair are subsets of the atoms or bonds of the same molecules, so there is one group: molecules are drawn once and both subsets inherit the same multiplicities, which preserves the dependence between the two sides.

Replicates of W_1. A point cloud drawn m times carries multiplicity m into every subset it enters. Let X and Y be the two sets of values of one W_1 term, x_i the value with index i, and $j(i)$ the point cloud it comes from. The resampled distribution of X in replicate r is the weighted empirical distribution

$$F^*_{X,r}(t) = \frac{\sum_{i \in X} m_{r,j(i)}\, \mathbf{1}[x_i \leq t]}{\sum_{i \in X} m_{r,j(i)}}, \tag{18}$$

and the replicate of the distance is

$$W^*_{1,r}(X, Y) = \int \left| F^*_{X,r}(t) - F^*_{Y,r}(t) \right| dt. \tag{19}$$

Under the union aggregation, i runs over the curvature values, so each value of a point cloud receives that point cloud's multiplicity. Under the per-point-cloud mean, i runs over the point clouds, x_i is the mean of point cloud i, which resampling leaves unchanged, and $j(i) = i$. Setting every $m_{r,j} = 1$ in Equation 18 recovers the full-sample estimate. The integral in Equation 19 is a finite sum over the pooled sorted values, so a replicate is evaluated exactly and no resampled array is ever built.

Replicates of the scores. One set of $B = 500$ resamples, stored as a $B \times N$ matrix of multiplicities, is reused for every construction, subset, curvature, aggregation, and conditioning variable, which is possible because all constructions are built on the same point clouds. Within a replicate we

recompute every quantity that enters a reported score on the same resampled point clouds. Writing $\mathcal{J}_{p,b}$ for the point clouds of the populations of pair p that fall in bin b, the bin weights become

$$w^*_{p,b,r} = \sum_{j \in \mathcal{J}_{p,b}} m_{r,j}, \tag{20}$$

which change with the resample because they are counts of point clouds. For construction g and pair p the separation at fixed c is the weighted mean of the per-bin distances,

$$S^*_{\mathrm{sep},r}(g, p \mid c) = \frac{\sum_{b \in \mathcal{B}_{g,p}} w^*_{p,b,r}\, W^*_{1,r}(g,p,b)}{\sum_{b \in \mathcal{B}_{g,p}} w^*_{p,b,r}}, \tag{21}$$

and its same-sample counterpart $S^*_{\mathrm{sep},r}(g, p \mid \mathcal{B})$ is a single distance $W^*_{1,r}$ computed without binning: on each side of the pair the values from all retained bins $b \in \mathcal{B}_{g,p}$ are pooled into one set, and the distance is taken between the two pooled sets, so both scores rest on the same values. The scores of a construction follow by averaging over its pairs and then forming the ratio of Equation 6,

$$\text{preserved-sep}^*_{c,r}(g) = \frac{|\mathcal{P}_g|^{-1} \sum_{p \in \mathcal{P}_g} S^*_{\mathrm{sep},r}(g, p \mid c)}{|\mathcal{P}_g|^{-1} \sum_{p \in \mathcal{P}_g} S^*_{\mathrm{sep},r}(g, p \mid \mathcal{B})}. \tag{22}$$

The averages and the ratio are taken inside the replicate, so each interval refers to the quantity that is reported. The unconditioned score is the special case of a single bin that holds every point cloud, for which $S_{\mathrm{sep}}(g \mid \mathcal{B})$ is the unconditioned separation.

Two properties follow from Equation 22 and from the shared matrix of multiplicities. First, the numerator and denominator of preserved-sep$_c$ are functions of the same m_r, so the interval on preserved-sep$_c$ carries their correlation. Second, comparisons between constructions are paired: replicate r contains the same point clouds for every construction, so the uncertainty of a difference between two constructions g and g' is obtained from the replicate wise differences

$$\Delta^*_r = \theta^*_r(g) - \theta^*_r(g'), \qquad r = 1, \ldots, B, \tag{23}$$

and not from the two separate intervals. Rank statistics over constructions are treated in the same way.

Fixed selection, standard error, and interval. The bin edges, the retained sets $\mathcal{B}_{g,p}$ (the bins in which both populations pass $N_{\min}$ and both sides hold at least N_{val} values), and the pair sets $\mathcal{P}_g$ are held at their full-sample values. Every replicate therefore estimates the same quantity, and the intervals are conditional on this selection: they do not include the variability of a bin entering or leaving the retained set. Let $\hat{\theta}$ be a score on the full sample, $\theta^*_1, \ldots, \theta^*_B$ its replicates, $\bar{\theta}^*$ their mean, and q_α their empirical α-quantile. The bootstrap standard error is the standard deviation of the replicates,

$$\widehat{\mathrm{SE}} = \sqrt{\frac{1}{B-1} \sum_{r=1}^{B} (\theta^*_r - \bar{\theta}^*)^2}, \qquad \bar{\theta}^* = \frac{1}{B} \sum_{r=1}^{B} \theta^*_r, \tag{24}$$

and we report the 95% basic bootstrap interval

$$\left[\, 2\hat{\theta} - q_{0.975},\;\; 2\hat{\theta} - q_{0.025} \,\right], \tag{25}$$

truncated below at zero, since both a separation and preserved-sep$_c$ are non-negative. A replicate in which one side of a term receives zero total multiplicity is undefined for that term and is omitted, in which case B in Equation 24 denotes the number of defined replicates. When conditioning removes the separation, W_1 is biased upward by Equation 8 and its bootstrap is unreliable (Sommerfeld & Munk, 2018).

Across-bin variance. The intervals above describe how well a score is determined by the sample. A separate question is how strongly the separation depends on the conditioning variable, which we summarize by the sample variance of the per-bin scores. Let $\hat{S}_{g,b}$ be the score of construction g in bin b (the mean of W_1 over the pairs retained in that bin) and $\mathcal{B}_g$ the bins in which it exists. Then

$$s^2_g = \frac{1}{|\mathcal{B}_g| - 1} \sum_{b \in \mathcal{B}_g} (\hat{S}_{g,b} - \bar{S}_g)^2, \qquad \bar{S}_g = \frac{1}{|\mathcal{B}_g|} \sum_{b \in \mathcal{B}_g} \hat{S}_{g,b}, \tag{26}$$

where the mean over bins is unweighted, since the question is how much the score changes across the range of the conditioning variable and not across point clouds. This is a descriptive measure of heterogeneity, not an uncertainty, and it enters no interval. It is nevertheless affected by sampling noise. The bins hold disjoint point clouds, so the $\hat{S}_{g,b}$ are independent to a good approximation, and

$$\mathbb{E}\big[s_g^2\big] = \sigma_g^2 + \frac{1}{|\mathcal{B}_g|} \sum_{b \in \mathcal{B}_g} \mathrm{se}_{g,b}^2, \tag{27}$$

with σ_g^2 the variance of the true per-bin separations and $\mathrm{se}_{g,b}$ the standard error of $\hat{S}_{g,b}$. A construction with thin, noisy bins would otherwise appear more dependent on the conditioning variable than it is. We therefore report

$$\hat{\sigma}_g^2 = \max\Big(0,\ s_g^2 - \frac{1}{|\mathcal{B}_g|} \sum_{b \in \mathcal{B}_g} \widehat{\mathrm{SE}}_{g,b}^{\,2}\Big), \tag{28}$$

with $\widehat{\mathrm{SE}}_{g,b}$ from Equation 24 applied to the replicates of $\hat{S}_{g,b}$. The correction removes the noise term of Equation 27 but not the upward bias of Equation 8, which is larger in thin bins.

C.4 A WORKED EXAMPLE OF THE DISTRIBUTIONAL MEASURES

We trace § 3.1 and § 3.2 by hand on eight jets (Table 5), asking whether the curvature distributions of b and light jets differ among jets of the same size. Under FC with the per-graph mean, every edge of K_n has $\deg(i) = \deg(j) = n - 1$ and $|\triangle(i, j)| = n - 2$, so Equation 2 gives

$$\kappa_{\mathrm{FR}}(i, j) = 4 - \deg(i) - \deg(j) + 3\,|\triangle(i, j)| = 4 - 2(n - 1) + 3(n - 2) = n, \tag{29}$$

so each jet contributes its track count. The light jets have $n \in \{4, 4, 5, 10\}$ and the b jets $n \in \{5, 10, 10, 12\}$, so any separation must come from size. Each W_1 compares curvature values, whereas bins are set by n: a light jet with $n = 4$ falls in the 3-5 bin under every construction, carrying $\kappa_{\mathrm{FR}} = 4$ under FC and some other value under a sparse graph. That coincidence under FC is what makes it a control.

Raw score. For the single pair $\mathcal{P} = \{(b, \text{light})\}$, Equation 4 reduces to

$$S_{\mathrm{sep}}(g) = \frac{1}{|\mathcal{P}|} \sum_{(f_A, f_B) \in \mathcal{P}} W_1\left(\hat{\mu}^{g,f_A}, \hat{\mu}^{g,f_B}\right) = W_1\big(\{4, 4, 5, 10\}, \{5, 10, 10, 12\}\big), \tag{30}$$

with W_1 the area between the two empirical CDFs (§ 3),

$$W_1(\hat{\mu}, \hat{\nu}) = \int_{\mathbb{R}} \big|F_{\hat{\mu}}(t) - F_{\hat{\nu}}(t)\big|\, dt. \tag{31}$$

The CDF gives, at each value x, the fraction of jets with a value less than or equal to x. Considering the intervals between the possible values 4, 5, 10, 12,

$$F_{\text{light}}(x) = \begin{cases} 0, & x < 4, \\ \frac{1}{2}, & 4 \le x < 5, \\ \frac{3}{4}, & 5 \le x < 10, \\ 1, & x \ge 10, \end{cases} \tag{32}$$

$$F_b(x) = \begin{cases} 0, & x < 5, \\ \frac{1}{4}, & 5 \le x < 10, \\ \frac{3}{4}, & 10 \le x < 12, \\ 1, & x \ge 12. \end{cases} \tag{33}$$

The difference is then

$$F_{\text{light}}(x) - F_b(x) = \begin{cases} \frac{1}{2} - 0 = \frac{1}{2}, & 4 \le x < 5, \\ \frac{3}{4} - \frac{1}{4} = \frac{1}{2}, & 5 \le x < 10, \\ 1 - \frac{3}{4} = \frac{1}{4}, & 10 \le x < 12, \end{cases} \tag{34}$$

and outside those intervals the two CDFs are equal, so they do not contribute. Substituting,

$$S_{\text{sep}}(g) = \frac{1}{2} \cdot (5-4) + \frac{1}{2} \cdot (10-5) + \frac{1}{4} \cdot (12-10) \tag{35}$$

$$= \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 5 + \frac{1}{4} \cdot 2 \tag{36}$$

$$= 3.5. \tag{37}$$

For equal-size samples this agrees with the sorted- neighbor shortcut, $W_1 = \frac{1}{4} \sum_{i=1}^{4} \left| x_{\text{light}}^{(i)} - x_b^{(i)} \right|$ with $x^{(i)}$ the i-th smallest value of each sample, which gives $\frac{1}{4}(1+6+5+2) = 3.5$.

Conditioned score. We take the bins {3-5, 6-12}. A jet goes to the bin containing its track count, so the first bin holds the light jets with $n = 4, 4, 5$ and the b jet with $n = 5$, and the second holds the light jet with $n = 10$ and the b jets with $n = 10, 10, 12$. Both flavors have at least one jet in both bins, so both bins enter and all eight jets are retained. The conditioned score of § 3.2 is Equation 4 evaluated on the jets of one bin,

$$S_{\text{sep}}(g \mid b) = W_1\big(\hat{\mu}^{g,\text{light}}(b),\ \hat{\mu}^{g,b}(b)\big), \tag{38}$$

where $\hat{\mu}^{g,f}(b)$ is the distribution of the values of the jets of flavor f in bin b. The samples in a bin have unequal sizes, so the sorted- neighbor shortcut does not apply and we integrate the CDFs.

Bin 3-5. The light values are $\{4, 4, 5\}$ and the b value is $\{5\}$, so

$$F_{\text{light}}^{(1)}(x) = \begin{cases} 0, & x < 4, \\ \frac{2}{3}, & 4 \le x < 5, \\ 1, & x \ge 5, \end{cases} \qquad F_b^{(1)}(x) = \begin{cases} 0, & x < 5, \\ 1, & x \ge 5, \end{cases} \tag{39}$$

and the difference is $\frac{2}{3} - 0 = \frac{2}{3}$ on $[4, 5)$ and zero everywhere else. Hence

$$S_{\text{sep}}(g \mid \text{3-5}) = \tfrac{2}{3} \cdot (5-4) = \tfrac{2}{3}. \tag{40}$$

Bin 6–12. The light value is $\{10\}$ and the b values are $\{10, 10, 12\}$, so

$$F_{\text{light}}^{(2)}(x) = \begin{cases} 0, & x < 10, \\ 1, & x \ge 10, \end{cases} \qquad F_b^{(2)}(x) = \begin{cases} 0, & x < 10, \\ \frac{2}{3}, & 10 \le x < 12, \\ 1, & x \ge 12, \end{cases} \tag{41}$$

and the difference is $1 - \frac{2}{3} = \frac{1}{3}$ on $[10, 12)$ and zero everywhere else. Hence

$$S_{\text{sep}}(g \mid \text{6-12}) = \tfrac{1}{3} \cdot (12-10) = \tfrac{2}{3}. \tag{42}$$

Preserved fraction. Equation 6 compares the conditioned score to the pooled one on the same jets. Here all eight jets entered both, so the coverage is 100% and

$$\text{preserved-sep}_c(g) = \frac{S_{\text{sep}}(g \mid n)}{S_{\text{sep}}(g \mid \mathcal{B})} = \frac{2/3}{3.5} = 0.190. \tag{43}$$

Only 19% of FC's separation survives conditioning on n; the remaining 81% was size.

Table 5: The distributional measures on an eight-jet toy set under FC, where $\kappa_{\mathrm{FR}} = n$, so each jet contributes its own track count. Only 19% of the separation survives conditioning on n, recovering by construction the fact that nothing but size distinguishes the flavors here.

	light values	b values	W_1
all jets	4, 4, 5, 10	5, 10, 10, 12	3.500
bin $n \in [3, 5]$	4, 4, 5	5	0.667
bin $n \in [6, 12]$	10	10, 10, 12	0.667
$S_{\mathrm{sep}}(g \mid n)$, occupancy-weighted ($w_1 = w_2 = 4$)			0.667
preserved-sep$_n$ = $S_{\mathrm{sep}}(g \mid n)/S_{\mathrm{sep}}(g \mid \mathcal{B})$			0.190

D Data and Model Details

D.1 Input variables and targets

Table 6: Variables used in the QM9 analyses with two regression targets and per-atom input variables are listed below.

Variable	Definition
Inputs	
x, y, z	Atom's 3D position
Z	Element (atomic number)
M	Atomic mass
q	Mulliken partial charge on the atom
Labels	
HOMO-LUMO gap	Energy difference between the highest occupied and lowest unoccupied molecular orbitals
μ	Dipole moment - the magnitude of the molecule's electric dipole, measuring the separation of positive and negative charge

Table 7: Variables for JetSet with classification task for flavors and per-track variables listed below

Variable	Definition
Inputs	
$\Delta\eta$	Pseudorapidity distance between track and jet
$\Delta\phi$	Azimuthal distance between track and jet
p_T	Transverse momentum, i.e., the component of the momentum of the track perpendicular to the beam axis
Labels	
PV	Primary vertex–the location of the initial particle collision
SV	Secondary vertex–the locations where decays occur, away from the PV
Flavor	Ground-truth label representing the type of the initiating parton (light, charm (c), bottom (b), or tau (τ)).

D.2 Structural subsets of QM9

The three binary atom partitions of § D.2 are defined as follows, all read off the bond graph B and the elements. *Hydrogen vs heavy* splits atoms by element. *Ring vs chain* applies to heavy atoms, with ring membership defined as lying in a biconnected component of B with ≥ 3 atoms, independent of bond type or aromaticity conventions. *Terminal vs branch* also applies to heavy atoms, separating those with one heavy neighbor in B from those with two or more. Hydrogens and heavy atoms with no heavy neighbor belong to neither class.

D.3 Model architectures and training protocol

Inputs. The linear model, the two MLPs, and the boosted trees operate on a single fixed-length vector of summary statistics per point cloud. For jets, each per-track variable ($\Delta\eta$, $\Delta\phi$, p_T) is summarized by ten statistics (mean, median, standard deviation, min, max, skewness, and the 10th/25th/75th/90th percentiles) over the jet's tracks, plus the track count, giving 31 baseline features. The curvature features are the same ten statistics of the per-track ORC and FRC, for an additional 20 features. The corresponding QM9 feature vector is described in § 4.1. Summary values that are undefined for a given point cloud (e.g., the skewness of a variable that is constant within the point cloud) are replaced by zero. The summary statistics feature vector for QM9 mirrors that jet vector: ten statistics of each per-atom variable of Table 6 and the atom count, with the same statistics of per-atom ORC and FRC as curvature features. Summary values that are undefined for a point cloud, for example, the skewness of a constant variable, are replaced by zero.

The transformer and GNN instead receive the padded point clouds directly, with the two per-point curvature values appended as extra channels in the curvature setting; all inputs are standardized to zero mean and unit variance using training-set statistics computed over real, non-padded points, except the boosted trees, which use unscaled features since threshold splits are invariant to rescaling.

Summary-statistic models. The four summary-statistic models are as follows. The *linear model* is scikit-learn's `LogisticRegression` with an L_2 penalty ($C = 1$, `lbfgs` solver, iteration limit raised to 1000) for classification, and ridge regression (`Ridge`, $\alpha = 1$) for the QM9 regression targets. The two MLPs are fully connected networks of ReLU units with a single linear output (the logit for classification, the raw value for regression): the *small MLP* has one hidden layer of 64 units, and the *large MLP* has three hidden layers of 256, 128, and 64 units. The *gradient-boosted trees* (`XGBoost`, 300 trees of maximum depth 6, histogram tree builder) fit a sequence of small decision trees, each of which is trained to correct the errors of those before it.

Small transformer. We use a transformer encoder to process each point cloud independently. The model is invariant to permutations of the points and makes no use of the graph structure. Each point feature is first projected to a 64-dimensional embedding using a linear layer. This is followed by two standard `PyTorch` transformer encoder layers, each with 4 attention heads, a 128-dimensional feed-forward layer, dropout of 0.1,and prelayer normalization. Self-attention is computed over all points within a point cloud, with padded points masked out. The resulting per-point representations are averaged over the non-padded points and passed to a small prediction head (a 64-unit linear layer with ReLU, followed by a linear output layer), which produces a single scalar.

Graph neural network. The final model is a message-passing network, written directly in `PyTorch` without a graph library, that receives the point list as well as the constructed graph, with the same standardization and 64-dimensional embedding as the transformer. Each undirected edge is stored twice, once in each direction, so every node exchanges information with all of its neighbors. The model has three message-passing layers. At each layer, an edge function (a linear layer with ReLU) computes a 64-dimensional message from the concatenated representations of the two endpoints of the edge. Each node then averages its incoming messages, with padded edges excluded by masking; the average is concatenated with the node's current representation and passed through a second linear layer with ReLU to produce the updated representation. Padded nodes are re-zeroed after every layer. The final node representations are mean-pooled over the real nodes and passed to a prediction head of the same form as that of the transformer.

Training protocol and uncertainty. After every epoch, each network is scored on its early-stopping holdout, which is drawn per seed from the training point clouds and never overlaps the held-out evaluation set. When training stops, the parameters from the best epoch are restored and the model is evaluated once on the held-out set. The held-out set is therefore never used for model selection or early stopping. The reported QM9 MAEs are given in the original units of each target.

We quantify two sources of uncertainty: variation across model seeds, and finite-sample variation in the test set. For each model, we use five random seeds, $\{0, 1, 2, 4, 44\}$. For the neural networks, the seed determines the weight initialization, the minibatch shuffling, and the validation split of 10% used for early stopping; for the boosted trees, it determines the random state. Logistic regression is

Table 8: Training hyperparameters. Entries of the form JetSet / QM9 differ between the two datasets; all other settings are shared. Early stopping (ES) holds out 10% of the training point clouds per seed, monitors AUROC (classification) or MAE (regression) on this holdout, stops after the stated patience without improvement above the threshold, and restores the weights of the best epoch before evaluation. The QM9 regression targets are standardized with training-set statistics during training; predictions are returned to the original units before scoring.

	Linear	MLPs	Grad boosting	Transformer, GNN
Optimizer	`lbfgs` / closed form	AdamW	–	AdamW
Learning rate	–	10^{-3}	0.1	10^{-3} / 3×10^{-4} (Tr.), 10^{-3} (GNN)
Weight decay	–	10^{-4}	–	10^{-4}
Batch size	full	1024	full	512
Budget	≤1000 iter.	≤200 epochs	300 trees	≤30 / ≤150 epochs
ES patience	–	15	–	5 / 25
ES threshold	–	10^{-5}	–	10^{-4}
LR schedule	–	None	–	cosine
Gradient clipping	–	None	–	1.0
Loss	log-loss / least squares	BCE / MSE	log-loss / squared error	BCE / MSE

effectively deterministic under this configuration, but we apply the same five-seed protocol for consistency. To quantify test-set uncertainty, each trained model is evaluated on 50 bootstrap resamples of the test point clouds, drawn with replacement and without retraining. For each seed, we report the mean AUROC across the bootstrap resamples, and take the bootstrap standard deviation as the measure of finite-sample uncertainty. The final result is the mean across seeds, with the seed to seed and bootstrap uncertainties combined in quadrature, $\sigma = \sqrt{\sigma_{\text{seed}}^2 + \sigma_{\text{boot}}^2}$. The tables of Appendix G report, for each mean, the 95% percentile interval from resampling the five seeds with replacement 10,000 times, jointly for the with and without curvature runs so the pairing is preserved with bold marks intervals that exclude zero.

D.4 Compute

JetSet models were trained on a server with two NVIDIA A100 (40 GB) GPUs, a 64-core AMD EPYC 7662 CPU, and 720 GB of RAM. Graph construction and curvature computation were run in parallel over about 200 CPU cores on a compute node of a shared cluster; each point cloud is processed independently, so the parallelization does not affect any result. QM9 models, the Wasserstein-1 separation scores, and some JetSet GNN runs used a server with two NVIDIA H100 NVL (94 GB) GPUs, a 48-core AMD EPYC 9474F CPU, and 1.5 TB of RAM. PyTorch, GraphRicciCurvature, scikit-learn, XGBoost were used and specific versions can be found in the repo provided.

E QM9 Reference Bond Graph

E.1 Verification procedure

We take bonds from the MoleculeNet release of QM9 (Wu et al., 2018) (`gdb9.sdf`, as distributed by PyTorch Geometric (Fey & Lenssen, 2019)) and verify them independently against the SMILES strings in the original `xyz` files (Ramakrishnan et al., 2014). For every molecule we require two things. First, the SDF atom ordering must match the `xyz` file element by element, so that bonds and coordinates refer to the same atoms. Second, the connectivity implied by the SDF bonds, ignoring bond order, charge, aromaticity, and stereochemistry, must agree with one of the two SMILES strings in the `xyz` file: the SMILES from the GDB-17 enumeration, which does not depend on the coordinates, or the SMILES of the relaxed geometry. We ignore bond order, charge, aromaticity, and stereochemistry because our constructions produce unweighted edges, so the only question they can answer is whether two atoms are connected. Because the GDB-17 SMILES does not depend on 3D coordinates, the check is independent of any distance rule, including the ϵ-based construction under evaluation.

All 133,885 molecules align atom by atom. SDF connectivity agrees with the relaxed-geometry SMILES for 132,478 of them and with the GDB-17 SMILES for one more, so 1,406 molecules

(1.1%) fail the comparison. A further 3,054 molecules are flagged by the QM9 authors as failing their consistency check between the relaxed geometry and the GDB-17 SMILES. Nine of these already fail the SMILES comparison, so the second filter removes 3,045 additional molecules and the two filters cut largely distinct sets. 129,434 molecules survive.

F Additional Results: Diagnostic Scores

F.1 Positive control: edge recovery on QM9

Before asking whether curvature computed on a construction carries information, we ask whether the construction recovers the bond graph at all. This subsection is a control on the graph constructions, not an evaluation of curvature. The curvature scores of § 3 never see the bond graph, so comparing them against this control tests whether the diagnostic reaches ground truth when the answer is known, which is what motivates its use on jets, where no ground truth exists. We score each construction against the bond graph using precision (fraction of constructed edges that are bonds), recall (fraction of bonds recovered), and F1, averaged per molecule over the 129,434 verified QM9 molecules of § E.1; pooled scores differ by less than 0.02. No model is trained, so the scores are a direct comparison of two edge lists.

Predictions. FC has recall 1 by construction, with precision fixed by atom count alone. The ϵ-based construction at 1.8 Å, near typical bond lengths, should recover the bond graph closely in both directions; it cannot itself serve as ground truth, since it is one of the constructions under test. kNN is the hard case. It enforces a minimum degree of k, but every hydrogen has exactly one bond, so at least $k-1$ of each hydrogen's edges are guaranteed non-bonds. This gives a counting bound: if each of n atoms keeps its k nearest neighbors, the construction places at least $nk/2$ edges, while the bond graph has $n\bar{d}/2$, so

$$\text{precision} = \frac{\text{correct edges}}{\text{all edges}} \leq \frac{n\bar{d}/2}{nk/2} = \frac{\bar{d}}{k}, \tag{44}$$

independent of which neighbors are chosen and independent of the molecule beyond $\bar{d}$. Molecular graphs are close to trees ($\bar{d} \approx 2$), so the bound sits near $2/k$. We predict that kNN precision is poor, that the loss falls on hydrogen endpoint edges, and that it grows with the hydrogen fraction r^{H}.

Discussion. The ϵ-based construction at 1.8 Å recovers every bond (recall 1.000) at precision 0.836, and every error involves hydrogen: precision is 1.000 on heavy–heavy edges against 0.694 on edges with a hydrogen endpoint (Table 10). Of its 526,248 non-bond edges, all but 134 join two hydrogens bonded to the same heavy atom, at distances of 1.51–1.80 Å (median 1.77 Å), just inside the cutoff. Lowering the cutoff to 1.5 Å removes these but also removes nearly half of the heavy–heavy bonds, which lie between 1.5 and 1.8 Å, whereas raising it only adds non-bonds, so precision decreases monotonically toward the FC baseline.

kNN does not close the gap to the ϵ-based construction. At $k = 3$, precision is 0.530 (recall 0.942). At $k = 4$, recall reaches 1.000 but precision is only 0.415, against 0.836 for the ϵ-based construction. The degree-floor bound of equation 44 averages 0.689 at $k = 3$, and the remainder of the gap comes from non-mutual neighbor relations and from about 12% of heavy–heavy bonds (6% of all bonds) being missed (heavy–heavy recall 0.878, overall recall 0.942). All three predictions hold: precision is 0.393on hydrogen endpoint edges against 0.863 on heavy–heavy edges at $k = 3$, and precision decreases monotonically with r^{H} (Table 11), tracking composition rather than ring count.

The rigidity constructions (unique-k, Laman) show the same mismatch even more sharply. Precision on hydrogen-endpoint edges is exactly $1/k$ (0.500 for Laman, and 0.333 and 0.250 for unique-k at $k = 3$ and $k = 4$), since every hydrogen bond is recovered and edge count alone fixes the ratio. As a result, the $p = 1$ and $p = -1$ variants of each rigidity construction are indistinguishable by edge recovery, despite placing very different non-bonds (H—Hat $p \geq 0$ against H–heavy at $p = -1$).

Relation to the curvature scores. We focus on the hydrogen vs heavy pair because it is the only split in § D.2 defined by true degree, which is the one property a minimum-degree construction directly constrains. If a construction imposes a false minimum degree, hydrogens should be pushed

Table 9: Edge recovery against the reference bond graph on the 129,434 verified QM9 molecules, next to the hydrogen vs heavy curvature separation, for the 19 constructions on which the curvature scores are computed. Precision (P), recall (R), and F1 are means over molecules, so mean F1 is not the harmonic mean of the P and R columns. P (H) is the pooled precision on edges with a hydrogen endpoint. S_{sep} is the Wasserstein-1 separation between the curvature distributions of hydrogens and of heavy atoms (per-molecule mean, unconditioned), and it never sees the bond graph. FC recall is 1 by construction, and its precision is the baseline set by the number of atoms. Ranking the 19 constructions by F1 and by ORC separation gives Spearman $\rho = 0.88$ (0.87 with the FC excluded); by F1 and FRC separation, $\rho = 0.09$ (-0.03 with the FC excluded).

	Edge recovery				Curvature separation	
Construction	P	R	F1	P (H)	ORC S_{sep}	FRC S_{sep}
FC	0.125	1.000	0.222	0.075	0.000	0.00
kNN, $k = 2$	0.684	0.804	0.738	0.584	0.208	0.96
kNN, $k = 3$	0.530	0.942	0.678	0.393	0.189	1.37
kNN, $k = 4$	0.415	1.000	0.586	0.281	0.092	0.97
kNN, $k = 5$	0.339	1.000	0.506	0.231	0.022	0.08
ϵ-based, 1.5 Å	1.000	0.785	0.874	1.000	0.160	0.24
ϵ-based, 1.8 Å	0.836	1.000	0.908	0.694	0.437	2.46
ϵ-based, 2.5 Å	0.377	1.000	0.546	0.272	0.085	0.57
ϵ-based, 3.0 Å	0.261	1.000	0.412	0.178	0.033	0.27
ϵ-based, 4.0 Å	0.173	1.000	0.294	0.111	0.008	1.32
k_T threshold, $p = 1$	0.140	0.496	0.218	0.138	0.038	2.49
k_T threshold, $p = 0$	0.281	1.000	0.438	0.190	0.028	0.10
k_T threshold, $p = -1$	0.276	1.000	0.431	0.186	0.026	2.60
Laman, $p = 1$	0.541	0.957	0.691	0.500	0.221	1.61
Laman, $p = -1$	0.541	0.957	0.690	0.500	0.125	0.36
unique-k, $k = 3$, $p = 1$	0.388	0.998	0.559	0.333	0.111	0.93
unique-k, $k = 3$, $p = -1$	0.388	0.998	0.558	0.333	0.045	0.72
unique-k, $k = 4$, $p = 0$	0.302	1.000	0.463	0.250	0.049	0.18
unique-k, $k = 4$, $p = -1$	0.302	1.000	0.463	0.250	0.003	2.01

toward the heavy-atom curvature distribution as k grows, while a construction that recovers the bond graph should keep them apart. No other label gives this prediction.

Edge recovery counts which constructed edges are bonds. The curvature scores never see the bond graph: separation compares the curvature distributions of two populations of atoms or edges defined by the labels of § D.2, at fixed size and composition. The two measurements therefore ask different questions of the same graph, and the control lets us check whether they agree. The right-hand columns of Table 9 report the hydrogen vs heavy pair, which is the pair the degree mismatch falls on. § F.3 reports all six pairs.

Under ORC, separation tracks edge recovery without seeing it. It peaks at 1.8 Å (0.437), falls to 0.160 at 1.5 Å (where heavy–heavy bonds start going missing) and to 0.09, 0.03, and 0.01 at 2.5, 3.0, and 4.0 Å. Across kNN it decreases monotonically (0.21, 0.19, 0.09, 0.02 for $k = 2$–5), and likewise for the rigidity constructions at $p = +1$ (0.22, 0.11,0.05 for $k = 2, 3, 4$). Ranking the 19 constructions by F1 and by ORC separation gives a Spearman rank correlation of $\rho = 0.88$, and 0.87 with the FC excluded (Table 14). By F1 and FRC separation, the correlation is indistinguishable from zero ($\rho = 0.09$, and -0.03 with the FC excluded,), underscoring that FRC does not track edge recovery the way ORC does. The diagnostic also resolves what edge recovery cannot: the $p = +1$ and $p = -1$ rigidity variants have identical edge recovery scores but differ by a factor of two or more in ORC separation (for example, 0.22 against 0.13 for Laman).

Limitations. Precision and recall ignore bond order, so confusing single and double bonds is invisible here. The verified set excludes molecules whose geometry and connectivity disagree. The kNN prediction concerns the mean degree of the true structure, so on point clouds with different degree distributions the same construction may behave differently, which is why we report QM9 and JetSet results separately rather than pooling them.

F.2 Edge recovery for all constructions

Table 10 reports every construction and parameter setting, scored on the same 129,434 verified molecules. In addition to the columns of Table 9, it reports (i) recall restricted to heavy–heavy bonds, (ii) the ratio of bonds to constructed edges, that is, the precision a construction would have if it recovered every bond, and (iii) the fraction of constructed edges that have a hydrogen endpoint. Columns (i) and (ii) decompose precision into the two loss mechanisms of § F.1: precision is recall times the bond-to-edge ratio, so a construction whose ratio equals its precision has lost nothing to missed bonds, and one whose ratio exceeds its precision has. The Bound column is equation 44, capped at 1 per molecule and averaged. It applies to kNN only.

Recall restricted to bonds involving a hydrogen is 1 for every construction in the table, which is why it is not shown: a hydrogen's nearest atom is always its bonding neighbor, so every construction recovers every bond to hydrogen. All recall losses in the table are therefore losses of heavy–heavy bonds. Most recall values shown as 1.000 are exact; two are rounded (k_T threshold with $p = -1$, and unique-k with $k = 4$, $p = -1$, both of which fall just short of 1), and only FC has recall 1 by construction.

Table 10: Edge recovery for all constructions on the 129,434 verified QM9 molecules. Precision (P), recall (R), and F1 are means over molecules, so mean F1 is not the harmonic mean of the P and R columns. P (H) and P (heavy) are pooled precisions on edges with a hydrogen endpoint and on heavy–heavy edges. R (heavy) is pooled recall on heavy–heavy bonds (recall on bonds to hydrogen is 1 for every row). Bonds/edges is the ratio of bonds to constructed edges, mean over molecules, which equals P when R is 1. H edges is the fraction of constructed edges with a hydrogen endpoint. Bound is the degree-floor ceiling equation 44 for kNN, capped at 1 per molecule and averaged. The k_T threshold rows use $q = 0.5$, $k_{\min} = 2$, and $k_{\max} = 7$.

Construction	P	R	F1	P (H)	P (heavy)	R (heavy)	Bonds/edges	H edges	Bound
FC	0.125	1.000	0.222	0.075	0.273	1.000	0.125	0.782	–
kNN, $k = 2$	0.684	0.804	0.738	0.584	0.935	0.601	0.854	0.724	0.994
kNN, $k = 3$	0.530	0.942	0.678	0.393	0.863	0.878	0.564	0.711	0.689
kNN, $k = 4$	0.415	1.000	0.586	0.281	0.774	1.000	0.415	0.730	0.517
kNN, $k = 5$	0.339	1.000	0.506	0.231	0.623	1.000	0.339	0.726	0.413
ϵ-based, 1.5 Å	1.000	0.785	0.874	1.000	1.000	0.549	1.000	0.642	–
ϵ-based, 1.8 Å	0.836	1.000	0.908	0.694	1.000	1.000	0.836	0.586	–
ϵ-based, 2.5 Å	0.377	1.000	0.546	0.272	0.577	1.000	0.377	0.676	–
ϵ-based, 3.0 Å	0.261	1.000	0.412	0.178	0.422	1.000	0.261	0.699	–
ϵ-based, 4.0 Å	0.173	1.000	0.294	0.111	0.316	1.000	0.173	0.736	–
k_T threshold, $p = 1$	0.140	0.496	0.218	0.138	0.768	0.015	0.286	0.997	–
k_T threshold, $p = 0$	0.281	1.000	0.438	0.190	0.487	1.000	0.281	0.716	–
k_T threshold, $p = -1$	0.276	1.000	0.431	0.186	0.455	1.000	0.276	0.707	–
Laman, $p = 1$	0.541	0.957	0.691	0.500	0.592	0.918	0.565	0.559	–
Laman, $p = 0$	0.541	0.957	0.691	0.500	0.592	0.918	0.565	0.559	–
Laman, $p = -1$	0.541	0.957	0.690	0.500	0.591	0.918	0.565	0.559	–
unique-k, $k = 2$, $p = -1$	0.541	0.957	0.690	0.500	0.591	0.918	0.565	0.559	–
unique-k, $k = 3$, $p = 1$	0.388	0.998	0.559	0.333	0.459	0.996	0.389	0.576	–
unique-k, $k = 3$, $p = 0$	0.388	0.998	0.559	0.333	0.459	0.996	0.389	0.576	–
unique-k, $k = 3$, $p = -1$	0.388	0.998	0.558	0.333	0.459	0.995	0.389	0.576	–
unique-k, $k = 4$, $p = 0$	0.302	1.000	0.463	0.250	0.373	1.000	0.302	0.595	–
unique-k, $k = 4$, $p = -1$	0.302	1.000	0.463	0.250	0.373	1.000	0.302	0.595	–

Table 11: kNN precision ($k = 3$) by hydrogen fraction, mean over the molecules in each bin, together with the degree-floor bound equation 44 and the mean number of rings. Precision falls with r^{H} in every bin. The bound, which depends on ring count rather than on composition, does not. The last bin contains only 9 molecules.

r^{H} bin	Molecules	Mean rings	Precision	Bound
$[0, 0.3]$	2,478	1.04	0.556	0.668
$(0.3, 0.4]$	16,275	1.38	0.545	0.684
$(0.4, 0.5]$	46,968	1.71	0.532	0.694
$(0.5, 0.6]$	52,921	1.69	0.525	0.690
$(0.6, 0.7]$	10,783	1.22	0.519	0.673
$(0.7, 1]$	9	0.00	0.504	0.594

F.3 Separation for QM9 population pairs

Table 12 reports the FRC separation for every population pair of § D.2; the ORC hydrogen vs heavy column appears in Table 9.

Table 12: FRC separation S_{sep} for every population pair of § D.2, on the 19 QM9 constructions. A dash marks a pair for which one population is empty under that construction: the ϵ-based construction at 1.5 Å places no non-bond and no H—Hedges, and the rigidity constructions at $p = -1$ place almost no H—Hedges. Aggregation is the per-molecule mean; the union aggregation gives the same rankings.

Construction	H vs heavy	ring vs chain	terminal vs branch	bond vs non-bond	H-heavy vs heavy–heavy	H—Hvs heavy–heavy
FC	0.00	0.08	0.03	0.00	0.00	0.01
kNN, $k = 2$	0.96	0.73	0.99	0.47	1.78	2.47
kNN, $k = 3$	1.37	1.41	1.46	0.28	1.86	3.96
kNN, $k = 4$	0.97	1.67	1.86	1.28	1.08	3.93
kNN, $k = 5$	0.08	1.11	1.22	2.43	0.81	0.57
ϵ-based, 1.5 Å	0.24	0.27	0.20	–	0.77	–
ϵ-based, 1.8 Å	2.46	1.67	2.19	4.13	3.15	5.73
ϵ-based, 2.5 Å	0.57	0.25	0.77	2.05	0.57	2.20
ϵ-based, 3.0 Å	0.27	0.81	0.74	4.42	0.43	0.57
ϵ-based, 4.0 Å	1.32	2.28	2.23	4.11	1.63	2.34
k_T threshold, $p = 1$	2.49	0.88	1.00	1.05	4.06	3.88
k_T threshold, $p = 0$	0.10	0.31	0.62	3.60	0.40	0.36
k_T threshold, $p = -1$	2.60	0.35	0.25	4.53	5.88	–
Laman, $p = 1$	1.61	0.51	0.51	1.05	1.90	3.77
Laman, $p = -1$	0.36	0.37	0.34	2.21	1.12	–
unique-k, $k = 3$, $p = 1$	0.93	0.56	0.56	1.98	0.32	3.80
unique-k, $k = 3$, $p = -1$	0.72	0.51	0.50	2.80	0.80	–
unique-k, $k = 4$, $p = 0$	0.18	0.36	0.41	2.75	1.24	1.58
unique-k, $k = 4$, $p = -1$	2.01	0.28	0.31	3.46	3.24	–

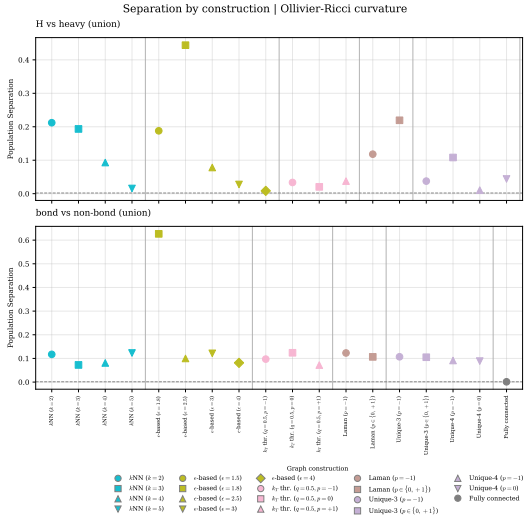


Figure 3: Unconditioned ORCseparation on QM(for all 19 construction on hydrogen vs heavy atom (top) and bond vs non-bond (bottom). Constructions are grouped by family, with the FC at the right. The 1.8 Å ϵ-based graph separates both pairs by a wide margin (0.44 and 0.63), and within each family separation falls as edges are added: along k for kNN, along ϵ for the radius graphs, and from unique-3 to unique-4. The FC sits at 0.002 and 0.001, the residual left when curvature depends on the atom count alone. The 95% bootstrap intervals (§ C.3, molecules resampled) are not drawn because they are narrower than the markers: at most 0.0007 in absolute terms and 1.8% of the plotted value, since every score pools 129, 434 molecules.

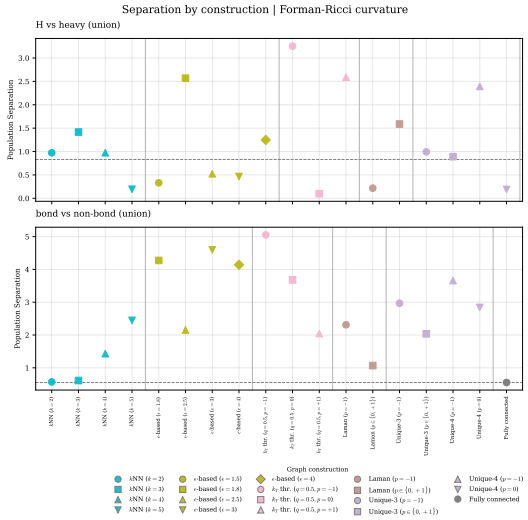


Figure 4: As Figure 3, under FRC. FRC is in curvature units that scale with degree, so its values grow with the number of edges rather than with bond recovery: on hydrogen vs heavy the k_T threshold at $p = 0$ and $p = +1$ (F1 of 0.44 and 0.22) and unique-4 at $p = -1$ score as high as or higher than the 1.8 Å graph, and on bond vs non-bond the 3.0 and 4.0 Å graphs match it. The 95% bootstrap intervals are again narrower than the markers, at most 0.01 in absolute terms and 4.1% of the plotted value, and are not included.

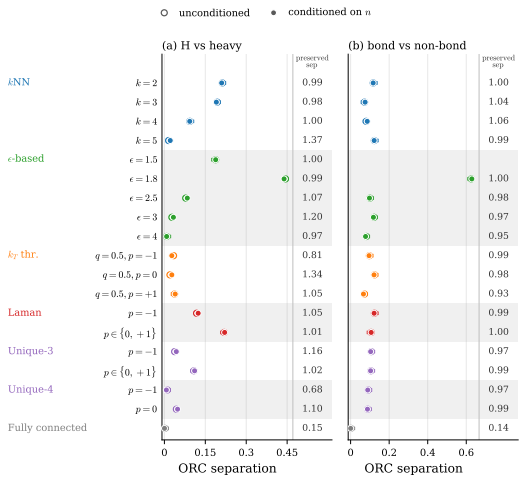


Figure 5: ORC separation for two molecular subsets: (a) hydrogen vs heavy-atom nodes and (b) bonded vs non-bonded edges. The $\epsilon = 1.8$ Å graph gives by far the strongest separation in both. Conditioning on n leaves separation largely unchanged: the preserved fraction is mostly near unity, with larger deviations in (a) for constructions whose separation is already small. The exception is the FC. Its separation is near zero to begin with, and only about 15% survives conditioning, indicating that what little separation it shows is driven mainly by molecule size.

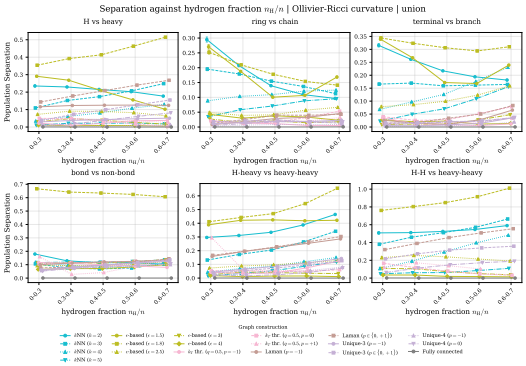


Figure 6: Within-bin ORC separation on QM9 against the hydrogen fraction n $r^{\mathrm{H}} = n_{\mathrm{H}}/n$, for all constructions, union aggregation. One panel per population pair of § D.2; each point is the W_1 between the two populations within one r^{H} bin of Table 4, with 95% bootstrap intervals. On the three hydrogen pairs (top left, middle bottom, and bottom right) its separation rises with r^{H}, as does that of Laman and kNN at $k = 3$, whereas kNN at $k = 2$ falls. Ring vs chain and terminal vs branch are separated clearly only by $\epsilon = 1.8$Å and kNN at $k = 2$, and ring vs chain falls since molecules with a lot of hydrogen have fewer rings (Table 11). Bond vs non-bond is flat for every construction. FC sits at 0 throughout.

Table 13: Graph sparsity by construction method, averaged over 129,434 molecules. Rows are ordered by the mean number of edges per molecule. The final column gives the ratio of graph edges to chemical bonds.

Graph type	Edges / mol	Mean degree	Edges / bond
kNN ($k = 2$)	21.855	2.425	1.176
kNN ($k = 3$)	33.084	3.671	1.779
kNN ($k = 4$)	45.005	4.986	2.416
kNN ($k = 5$)	55.087	6.107	2.960
ϵ-based ($\epsilon = 1.5$ Å)	14.401	1.616	0.785
ϵ-based ($\epsilon = 1.8$ Å)	22.710	2.493	1.209
ϵ-based ($\epsilon = 2.5$ Å)	50.255	5.529	2.675
ϵ-based ($\epsilon = 3$ Å)	74.192	8.082	3.913
ϵ-based ($\epsilon = 4$ Å)	112.892	12.256	5.927
Laman ($p = 0$)	33.074	3.658	1.774
Laman ($p = -1$)	33.074	3.658	1.774
Laman ($p = +1$)	33.074	3.658	1.774
Unique-2 ($p = -1$)	33.074	3.658	1.774
Unique-3 ($p = 0$)	48.112	5.315	2.578
Unique-3 ($p = +1$)	48.112	5.315	2.578
Unique-3 ($p = -1$)	48.112	5.315	2.578
Unique-4 ($p = 0$)	62.149	6.858	3.326
Unique-4 ($p = -1$)	62.149	6.858	3.326
k_T threshold ($p = +1$)	66.920	7.318	3.547
k_T threshold ($p = 0$)	67.920	7.427	3.600
k_T threshold ($p = -1$)	70.362	7.641	3.706
Fully connected	158.003	17.037	8.263

Table 14: Spearman rank correlation between bond recovery F1 and curvature separation (per molecule mean) over the QM9 constructions, for every population pair of § D.2. n is the number of constructions on which the pair is defined (H–H vs heavy- -heavy edges and bond vs non-bond are undefined where a construction places no such edges,Table 12). The second value in each cell excludes the complete graph. ORC tracks recovery on every pair except bond vs non-bond. FRC tracks it on none of the node pairs.

Pair	n	ORC ρ	FRC ρ
H vs heavy	19	0.88 / 0.87	0.09 / −0.03
ring vs chain	19	0.82 / 0.80	0.13 / 0.00
terminal vs branch	19	0.76 / 0.73	0.10 / −0.04
bond vs non-bond	18	0.28 / 0.15	−0.25 / −0.48
H-heavy vs heavy–heavy	19	0.82 / 0.81	0.09 / −0.04
H—Hvs heavy–heavy	14	0.92 / 0.92	0.58 / 0.53

F.4 Separation figures for JetSet

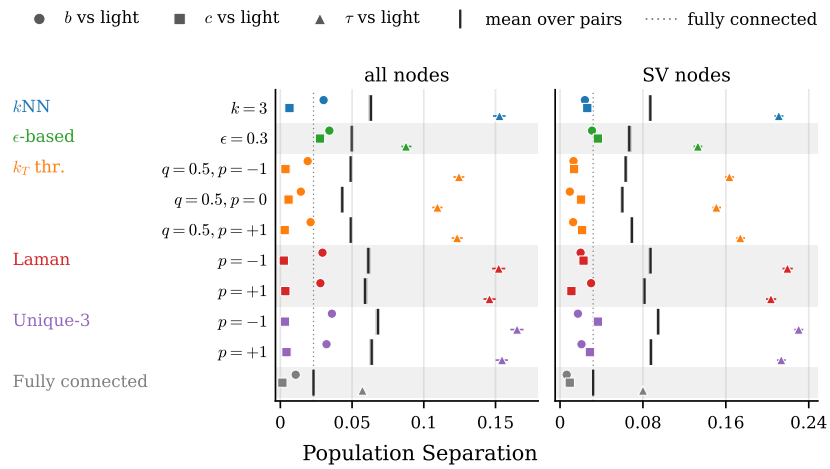


Figure 7: Unconditioned ORC separation by flavor pair. Pair terms of Equation 4 (markers) and their mean (bars) for the default constructions over all nodes (left) and SV nodes (right), union aggregation. The dotted line is the FC's mean.

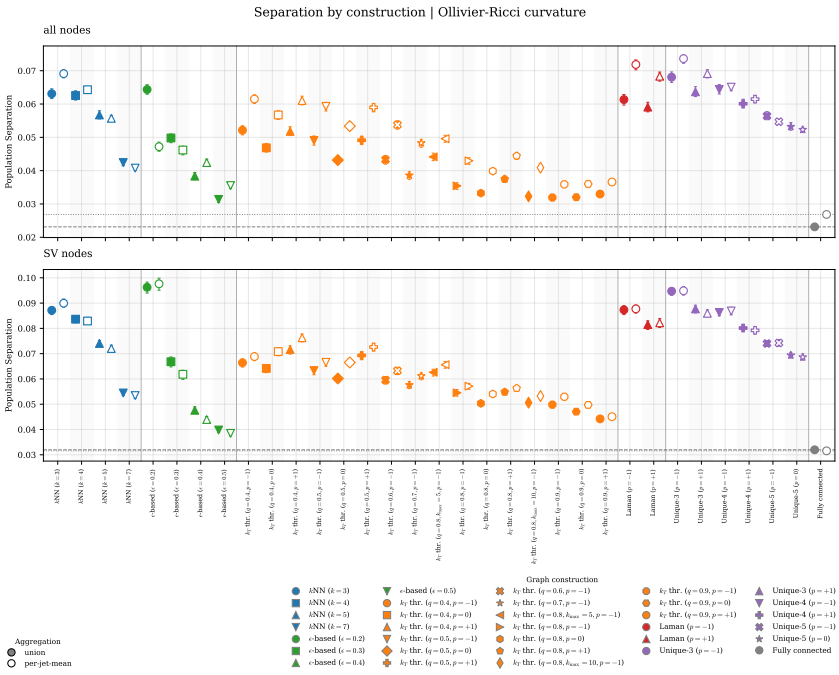


Figure 8: Unconditioned ORC separation by construction on SV nodes (top) and PV-SV edges (bottom), mean over the three flavor pairs, for all 33 constructions under union (filled) and per-jet-mean (open) aggregation, The horizontal lines are the FC graph. The two aggregations agree on the ordering except on PV-SV edges, where the per-jet mean raises the sparse constructions.

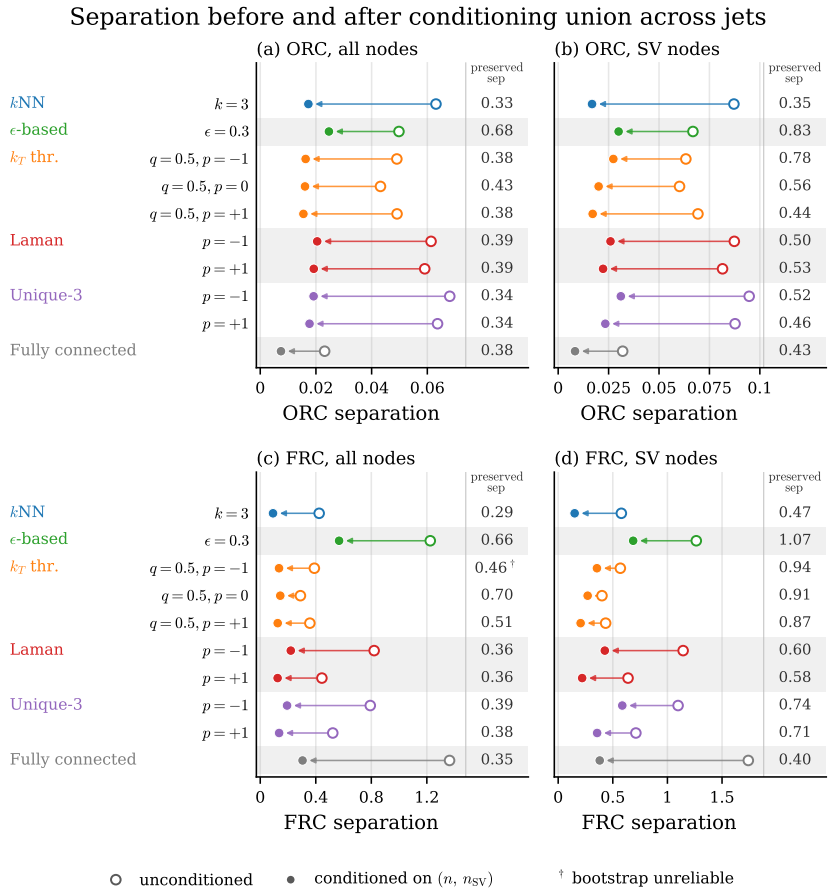


Figure 9: **Flavor separation before and after conditioning on size and composition.** Open circles: pooled separation; filled circles: binned separation at fixed (n, n_{sv}) (Equation 5) averaged over the three flavor pairs. The right column is preserved-sep$_{(n,n_{\text{sv}})}(g)$ (Equation 6). (a) ORC over all nodes, (b) ORC over SV nodes, (c) FRC over all nodes, (d) FRC over SV nodes. Rows are grouped by construction, with parameters as labeled. The FC graph marks the level left when no separation survives. FRC axes are not comparable across constructions, but their ratio is. †: bootstrap unreliable near the W_1 bias floor, read as an upper bound (§ C.3). Intervals are 95% bootstrap; bins and coverage are in § C.2.

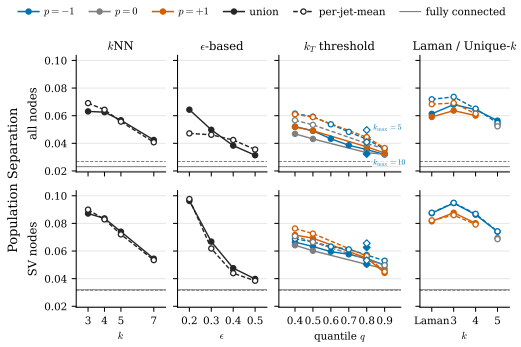


Figure 10: Population separation of ORC across graph construction methods. Columns show graph constructions kNN, ϵ-based, k_T threshold, and Laman/unique-k graphs, swept over their respective parameters (k, ϵ, quantile q, and k). The top row uses curvature from all nodes; the bottom row restricts to SV ndoes. Colors indicate the clustering exponent $p \in \{-1, 0, +1\}$ used to build the constructions where applicable. Solid lines with filled markers give the separation of the pooled (union) curvature distribution; dashed lines with open markers give the per-jet mean. Horizontal gray lines mark the fully connected baseline (solid: union, dashed: per-jet mean). Every sparse construction exceeds the fully connected baseline, with separation generally decreasing as graphs become denser, and SV-node curvature separating the populations more strongly than all-node curvature.

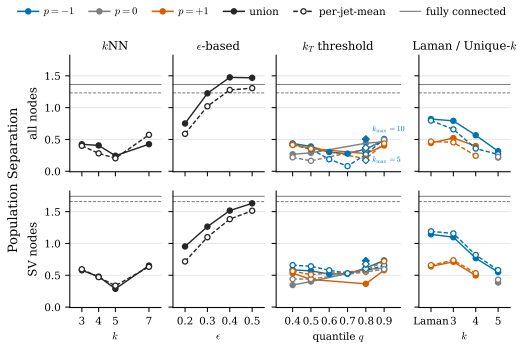


Figure 11: Population separation of FRC across graph construction methods. This has the same conventions and layout as Figure 10. Compared to ORC, fully connected graph gives highest separation for FRC. Only the ϵ-based graphs approach this baseline, with separation rising steadily with ϵ. kNN, k_T-threshold, Laman and unique-k constructions remain well below it across all parameters. As for ORC, SV node curvature gives larger separation than all nodes curvatures.

F.5 SEPARATION: UNION AGGREGATION ACROSS JETS PER BIN

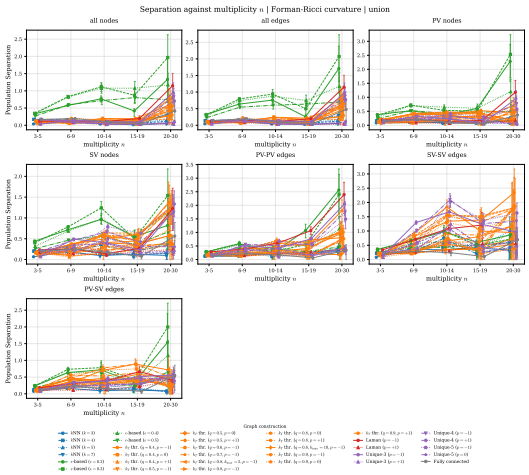


Figure 12: Within-bin FRC separation against multiplicity, all constructions, union aggregation. One panel per structural subset; each point is the mean over the three flavor pairs of the within-bin W_1, with 95% bootstrap intervals. The FRC separation of the ϵ-based graphs grows with n on all nodes and all edges, since degree under a fixed ϵ-based grows with n, whereas the other families stay near the FC until the sparsest bin (n = 20–30), where intervals widen.

F.6 SEPARATION: MEAN PER JET AGGREGATION ACROSS JETS PER BIN

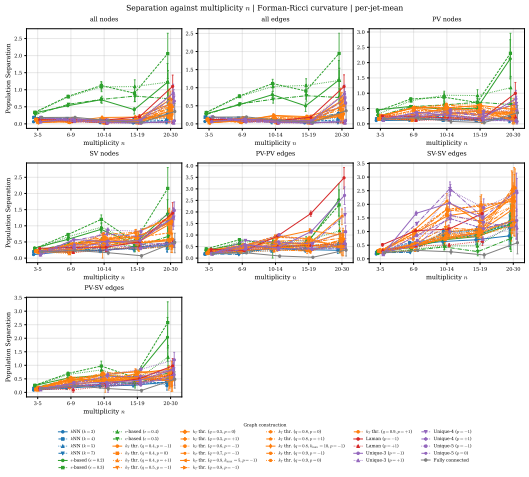


Figure 13: Within-bin FRC separation against multiplicity, all constructions, per-jet-mean aggregation. Panels and markers as in Figure 12; each jet contributes one mean per structural subset. The ordering matches the union aggregation: the ϵ-based graphs lead on all nodes and all edges, and the intervals widen in the sparsest bin.

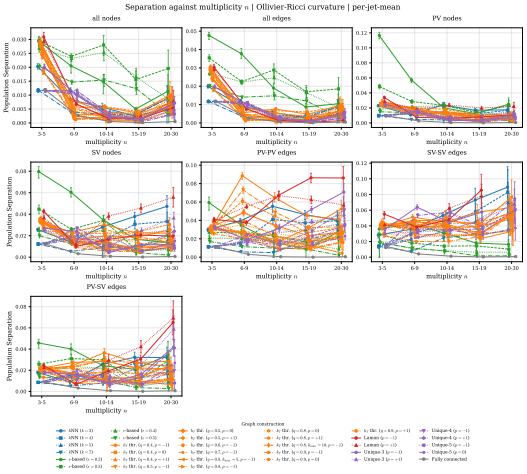


Figure 14: Within-bin ORC separation against multiplicity, all constructions, per-jet-mean aggregation. Panels and markers as in Figure 20; each jet contributes one mean per structural subset. The ϵ-based graphs at $\epsilon = 0.2$ and 0.3 stay above the other constructions on all nodes and all edges from $n \geq 6$, and the complete graph falls to the W_1 noise level from $n \geq 10$.

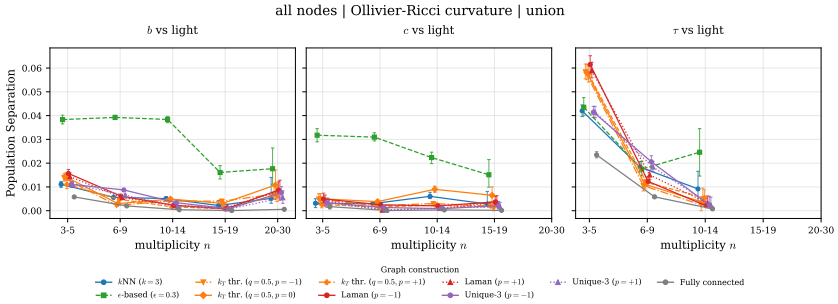


Figure 15: Within-bin ORC separation against multiplicity, by flavor pair. Each panel is one pair, each curve one default construction, and each point the W_1 between the two flavors' all nodes ORC distributions within one bin of n (union aggregation), with 95% bootstrap intervals. Bins in which a flavor has fewer than $N_{\min}$ jets are omitted, which is why the c and τ curves end early. The ϵ-based graph leads every bin on b and c vs light. On τ vs light the rigid and k_T graphs lead at $n \leq 5$ and the constructions tie at 6–9. The FC falls below 0.001, and the W_1 noise level, from $n \geq 10$.

F.7 Separation: union aggregation across jets

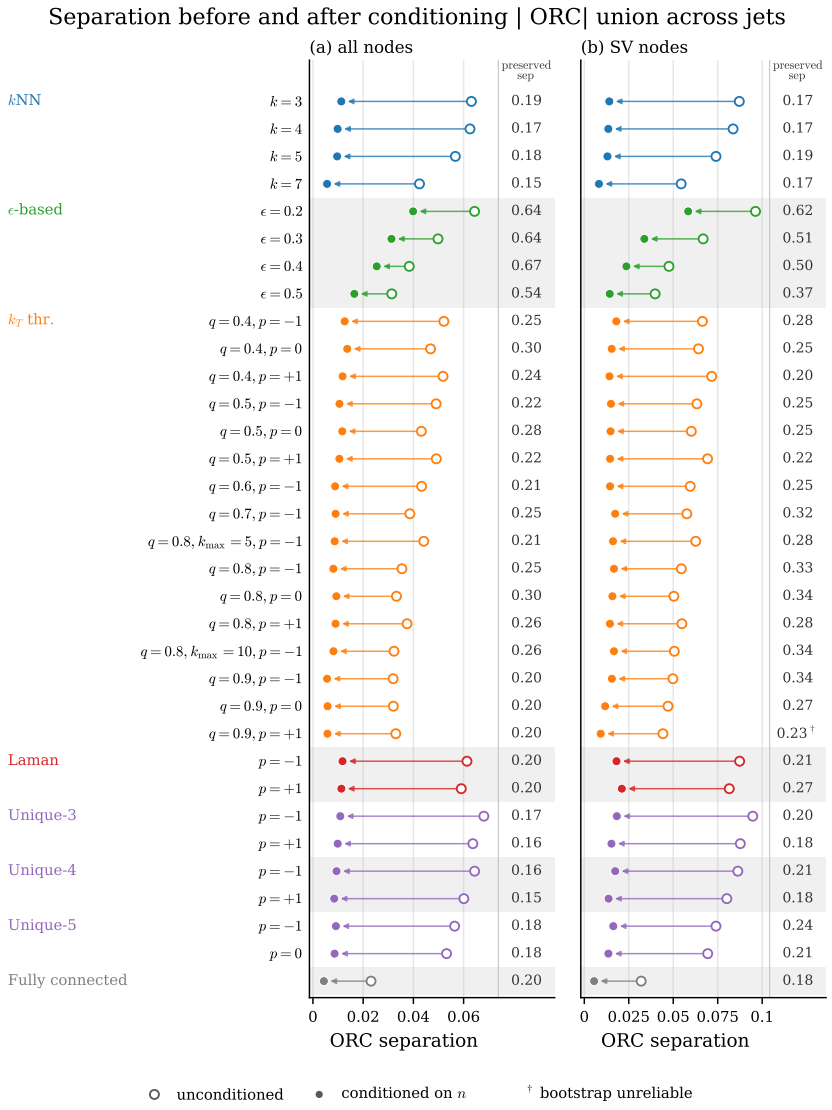


Figure 16: **ORC separation for JetSet at fixed multiplicity.** Population separation over all nodes, computed within bins of multiplicity n and averaged over the three flavor pairs for *union* aggregation on default graphs. Error bars are 95% bootstrap intervals ($B = 500$, jets resampled within flavor). Here points within a bin are offset for legibility. The complete graph (gray) is the control: its curvature is a function of n alone, so its value in a bin is the separation that the width of that bin lets through. The ϵ-based construction stays above all others in every bin. The rest fall toward the control as n grows.

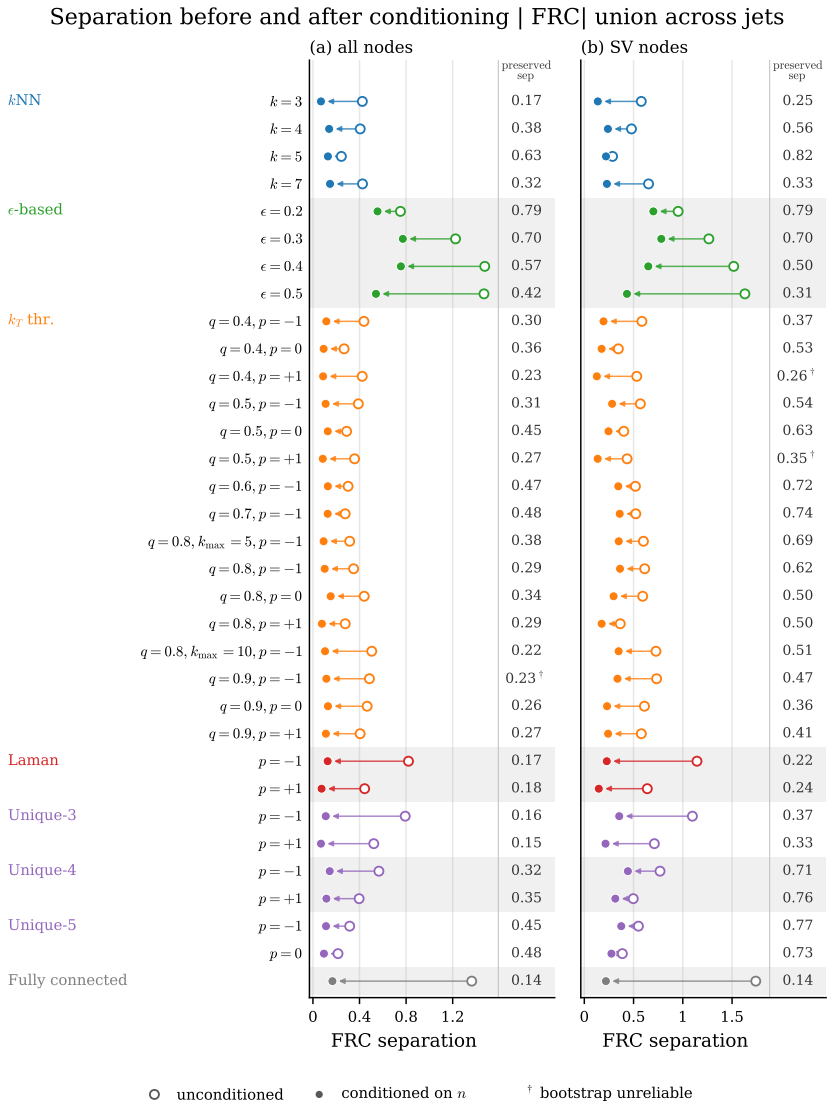


Figure 17: **FRC separation for JetSet at fixed multiplicity.** Population separation over all nodes, computed within bins of multiplicity n and averaged over the three flavor pairs for *union* aggregation on default graphs. Error bars are 95% bootstrap intervals ($B = 500$, jets resampled within flavor). Here points within a bin are offset for legibility. The complete graph (gray) is the control: its curvature is a function of n alone, so its value in a bin is the separation that the width of that bin lets through. The ϵ-based construction stays above all others in every bin. The rest fall toward the control as n grows.

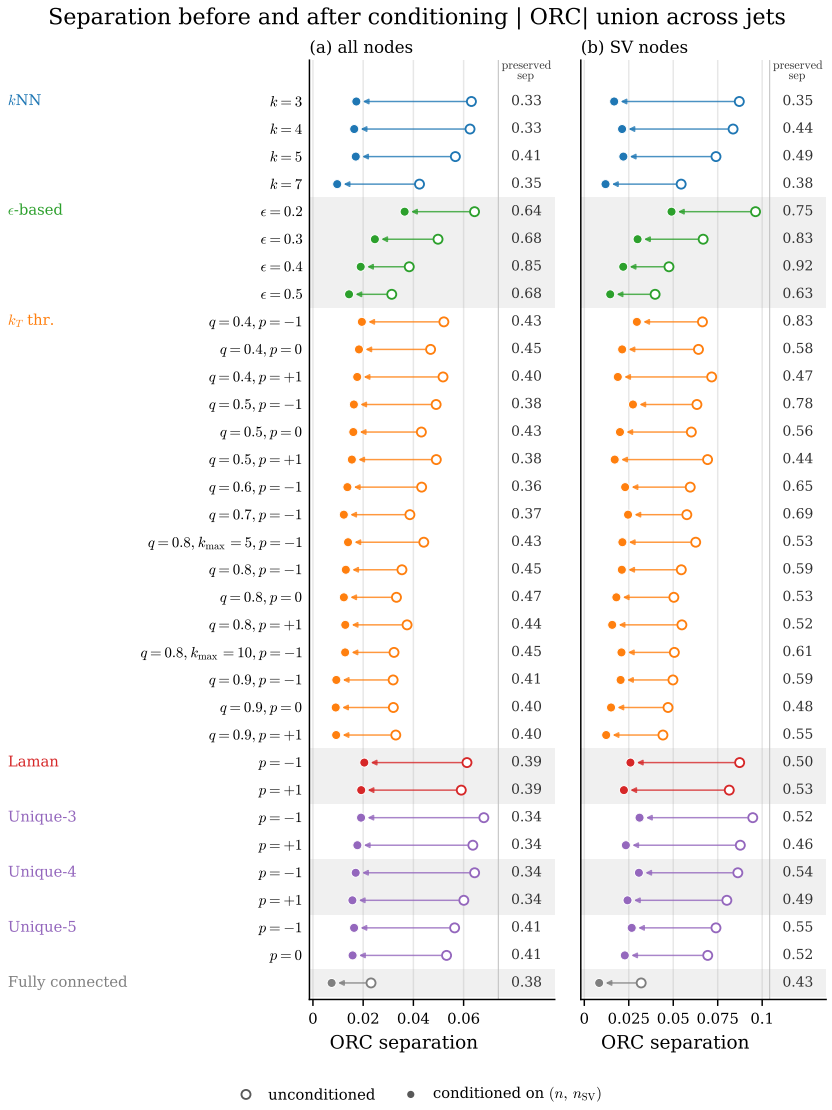


Figure 18: **ORC separation for JetSet at fixed joint composition** (n, n_{sv})**.** Population separation over all nodes, computed within bins of multiplicity (n, n_{sv}) and averaged over the three flavor pairs for *union* aggregation on default graphs. Error bars are 95% bootstrap intervals ($B = 500$, jets resampled within flavor). Here points within a bin are offset for legibility. The complete graph (gray) is the control: its curvature is a function of n alone, so its value in a bin is the separation that the width of that bin lets through. The ϵ-based construction stays above all others in every bin. The rest fall toward the control as n grows.

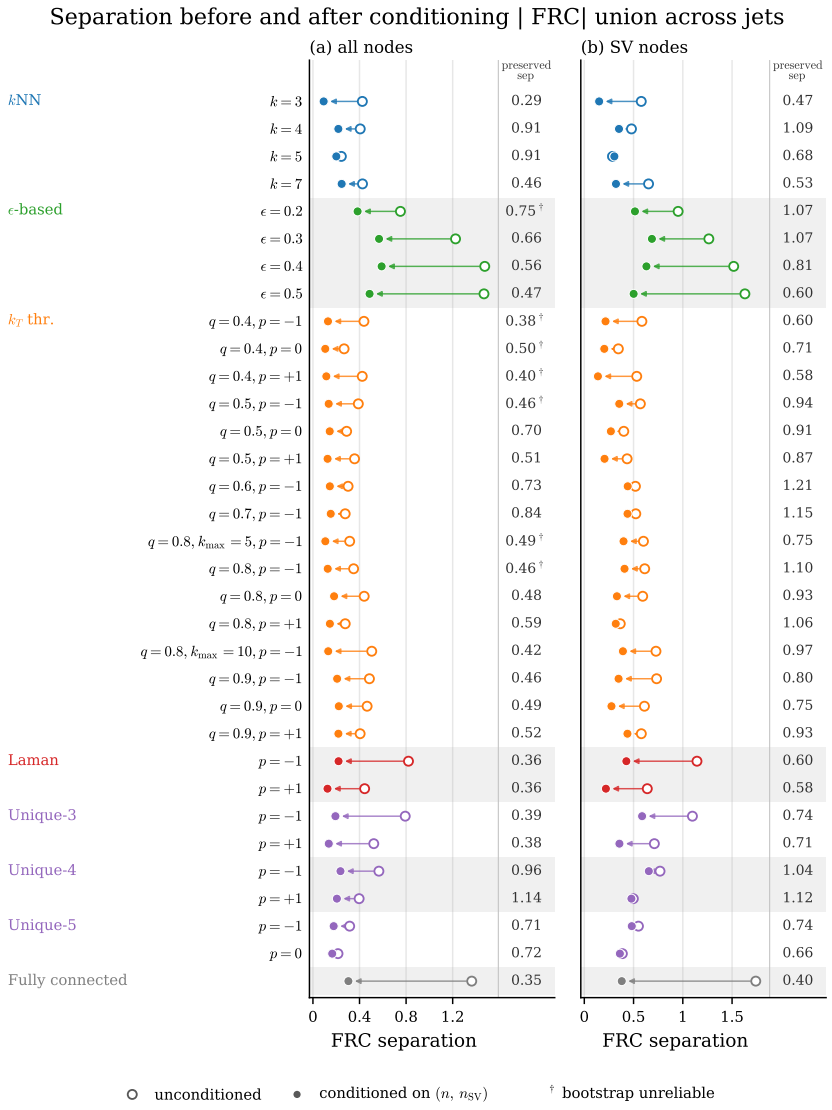


Figure 19: **FRC separation for JetSet at fixed joint composition** (n, n_{sv})**.** Population separation over all nodes, computed within bins of multiplicity (n, n_{sv}) and averaged over the three flavor pairs for *union* aggregation on default graphs. Error bars are 95% bootstrap intervals ($B = 500$, jets resampled within flavor). Here points within a bin are offset for legibility. The complete graph (gray) is the control: its curvature is a function of n alone, so its value in a bin is the separation that the width of that bin lets through. The ϵ-based construction stays above all others in every bin. The rest fall toward the control as n grows.

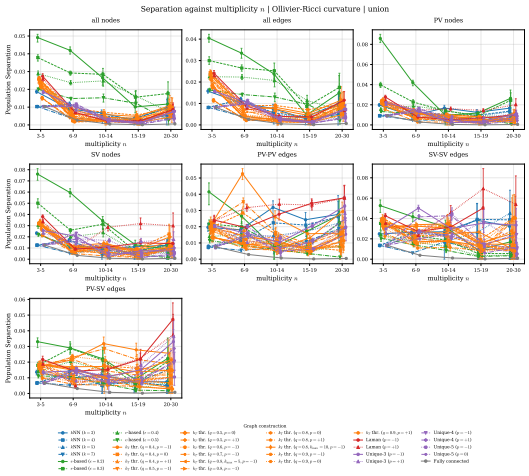


Figure 20: Within-bin ORC separation against multiplicity, all constructions, union aggregation. One panel per structural subset; each point is the mean over the three flavor pairs of the within-bin W_1, with 95% bootstrap intervals. The ϵ-based graphs at $\epsilon = 0.2$ and 0.3 stay above every other construction on all nodes and all edges from $n \geq 6$, and the complete graph falls to the W_1 noise level from $n \geq 10$.

F.8 Separation: mean per jet aggregation



Figure 21: **ORC separation for JetSet at fixed joint composition n.** Population separation over all nodes, computed within bins of multiplicity n and averaged over the three flavor pairs for *mean per jet* aggregation on default graphs. Error bars are 95% bootstrap intervals ($B = 500$, jets resampled within flavor). Here points within a bin are offset for legibility. The complete graph (gray) is the control: its curvature is a function of (n, n_{sv}) alone, so its value in a bin is the separation that the width of that bin lets through. The ϵ-based construction stays above all others in every bin. The rest fall toward the control as n grows.

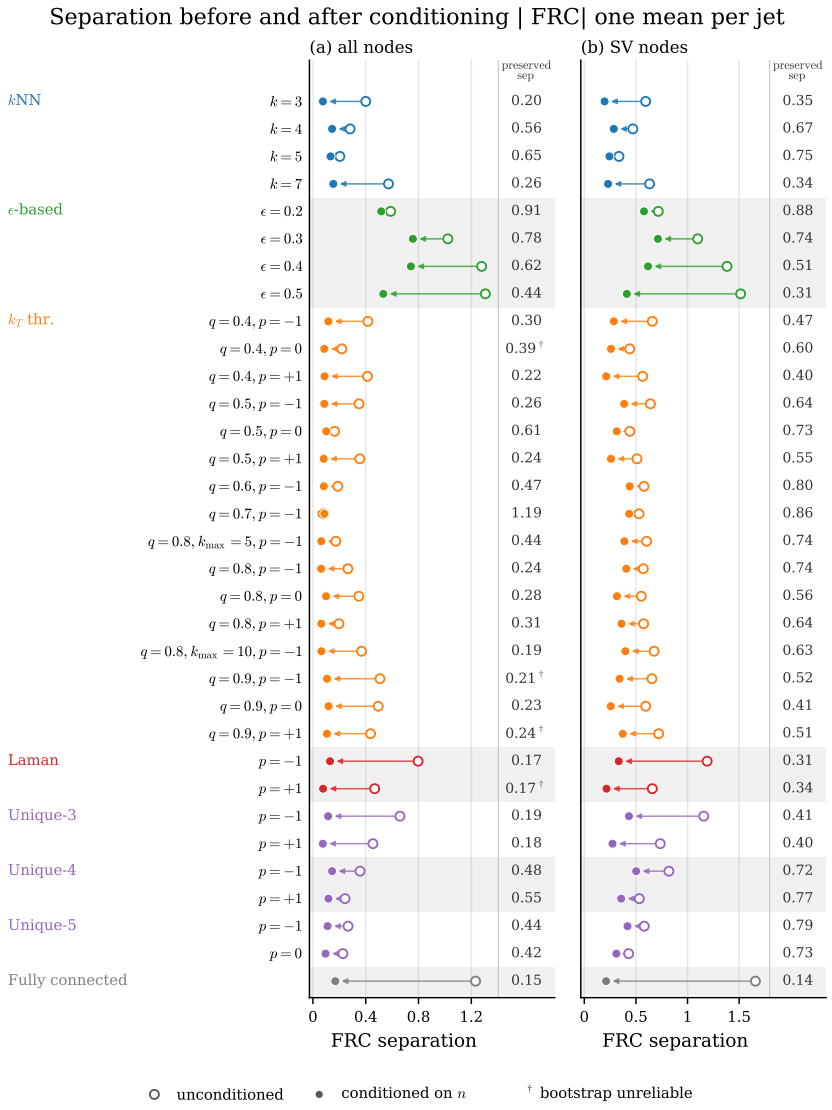


Figure 22: **FRC separation for JetSet at fixed joint composition n.** Population separation over all nodes, computed within bins of multiplicity n and averaged over the three flavor pairs for *mean per jet* aggregation on default graphs. Error bars are 95% bootstrap intervals ($B = 500$, jets resampled within flavor). Here points within a bin are offset for legibility. The complete graph (gray) is the control: its curvature is a function of (n, n_{sv}) alone, so its value in a bin is the separation that the width of that bin lets through. The ϵ-based construction stays above all others in every bin. The rest fall toward the control as n grows.

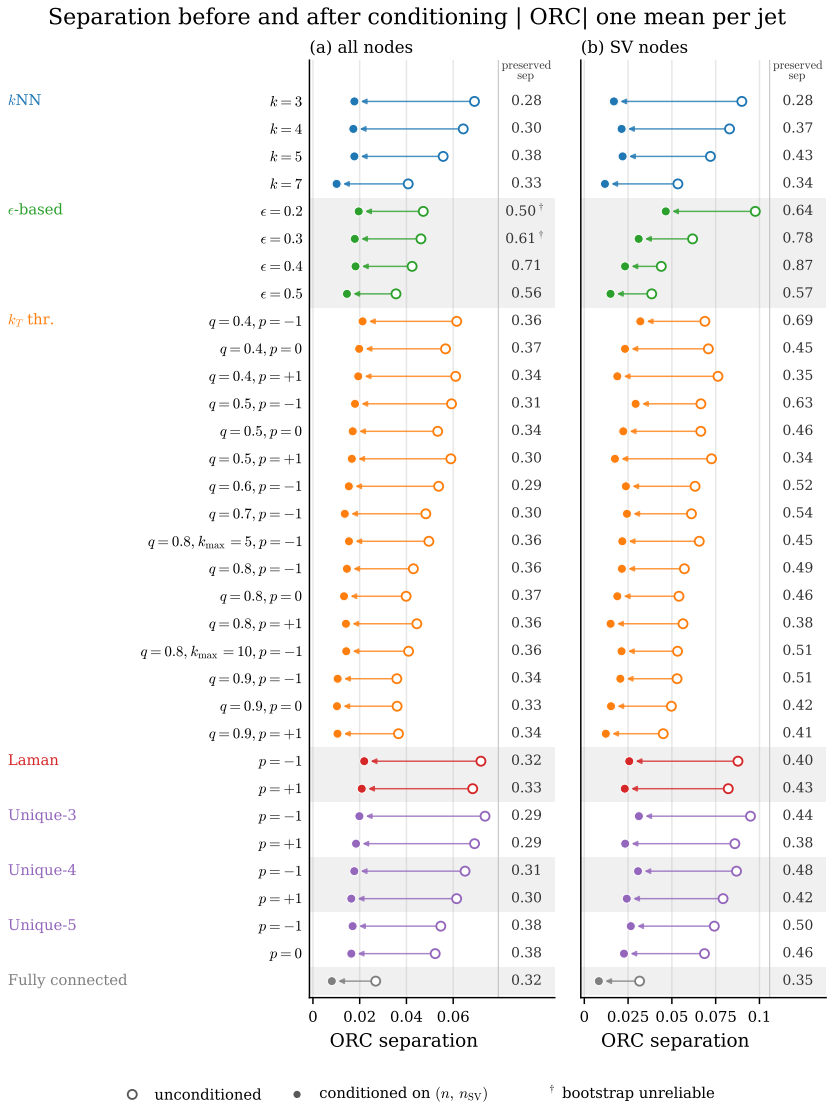


Figure 23: **ORC separation for JetSet at fixed joint composition** (n, n_{sv})**.** Population separation over all nodes, computed within bins of multiplicity (n, n_{sv}) and averaged over the three flavor pairs for *mean per jet* aggregation on default graphs. Error bars are 95% bootstrap intervals ($B = 500$, jets resampled within flavor). Here points within a bin are offset for legibility. The complete graph (gray) is the control: its curvature is a function of (n, n_{sv}) alone, so its value in a bin is the separation that the width of that bin lets through. The ϵ-based construction stays above all others in every bin. The rest fall toward the control as n grows.

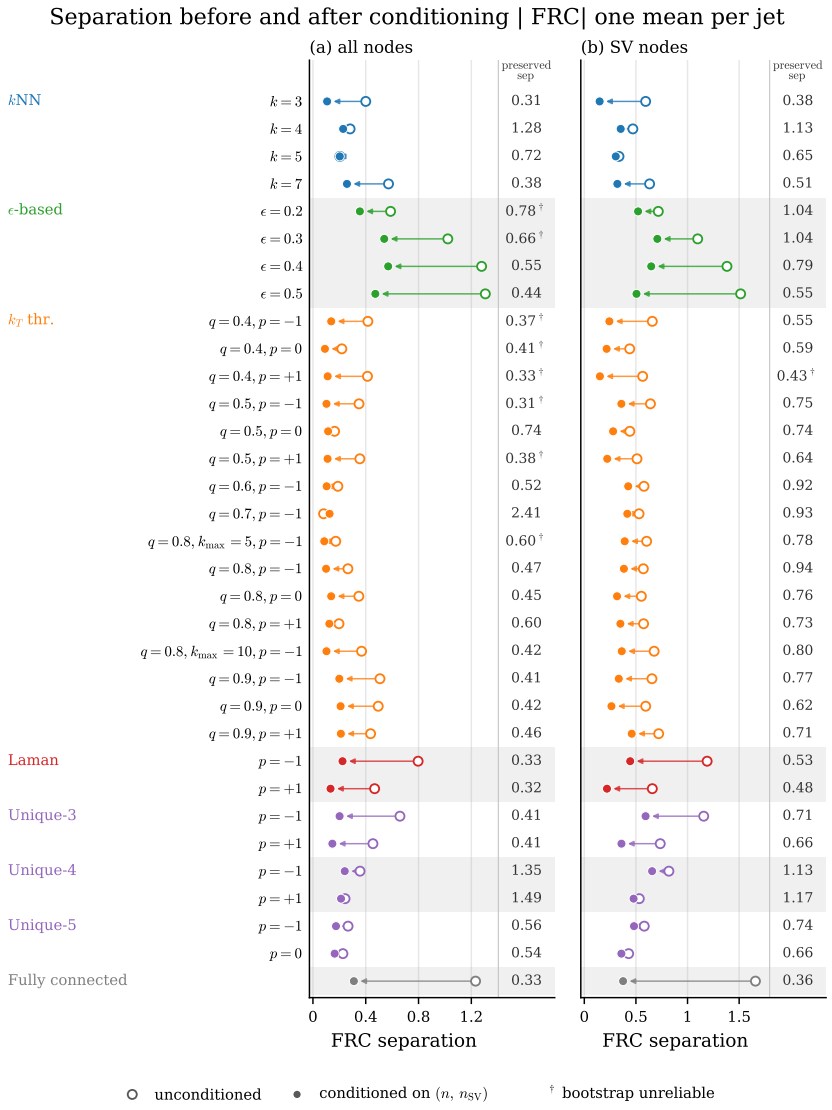


Figure 24: **FRC separation for JetSet at fixed joint composition** (n, n_{sv})**.** Population separation over all nodes, computed within bins of multiplicity (n, n_{sv}) and averaged over the three flavor pairs for *mean per jet* aggregation on default graphs. Error bars are 95% bootstrap intervals ($B = 500$, jets resampled within flavor). Here points within a bin are offset for legibility. The complete graph (gray) is the control: its curvature is a function of n alone, so its value in a bin is the separation that the width of that bin lets through. The ϵ-based construction stays above all others in every bin. The rest fall toward the control as n grows.

F.9 SEPARATION TABLES FOR JETSET

Table 15: Unconditioned separation S_{sep} under Ollivier-Ricci curvature (union aggregation). Bold marks the most separating construction per subset, excluding FC. Sub- and superscripts give the bootstrap interval where available.

Construction	all nodes	all edges	PV nodes	SV nodes	PV-PV edges	SV-SV edges	PV-SV edges
kNN ($k = 3$)	$0.063^{+0.001}_{-0.001}$	$0.059^{+0.002}_{-0.001}$	$0.050^{+0.002}_{-0.002}$	$0.087^{+0.002}_{-0.002}$	$0.034^{+0.004}_{-0.004}$	$0.091^{+0.002}_{-0.003}$	$\mathbf{0.065^{+0.002}_{-0.002}}$
kNN ($k = 4$)	$0.063^{+0.001}_{-0.001}$	$0.056^{+0.002}_{-0.001}$	$0.054^{+0.002}_{-0.002}$	$0.084^{+0.001}_{-0.001}$	$0.038^{+0.004}_{-0.004}$	$0.086^{+0.002}_{-0.002}$	$0.064^{+0.002}_{-0.002}$
kNN ($k = 5$)	$0.057^{+0.001}_{-0.001}$	$0.050^{+0.002}_{-0.001}$	$0.051^{+0.002}_{-0.002}$	$0.074^{+0.001}_{-0.001}$	$0.038^{+0.004}_{-0.004}$	$0.076^{+0.002}_{-0.002}$	$0.059^{+0.002}_{-0.002}$
kNN ($k = 7$)	$0.042^{+0.001}_{-0.001}$	$0.037^{+0.001}_{-0.001}$	$0.038^{+0.002}_{-0.002}$	$0.054^{+0.001}_{-0.001}$	$0.030^{+0.004}_{-0.004}$	$0.057^{+0.002}_{-0.002}$	$0.044^{+0.002}_{-0.002}$
ϵ-based ($\epsilon = 0.2$)	$0.064^{+0.001}_{-0.001}$	$\mathbf{0.061^{+0.002}_{-0.002}}$	$0.056^{+0.002}_{-0.002}$	$\mathbf{0.096^{+0.002}_{-0.002}}$	$0.030^{+0.003}_{-0.004}$	$0.081^{+0.004}_{-0.003}$	$0.042^{+0.003}_{-0.003}$
ϵ-based ($\epsilon = 0.3$)	$0.050^{+0.001}_{-0.001}$	$0.037^{+0.001}_{-0.002}$	$0.030^{+0.002}_{-0.002}$	$0.067^{+0.002}_{-0.002}$	$0.016^{+0.002}_{-0.003}$	$0.062^{+0.003}_{-0.003}$	$0.025^{+0.001}_{-0.002}$
ϵ-based ($\epsilon = 0.4$)	$0.038^{+0.001}_{-0.001}$	$0.028^{+0.001}_{-0.001}$	$0.021^{+0.001}_{-0.001}$	$0.048^{+0.001}_{-0.001}$	$0.013^{+0.001}_{-0.002}$	$0.048^{+0.002}_{-0.002}$	$0.022^{+0.001}_{-0.001}$
ϵ-based ($\epsilon = 0.5$)	$0.031^{+0.001}_{-0.001}$	$0.023^{+0.001}_{-0.001}$	$0.020^{+0.001}_{-0.001}$	$0.040^{+0.001}_{-0.001}$	$0.012^{+0.001}_{-0.001}$	$0.043^{+0.001}_{-0.002}$	$0.020^{+0.001}_{-0.001}$
k_T thr. ($q = 0.4, p = -1$)	$0.052^{+0.001}_{-0.001}$	$0.047^{+0.001}_{-0.001}$	$0.043^{+0.002}_{-0.002}$	$0.066^{+0.001}_{-0.002}$	$0.034^{+0.004}_{-0.004}$	$0.073^{+0.002}_{-0.002}$	$0.040^{+0.002}_{-0.002}$
k_T thr. ($q = 0.4, p = 0$)	$0.047^{+0.001}_{-0.001}$	$0.042^{+0.001}_{-0.001}$	$0.037^{+0.002}_{-0.002}$	$0.064^{+0.001}_{-0.002}$	$0.023^{+0.003}_{-0.003}$	$0.066^{+0.002}_{-0.002}$	$0.043^{+0.002}_{-0.002}$
k_T thr. ($q = 0.4, p = +1$)	$0.052^{+0.001}_{-0.001}$	$0.048^{+0.002}_{-0.001}$	$0.042^{+0.002}_{-0.002}$	$0.072^{+0.002}_{-0.002}$	$0.033^{+0.003}_{-0.004}$	$0.086^{+0.002}_{-0.003}$	$0.053^{+0.002}_{-0.003}$
k_T thr. ($q = 0.5, p = -1$)	$0.049^{+0.001}_{-0.001}$	$0.041^{+0.001}_{-0.001}$	$0.042^{+0.002}_{-0.002}$	$0.063^{+0.001}_{-0.002}$	$0.026^{+0.003}_{-0.003}$	$0.064^{+0.002}_{-0.002}$	$0.035^{+0.002}_{-0.002}$
k_T thr. ($q = 0.5, p = 0$)	$0.043^{+0.001}_{-0.001}$	$0.036^{+0.001}_{-0.001}$	$0.035^{+0.002}_{-0.002}$	$0.060^{+0.001}_{-0.002}$	$0.022^{+0.003}_{-0.003}$	$0.059^{+0.002}_{-0.002}$	$0.036^{+0.002}_{-0.002}$
k_T thr. ($q = 0.5, p = +1$)	$0.049^{+0.001}_{-0.001}$	$0.042^{+0.001}_{-0.001}$	$0.041^{+0.002}_{-0.002}$	$0.069^{+0.001}_{-0.002}$	$0.032^{+0.003}_{-0.003}$	$0.086^{+0.002}_{-0.003}$	$0.048^{+0.002}_{-0.002}$
k_T thr. ($q = 0.6, p = -1$)	$0.043^{+0.001}_{-0.001}$	$0.034^{+0.001}_{-0.001}$	$0.038^{+0.002}_{-0.002}$	$0.060^{+0.001}_{-0.002}$	$0.021^{+0.003}_{-0.003}$	$0.061^{+0.002}_{-0.002}$	$0.030^{+0.002}_{-0.002}$
k_T thr. ($q = 0.7, p = -1$)	$0.039^{+0.001}_{-0.001}$	$0.029^{+0.001}_{-0.001}$	$0.034^{+0.001}_{-0.002}$	$0.058^{+0.001}_{-0.001}$	$0.021^{+0.002}_{-0.002}$	$0.061^{+0.002}_{-0.002}$	$0.027^{+0.001}_{-0.002}$
k_T thr. ($q = 0.8, k_{\max} = 5, p = -1$)	$0.044^{+0.001}_{-0.001}$	$0.037^{+0.001}_{-0.001}$	$0.039^{+0.002}_{-0.002}$	$0.063^{+0.001}_{-0.001}$	$0.027^{+0.003}_{-0.003}$	$0.072^{+0.002}_{-0.002}$	$0.039^{+0.002}_{-0.002}$
k_T thr. ($q = 0.8, p = -1$)	$0.035^{+0.001}_{-0.001}$	$0.027^{+0.001}_{-0.001}$	$0.032^{+0.001}_{-0.001}$	$0.055^{+0.001}_{-0.001}$	$0.022^{+0.002}_{-0.002}$	$0.063^{+0.002}_{-0.002}$	$0.028^{+0.001}_{-0.001}$
k_T thr. ($q = 0.8, p = 0$)	$0.033^{+0.001}_{-0.001}$	$0.025^{+0.001}_{-0.001}$	$0.029^{+0.001}_{-0.001}$	$0.050^{+0.001}_{-0.001}$	$0.020^{+0.002}_{-0.002}$	$0.052^{+0.002}_{-0.002}$	$0.027^{+0.001}_{-0.001}$
k_T thr. ($q = 0.8, p = +1$)	$0.037^{+0.001}_{-0.001}$	$0.030^{+0.001}_{-0.001}$	$0.032^{+0.001}_{-0.001}$	$0.055^{+0.001}_{-0.001}$	$0.022^{+0.002}_{-0.002}$	$0.068^{+0.002}_{-0.002}$	$0.034^{+0.002}_{-0.002}$
k_T thr. ($q = 0.8, k_{\max} = 10, p = -1$)	$0.032^{+0.001}_{-0.001}$	$0.022^{+0.001}_{-0.001}$	$0.029^{+0.001}_{-0.001}$	$0.051^{+0.001}_{-0.001}$	$0.018^{+0.001}_{-0.002}$	$0.060^{+0.002}_{-0.002}$	$0.023^{+0.001}_{-0.001}$
k_T thr. ($q = 0.9, p = -1$)	$0.032^{+0.001}_{-0.001}$	$0.026^{+0.001}_{-0.001}$	$0.028^{+0.001}_{-0.001}$	$0.050^{+0.001}_{-0.001}$	$0.024^{+0.002}_{-0.002}$	$0.062^{+0.003}_{-0.003}$	$0.029^{+0.001}_{-0.001}$
k_T thr. ($q = 0.9, p = 0$)	$0.032^{+0.001}_{-0.001}$	$0.026^{+0.001}_{-0.001}$	$0.028^{+0.001}_{-0.001}$	$0.047^{+0.001}_{-0.001}$	$0.022^{+0.003}_{-0.003}$	$0.052^{+0.002}_{-0.002}$	$0.029^{+0.001}_{-0.001}$
k_T thr. ($q = 0.9, p = +1$)	$0.033^{+0.001}_{-0.001}$	$0.027^{+0.001}_{-0.001}$	$0.028^{+0.001}_{-0.001}$	$0.044^{+0.001}_{-0.001}$	$0.021^{+0.003}_{-0.003}$	$0.051^{+0.002}_{-0.002}$	$0.030^{+0.001}_{-0.001}$
Laman ($p = -1$)	$0.061^{+0.002}_{-0.002}$	$0.059^{+0.002}_{-0.002}$	$0.051^{+0.002}_{-0.002}$	$0.087^{+0.002}_{-0.002}$	$0.034^{+0.004}_{-0.004}$	$0.096^{+0.002}_{-0.002}$	$0.054^{+0.002}_{-0.002}$
Laman ($p = +1$)	$0.059^{+0.001}_{-0.001}$	$0.058^{+0.002}_{-0.001}$	$0.046^{+0.002}_{-0.002}$	$0.081^{+0.001}_{-0.002}$	$0.033^{+0.004}_{-0.004}$	$0.095^{+0.002}_{-0.002}$	$0.057^{+0.002}_{-0.003}$
Unique-3 ($p = -1$)	$\mathbf{0.068^{+0.002}_{-0.002}}$	$0.060^{+0.002}_{-0.002}$	$\mathbf{0.061^{+0.002}_{-0.003}}$	$0.095^{+0.001}_{-0.002}$	$0.044^{+0.004}_{-0.004}$	$\mathbf{0.104^{+0.002}_{-0.002}}$	$0.062^{+0.003}_{-0.003}$
Unique-3 ($p = +1$)	$0.064^{+0.002}_{-0.001}$	$0.057^{+0.002}_{-0.001}$	$0.054^{+0.002}_{-0.002}$	$0.088^{+0.001}_{-0.002}$	$0.040^{+0.004}_{-0.004}$	$0.094^{+0.002}_{-0.002}$	$0.062^{+0.002}_{-0.002}$
Unique-4 ($p = -1$)	$0.064^{+0.001}_{-0.001}$	$0.056^{+0.002}_{-0.001}$	$0.060^{+0.003}_{-0.003}$	$0.086^{+0.001}_{-0.001}$	$\mathbf{0.044^{+0.004}_{-0.004}}$	$0.091^{+0.002}_{-0.002}$	$0.064^{+0.002}_{-0.003}$
Unique-4 ($p = +1$)	$0.060^{+0.001}_{-0.001}$	$0.052^{+0.002}_{-0.001}$	$0.054^{+0.002}_{-0.002}$	$0.080^{+0.001}_{-0.001}$	$0.041^{+0.004}_{-0.004}$	$0.085^{+0.002}_{-0.002}$	$0.060^{+0.002}_{-0.002}$
Unique-5 ($p = -1$)	$0.056^{+0.001}_{-0.001}$	$0.049^{+0.002}_{-0.001}$	$0.053^{+0.002}_{-0.002}$	$0.074^{+0.001}_{-0.001}$	$0.041^{+0.004}_{-0.004}$	$0.077^{+0.002}_{-0.002}$	$0.058^{+0.002}_{-0.002}$
Unique-5 ($p = 0$)	$0.053^{+0.001}_{-0.001}$	$0.046^{+0.002}_{-0.001}$	$0.049^{+0.002}_{-0.002}$	$0.069^{+0.001}_{-0.001}$	$0.038^{+0.004}_{-0.004}$	$0.073^{+0.002}_{-0.002}$	$0.054^{+0.002}_{-0.002}$
Fully connected	$0.023^{+0.001}_{-0.001}$	$0.017^{+0.001}_{-0.001}$	$0.019^{+0.001}_{-0.001}$	$0.032^{+0.001}_{-0.001}$	$0.012^{+0.001}_{-0.001}$	$0.034^{+0.001}_{-0.001}$	$0.020^{+0.001}_{-0.001}$

Table 16: Unconditioned separation S_{sep} under Forman-Ricci curvature (union aggregation). Bold marks the most separating construction per subset, excluding FC. Sub- and superscripts give the bootstrap interval where available.

Construction	all nodes	all edges	PV nodes	SV nodes	PV-PV edges	SV-SV edges	PV-SV edges
kNN ($k = 3$)	$0.424^{+0.011}_{-0.010}$	$0.430^{+0.013}_{-0.011}$	$0.362^{+0.016}_{-0.019}$	$0.578^{+0.012}_{-0.010}$	$0.309^{+0.027}_{-0.031}$	$0.638^{+0.020}_{-0.022}$	$0.517^{+0.021}_{-0.018}$
kNN ($k = 4$)	$0.406^{+0.012}_{-0.013}$	$0.451^{+0.015}_{-0.014}$	$0.405^{+0.019}_{-0.019}$	$0.479^{+0.012}_{-0.014}$	$0.425^{+0.038}_{-0.039}$	$0.554^{+0.018}_{-0.022}$	$0.587^{+0.022}_{-0.020}$
kNN ($k = 5$)	$0.244^{+0.011}_{-0.010}$	$0.368^{+0.016}_{-0.016}$	$0.307^{+0.014}_{-0.016}$	$0.287^{+0.011}_{-0.014}$	$0.432^{+0.040}_{-0.042}$	$0.470^{+0.018}_{-0.018}$	$0.517^{+0.018}_{-0.018}$
kNN ($k = 7$)	$0.426^{+0.017}_{-0.018}$	$0.345^{+0.018}_{-0.017}$	$0.316^{+0.022}_{-0.025}$	$0.652^{+0.017}_{-0.019}$	$0.324^{+0.025}_{-0.030}$	$0.822^{+0.031}_{-0.032}$	$0.455^{+0.022}_{-0.026}$
ϵ-based ($\epsilon = 0.2$)	$0.751^{+0.014}_{-0.016}$	$1.064^{+0.045}_{-0.041}$	$0.684^{+0.031}_{-0.032}$	$0.953^{+0.018}_{-0.021}$	$0.673^{+0.133}_{-0.140}$	$1.262^{+0.046}_{-0.051}$	$1.024^{+0.063}_{-0.059}$
ϵ-based ($\epsilon = 0.3$)	$1.225^{+0.022}_{-0.023}$	$1.448^{+0.031}_{-0.057}$	$1.129^{+0.048}_{-0.050}$	$1.264^{+0.021}_{-0.024}$	$1.052^{+0.157}_{-0.179}$	$1.675^{+0.063}_{-0.076}$	$1.522^{+0.076}_{-0.077}$
ϵ-based ($\epsilon = 0.4$)	$\mathbf{1.475^{+0.026}_{-0.028}}$	$\mathbf{1.621^{+0.059}_{-0.065}}$	$1.413^{+0.063}_{-0.061}$	$1.517^{+0.027}_{-0.028}$	$1.324^{+0.180}_{-0.208}$	$1.891^{+0.069}_{-0.086}$	$\mathbf{1.764^{+0.093}_{-0.088}}$
ϵ-based ($\epsilon = 0.5$)	$1.469^{+0.028}_{-0.030}$	$1.564^{+0.059}_{-0.061}$	$\mathbf{1.416^{+0.060}_{-0.062}}$	$\mathbf{1.630^{+0.031}_{-0.034}}$	$\mathbf{1.342^{+0.161}_{-0.184}}$	$\mathbf{2.006^{+0.077}_{-0.095}}$	$1.734^{+0.093}_{-0.094}$
k_T thr. ($q = 0.4, p = -1$)	$0.438^{+0.014}_{-0.014}$	$0.494^{+0.019}_{-0.020}$	$0.478^{+0.020}_{-0.023}$	$0.585^{+0.020}_{-0.021}$	$0.370^{+0.053}_{-0.059}$	$0.603^{+0.026}_{-0.030}$	$0.564^{+0.024}_{-0.029}$
k_T thr. ($q = 0.4, p = 0$)	$0.268^{+0.008}_{-0.008}$	$0.332^{+0.011}_{-0.011}$	$0.261^{+0.010}_{-0.013}$	$0.346^{+0.012}_{-0.016}$	$0.233^{+0.028}_{-0.031}$	$0.448^{+0.018}_{-0.026}$	$0.439^{+0.021}_{-0.020}$
k_T thr. ($q = 0.4, p = +1$)	$0.423^{+0.014}_{-0.012}$	$0.481^{+0.018}_{-0.018}$	$0.407^{+0.018}_{-0.022}$	$0.532^{+0.014}_{-0.014}$	$0.434^{+0.047}_{-0.057}$	$0.714^{+0.033}_{-0.038}$	$0.560^{+0.026}_{-0.027}$
k_T thr. ($q = 0.5, p = -1$)	$0.390^{+0.011}_{-0.013}$	$0.483^{+0.019}_{-0.020}$	$0.475^{+0.017}_{-0.020}$	$0.569^{+0.019}_{-0.021}$	$0.399^{+0.054}_{-0.057}$	$0.638^{+0.035}_{-0.041}$	$0.591^{+0.026}_{-0.027}$
k_T thr. ($q = 0.5, p = 0$)	$0.289^{+0.007}_{-0.008}$	$0.404^{+0.012}_{-0.013}$	$0.317^{+0.012}_{-0.013}$	$0.401^{+0.015}_{-0.016}$	$0.275^{+0.029}_{-0.039}$	$0.503^{+0.020}_{-0.030}$	$0.496^{+0.023}_{-0.022}$
k_T thr. ($q = 0.5, p = +1$)	$0.357^{+0.011}_{-0.011}$	$0.442^{+0.019}_{-0.018}$	$0.379^{+0.017}_{-0.021}$	$0.434^{+0.012}_{-0.014}$	$0.431^{+0.044}_{-0.057}$	$0.736^{+0.043}_{-0.049}$	$0.527^{+0.028}_{-0.028}$
k_T thr. ($q = 0.6, p = -1$)	$0.302^{+0.008}_{-0.009}$	$0.438^{+0.017}_{-0.018}$	$0.419^{+0.011}_{-0.013}$	$0.519^{+0.016}_{-0.018}$	$0.422^{+0.044}_{-0.052}$	$0.703^{+0.039}_{-0.046}$	$0.619^{+0.026}_{-0.026}$
k_T thr. ($q = 0.7, p = -1$)	$0.277^{+0.006}_{-0.008}$	$0.419^{+0.016}_{-0.017}$	$0.444^{+0.013}_{-0.015}$	$0.523^{+0.015}_{-0.015}$	$0.470^{+0.045}_{-0.052}$	$0.786^{+0.048}_{-0.050}$	$0.628^{+0.029}_{-0.029}$
k_T thr. ($q = 0.8, k_{\max} = 5, p = -1$)	$0.315^{+0.013}_{-0.011}$	$0.463^{+0.021}_{-0.019}$	$0.466^{+0.014}_{-0.018}$	$0.600^{+0.020}_{-0.021}$	$0.672^{+0.041}_{-0.061}$	$0.847^{+0.038}_{-0.047}$	$0.671^{+0.027}_{-0.026}$
k_T thr. ($q = 0.8, p = -1$)	$0.349^{+0.006}_{-0.010}$	$0.445^{+0.018}_{-0.019}$	$0.461^{+0.018}_{-0.020}$	$0.613^{+0.014}_{-0.015}$	$0.575^{+0.042}_{-0.047}$	$0.925^{+0.049}_{-0.045}$	$0.610^{+0.026}_{-0.031}$
k_T thr. ($q = 0.8, p = 0$)	$0.440^{+0.010}_{-0.010}$	$0.465^{+0.015}_{-0.015}$	$0.449^{+0.016}_{-0.016}$	$0.593^{+0.014}_{-0.014}$	$0.345^{+0.033}_{-0.036}$	$0.744^{+0.028}_{-0.030}$	$0.572^{+0.023}_{-0.024}$
k_T thr. ($q = 0.8, p = +1$)	$0.277^{+0.009}_{-0.011}$	$0.382^{+0.019}_{-0.019}$	$0.277^{+0.013}_{-0.016}$	$0.365^{+0.011}_{-0.014}$	$0.376^{+0.027}_{-0.043}$	$0.741^{+0.042}_{-0.048}$	$0.528^{+0.025}_{-0.025}$
k_T thr. ($q = 0.8, k_{\max} = 10, p = -1$)	$0.505^{+0.011}_{-0.011}$	$0.606^{+0.023}_{-0.022}$	$0.553^{+0.026}_{-0.022}$	$0.727^{+0.013}_{-0.012}$	$0.412^{+0.070}_{-0.076}$	$0.947^{+0.039}_{-0.041}$	$0.783^{+0.040}_{-0.042}$
k_T thr. ($q = 0.9, p = -1$)	$0.485^{+0.015}_{-0.015}$	$0.485^{+0.024}_{-0.023}$	$0.488^{+0.025}_{-0.026}$	$0.734^{+0.019}_{-0.018}$	$0.627^{+0.044}_{-0.049}$	$1.032^{+0.045}_{-0.044}$	$0.618^{+0.029}_{-0.034}$
k_T thr. ($q = 0.9, p = 0$)	$0.465^{+0.013}_{-0.015}$	$0.432^{+0.019}_{-0.018}$	$0.384^{+0.020}_{-0.021}$	$0.610^{+0.012}_{-0.014}$	$0.351^{+0.038}_{-0.039}$	$0.758^{+0.028}_{-0.030}$	$0.564^{+0.024}_{-0.031}$
k_T thr. ($q = 0.9, p = +1$)	$0.405^{+0.015}_{-0.015}$	$0.414^{+0.022}_{-0.020}$	$0.341^{+0.017}_{-0.019}$	$0.579^{+0.016}_{-0.016}$	$0.339^{+0.030}_{-0.041}$	$0.887^{+0.042}_{-0.047}$	$0.570^{+0.028}_{-0.030}$
Laman ($p = -1$)	$0.820^{+0.020}_{-0.021}$	$0.824^{+0.025}_{-0.023}$	$0.812^{+0.034}_{-0.038}$	$1.144^{+0.022}_{-0.025}$	$0.639^{+0.072}_{-0.074}$	$1.310^{+0.025}_{-0.043}$	$0.879^{+0.034}_{-0.035}$
Laman ($p = +1$)	$0.444^{+0.011}_{-0.010}$	$0.445^{+0.013}_{-0.011}$	$0.367^{+0.017}_{-0.017}$	$0.639^{+0.010}_{-0.012}$	$0.290^{+0.034}_{-0.033}$	$0.727^{+0.016}_{-0.019}$	$0.477^{+0.018}_{-0.019}$
Unique-3 ($p = -1$)	$0.792^{+0.021}_{-0.020}$	$0.822^{+0.024}_{-0.025}$	$0.852^{+0.034}_{-0.037}$	$1.098^{+0.023}_{-0.024}$	$0.845^{+0.074}_{-0.080}$	$1.477^{+0.036}_{-0.045}$	$0.987^{+0.032}_{-0.035}$
Unique-3 ($p = +1$)	$0.522^{+0.013}_{-0.012}$	$0.518^{+0.016}_{-0.014}$	$0.538^{+0.024}_{-0.023}$	$0.711^{+0.013}_{-0.013}$	$0.508^{+0.042}_{-0.043}$	$0.821^{+0.021}_{-0.022}$	$0.643^{+0.022}_{-0.021}$
Unique-4 ($p = -1$)	$0.566^{+0.019}_{-0.018}$	$0.722^{+0.026}_{-0.024}$	$0.698^{+0.032}_{-0.033}$	$0.769^{+0.030}_{-0.030}$	$0.877^{+0.067}_{-0.071}$	$1.131^{+0.042}_{-0.043}$	$0.937^{+0.035}_{-0.034}$
Unique-4 ($p = +1$)	$0.397^{+0.013}_{-0.011}$	$0.463^{+0.017}_{-0.015}$	$0.496^{+0.023}_{-0.022}$	$0.497^{+0.018}_{-0.018}$	$0.584^{+0.047}_{-0.048}$	$0.618^{+0.027}_{-0.043}$	$0.651^{+0.021}_{-0.021}$
Unique-5 ($p = -1$)	$0.315^{+0.013}_{-0.013}$	$0.563^{+0.024}_{-0.021}$	$0.485^{+0.018}_{-0.021}$	$0.550^{+0.022}_{-0.022}$	$0.779^{+0.063}_{-0.065}$	$0.933^{+0.034}_{-0.039}$	$0.735^{+0.033}_{-0.031}$
Unique-5 ($p = 0$)	$0.215^{+0.010}_{-0.011}$	$0.350^{+0.016}_{-0.014}$	$0.350^{+0.014}_{-0.016}$	$0.386^{+0.016}_{-0.017}$	$0.544^{+0.042}_{-0.041}$	$0.620^{+0.034}_{-0.040}$	$0.524^{+0.020}_{-0.020}$
Fully connected	$1.364^{+0.033}_{-0.030}$	$1.421^{+0.065}_{-0.064}$	$1.246^{+0.051}_{-0.059}$	$1.741^{+0.035}_{-0.036}$	$1.219^{+0.157}_{-0.169}$	$2.167^{+0.084}_{-0.095}$	$1.694^{+0.088}_{-0.086}$

Table 17: Unconditioned separation S_{sep} under Ollivier-Ricci curvature (per-jet-mean aggregation). Bold marks the most separating construction per subset, excluding FC. Sub- and superscripts give the bootstrap interval where available.

Construction	all nodes	all edges	PV nodes	SV nodes	PV-PV edges	SV-SV edges	PV-SV edges
kNN ($k = 3$)	$0.069^{+0.001}_{-0.001}$	$0.072^{+0.001}_{-0.001}$	$0.058^{+0.002}_{-0.002}$	$0.090^{+0.002}_{-0.002}$	$0.047^{+0.003}_{-0.003}$	$0.081^{+0.018}_{-0.000}$	$0.064^{+0.014}_{-0.000}$
kNN ($k = 4$)	$0.064^{+0.001}_{-0.001}$	$0.065^{+0.001}_{-0.001}$	$0.058^{+0.001}_{-0.002}$	$0.083^{+0.001}_{-0.001}$	$0.050^{+0.002}_{-0.002}$	$0.077^{+0.012}_{-0.000}$	$0.065^{+0.011}_{-0.000}$
kNN ($k = 5$)	$0.056^{+0.001}_{-0.001}$	$0.056^{+0.001}_{-0.001}$	$0.051^{+0.001}_{-0.001}$	$0.072^{+0.001}_{-0.001}$	$0.048^{+0.002}_{-0.002}$	$0.068^{+0.011}_{-0.000}$	$0.058^{+0.009}_{-0.000}$
kNN ($k = 7$)	$0.041^{+0.001}_{-0.001}$	$0.041^{+0.001}_{-0.001}$	$0.037^{+0.001}_{-0.001}$	$0.053^{+0.001}_{-0.001}$	$0.036^{+0.001}_{-0.001}$	$0.052^{+0.012}_{-0.000}$	$0.041^{+0.006}_{-0.000}$
ϵ-based ($\epsilon = 0.2$)	$0.047^{+0.001}_{-0.001}$	$\mathbf{0.082^{+0.002}_{-0.002}}$	$\mathbf{0.076^{+0.002}_{-0.002}}$	$\mathbf{0.098^{+0.002}_{-0.003}}$	$\mathbf{0.065^{+0.004}_{-0.005}}$	$\mathbf{0.095^{+0.000}_{-0.017}}$	$0.070^{+0.000}_{-0.010}$
ϵ-based ($\epsilon = 0.3$)	$0.046^{+0.001}_{-0.001}$	$0.057^{+0.001}_{-0.001}$	$0.035^{+0.002}_{-0.002}$	$0.062^{+0.002}_{-0.002}$	$0.027^{+0.003}_{-0.003}$	$0.075^{+0.000}_{-0.020}$	$0.046^{+0.000}_{-0.014}$
ϵ-based ($\epsilon = 0.4$)	$0.042^{+0.001}_{-0.001}$	$0.043^{+0.001}_{-0.001}$	$0.027^{+0.001}_{-0.001}$	$0.044^{+0.001}_{-0.001}$	$0.024^{+0.002}_{-0.002}$	$0.062^{+0.000}_{-0.018}$	$0.041^{+0.000}_{-0.014}$
ϵ-based ($\epsilon = 0.5$)	$0.036^{+0.001}_{-0.001}$	$0.035^{+0.001}_{-0.001}$	$0.025^{+0.001}_{-0.001}$	$0.038^{+0.001}_{-0.001}$	$0.021^{+0.001}_{-0.001}$	$0.055^{+0.000}_{-0.012}$	$0.035^{+0.000}_{-0.009}$
k_T thr. ($q = 0.4, p = -1$)	$0.062^{+0.001}_{-0.001}$	$0.063^{+0.001}_{-0.001}$	$0.050^{+0.002}_{-0.002}$	$0.069^{+0.002}_{-0.002}$	$0.060^{+0.003}_{-0.003}$	$0.081^{+0.000}_{-0.006}$	$0.053^{+0.001}_{-0.004}$
k_T thr. ($q = 0.4, p = 0$)	$0.057^{+0.001}_{-0.001}$	$0.058^{+0.001}_{-0.001}$	$0.046^{+0.002}_{-0.002}$	$0.071^{+0.001}_{-0.002}$	$0.045^{+0.002}_{-0.003}$	$0.070^{+0.009}_{-0.000}$	$0.049^{+0.004}_{-0.000}$
k_T thr. ($q = 0.4, p = +1$)	$0.061^{+0.001}_{-0.001}$	$0.063^{+0.001}_{-0.001}$	$0.049^{+0.002}_{-0.002}$	$0.076^{+0.002}_{-0.002}$	$0.047^{+0.002}_{-0.003}$	$0.084^{+0.014}_{-0.000}$	$0.056^{+0.000}_{-0.005}$
k_T thr. ($q = 0.5, p = -1$)	$0.059^{+0.001}_{-0.001}$	$0.059^{+0.001}_{-0.001}$	$0.050^{+0.002}_{-0.002}$	$0.067^{+0.001}_{-0.002}$	$0.053^{+0.002}_{-0.003}$	$0.073^{+0.002}_{-0.002}$	$0.053^{+0.000}_{-0.005}$
k_T thr. ($q = 0.5, p = 0$)	$0.053^{+0.001}_{-0.001}$	$0.053^{+0.001}_{-0.001}$	$0.044^{+0.001}_{-0.002}$	$0.067^{+0.001}_{-0.001}$	$0.041^{+0.002}_{-0.003}$	$0.064^{+0.009}_{-0.000}$	$0.047^{+0.003}_{-0.001}$
k_T thr. ($q = 0.5, p = +1$)	$0.059^{+0.001}_{-0.001}$	$0.059^{+0.001}_{-0.001}$	$0.049^{+0.002}_{-0.002}$	$0.073^{+0.002}_{-0.002}$	$0.048^{+0.002}_{-0.002}$	$0.079^{+0.019}_{-0.000}$	$0.053^{+0.001}_{-0.003}$
k_T thr. ($q = 0.6, p = -1$)	$0.054^{+0.001}_{-0.001}$	$0.053^{+0.001}_{-0.001}$	$0.046^{+0.002}_{-0.002}$	$0.063^{+0.001}_{-0.002}$	$0.041^{+0.002}_{-0.002}$	$0.062^{+0.015}_{-0.000}$	$0.048^{+0.000}_{-0.005}$
k_T thr. ($q = 0.7, p = -1$)	$0.048^{+0.001}_{-0.001}$	$0.047^{+0.001}_{-0.001}$	$0.043^{+0.001}_{-0.002}$	$0.061^{+0.001}_{-0.001}$	$0.035^{+0.002}_{-0.002}$	$0.053^{+0.026}_{-0.000}$	$0.044^{+0.000}_{-0.003}$
k_T thr. ($q = 0.8, k_{\max} = 5, p = -1$)	$0.050^{+0.001}_{-0.001}$	$0.049^{+0.001}_{-0.001}$	$0.043^{+0.001}_{-0.001}$	$0.066^{+0.001}_{-0.001}$	$0.038^{+0.002}_{-0.002}$	$0.050^{+0.037}_{-0.000}$	$0.048^{+0.003}_{-0.000}$
k_T thr. ($q = 0.8, p = -1$)	$0.043^{+0.001}_{-0.001}$	$0.042^{+0.001}_{-0.001}$	$0.038^{+0.001}_{-0.001}$	$0.057^{+0.001}_{-0.001}$	$0.036^{+0.002}_{-0.002}$	$0.048^{+0.031}_{-0.000}$	$0.041^{+0.000}_{-0.003}$
k_T thr. ($q = 0.8, p = 0$)	$0.040^{+0.001}_{-0.001}$	$0.039^{+0.001}_{-0.001}$	$0.035^{+0.001}_{-0.001}$	$0.054^{+0.001}_{-0.001}$	$0.033^{+0.001}_{-0.002}$	$0.058^{+0.005}_{-0.000}$	$0.035^{+0.003}_{-0.000}$
k_T thr. ($q = 0.8, p = +1$)	$0.044^{+0.001}_{-0.001}$	$0.044^{+0.001}_{-0.001}$	$0.037^{+0.001}_{-0.001}$	$0.056^{+0.001}_{-0.001}$	$0.036^{+0.002}_{-0.002}$	$0.050^{+0.027}_{-0.000}$	$0.037^{+0.006}_{-0.000}$
k_T thr. ($q = 0.8, k_{\max} = 10, p = -1$)	$0.041^{+0.001}_{-0.001}$	$0.040^{+0.001}_{-0.001}$	$0.038^{+0.001}_{-0.001}$	$0.053^{+0.001}_{-0.001}$	$0.035^{+0.002}_{-0.002}$	$0.047^{+0.027}_{-0.000}$	$0.038^{+0.000}_{-0.004}$
k_T thr. ($q = 0.9, p = -1$)	$0.036^{+0.001}_{-0.001}$	$0.036^{+0.001}_{-0.001}$	$0.030^{+0.001}_{-0.001}$	$0.053^{+0.001}_{-0.001}$	$0.034^{+0.001}_{-0.001}$	$0.051^{+0.025}_{-0.000}$	$0.034^{+0.003}_{-0.000}$
k_T thr. ($q = 0.9, p = 0$)	$0.036^{+0.001}_{-0.001}$	$0.036^{+0.001}_{-0.001}$	$0.030^{+0.001}_{-0.001}$	$0.050^{+0.001}_{-0.001}$	$0.031^{+0.001}_{-0.002}$	$0.055^{+0.008}_{-0.000}$	$0.032^{+0.005}_{-0.000}$
k_T thr. ($q = 0.9, p = +1$)	$0.037^{+0.001}_{-0.001}$	$0.036^{+0.001}_{-0.001}$	$0.030^{+0.001}_{-0.001}$	$0.045^{+0.001}_{-0.001}$	$0.028^{+0.001}_{-0.001}$	$0.040^{+0.017}_{-0.000}$	$0.030^{+0.004}_{-0.000}$
Laman ($p = -1$)	$0.072^{+0.001}_{-0.002}$	$0.075^{+0.001}_{-0.002}$	$0.058^{+0.002}_{-0.002}$	$0.088^{+0.001}_{-0.002}$	$0.048^{+0.003}_{-0.003}$	$0.090^{+0.009}_{-0.000}$	$0.064^{+0.009}_{-0.000}$
Laman ($p = +1$)	$0.068^{+0.001}_{-0.001}$	$0.073^{+0.001}_{-0.002}$	$0.055^{+0.002}_{-0.002}$	$0.082^{+0.002}_{-0.002}$	$0.052^{+0.003}_{-0.003}$	$0.090^{+0.005}_{-0.000}$	$0.059^{+0.009}_{-0.000}$
Unique-3 ($p = -1$)	$\mathbf{0.074^{+0.001}_{-0.001}}$	$0.074^{+0.001}_{-0.001}$	$0.067^{+0.002}_{-0.002}$	$0.095^{+0.002}_{-0.002}$	$0.059^{+0.003}_{-0.003}$	$0.079^{+0.032}_{-0.000}$	$\mathbf{0.074^{+0.009}_{-0.000}}$
Unique-3 ($p = +1$)	$0.069^{+0.001}_{-0.001}$	$0.070^{+0.001}_{-0.001}$	$0.061^{+0.002}_{-0.002}$	$0.086^{+0.001}_{-0.002}$	$0.055^{+0.003}_{-0.003}$	$0.078^{+0.020}_{-0.000}$	$0.067^{+0.010}_{-0.000}$
Unique-4 ($p = -1$)	$0.065^{+0.001}_{-0.001}$	$0.064^{+0.001}_{-0.001}$	$0.062^{+0.001}_{-0.002}$	$0.087^{+0.001}_{-0.002}$	$0.059^{+0.002}_{-0.002}$	$0.070^{+0.033}_{-0.000}$	$0.070^{+0.007}_{-0.000}$
Unique-4 ($p = +1$)	$0.061^{+0.001}_{-0.001}$	$0.061^{+0.001}_{-0.001}$	$0.057^{+0.001}_{-0.002}$	$0.079^{+0.001}_{-0.001}$	$0.055^{+0.002}_{-0.002}$	$0.073^{+0.018}_{-0.000}$	$0.064^{+0.008}_{-0.000}$
Unique-5 ($p = -1$)	$0.055^{+0.001}_{-0.001}$	$0.054^{+0.001}_{-0.001}$	$0.052^{+0.001}_{-0.001}$	$0.074^{+0.001}_{-0.001}$	$0.052^{+0.002}_{-0.002}$	$0.063^{+0.028}_{-0.000}$	$0.059^{+0.006}_{-0.000}$
Unique-5 ($p = 0$)	$0.052^{+0.001}_{-0.001}$	$0.052^{+0.001}_{-0.001}$	$0.049^{+0.001}_{-0.001}$	$0.069^{+0.001}_{-0.001}$	$0.049^{+0.002}_{-0.002}$	$0.067^{+0.016}_{-0.000}$	$0.055^{+0.006}_{-0.000}$
Fully connected	$0.027^{+0.001}_{-0.001}$	$0.027^{+0.001}_{-0.001}$	$0.022^{+0.001}_{-0.001}$	$0.032^{+0.001}_{-0.001}$	$0.020^{+0.001}_{-0.001}$	$0.033^{+0.006}_{-0.000}$	$0.023^{+0.003}_{-0.000}$

Table 18: Unconditioned separation S_{sep} under Forman-Ricci curvature (per-jet-mean aggregation). Bold marks the most separating construction per subset, excluding FC. Sub- and superscripts give the bootstrap interval where available.

Construction	all nodes	all edges	PV nodes	SV nodes	PV-PV edges	SV-SV edges	PV-SV edges
kNN ($k = 3$)	$0.400^{+0.008}_{-0.008}$	$0.414^{+0.008}_{-0.008}$	$0.388^{+0.011}_{-0.013}$	$0.594^{+0.013}_{-0.012}$	$0.356^{+0.020}_{-0.023}$	$0.570^{+0.166}_{-0.000}$	$0.467^{+0.093}_{-0.000}$
kNN ($k = 4$)	$0.281^{+0.009}_{-0.009}$	$0.280^{+0.009}_{-0.009}$	$0.353^{+0.012}_{-0.012}$	$0.470^{+0.014}_{-0.018}$	$0.389^{+0.018}_{-0.020}$	$0.507^{+0.114}_{-0.000}$	$0.448^{+0.089}_{-0.000}$
kNN ($k = 5$)	$0.204^{+0.010}_{-0.010}$	$0.209^{+0.010}_{-0.010}$	$0.244^{+0.010}_{-0.011}$	$0.336^{+0.013}_{-0.013}$	$0.280^{+0.009}_{-0.015}$	$0.422^{+0.198}_{-0.000}$	$0.277^{+0.079}_{-0.000}$
kNN ($k = 7$)	$0.573^{+0.014}_{-0.015}$	$0.586^{+0.014}_{-0.015}$	$0.419^{+0.017}_{-0.019}$	$0.633^{+0.015}_{-0.014}$	$0.333^{+0.024}_{-0.026}$	$0.660^{+0.250}_{-0.000}$	$0.417^{+0.090}_{-0.000}$
ϵ-based ($\epsilon = 0.2$)	$0.588^{+0.010}_{-0.012}$	$0.670^{+0.011}_{-0.013}$	$0.641^{+0.016}_{-0.019}$	$0.717^{+0.016}_{-0.018}$	$0.701^{+0.033}_{-0.035}$	$0.985^{+0.133}_{-0.000}$	$0.708^{+0.000}_{-0.120}$
ϵ-based ($\epsilon = 0.3$)	$1.022^{+0.015}_{-0.016}$	$1.027^{+0.015}_{-0.016}$	$1.059^{+0.025}_{-0.025}$	$1.098^{+0.020}_{-0.019}$	$1.118^{+0.041}_{-0.042}$	$1.247^{+0.346}_{-0.000}$	$1.121^{+0.000}_{-0.104}$
ϵ-based ($\epsilon = 0.4$)	$1.278^{+0.018}_{-0.017}$	$1.271^{+0.019}_{-0.016}$	$1.272^{+0.027}_{-0.027}$	$1.382^{+0.026}_{-0.026}$	$1.311^{+0.044}_{-0.043}$	$1.328^{+0.565}_{-0.000}$	$\mathbf{1.308^{+0.089}_{-0.000}}$
ϵ-based ($\epsilon = 0.5$)	$\mathbf{1.306^{+0.019}_{-0.019}}$	$\mathbf{1.298^{+0.019}_{-0.019}}$	$\mathbf{1.279^{+0.026}_{-0.028}}$	$\mathbf{1.514^{+0.028}_{-0.029}}$	$\mathbf{1.320^{+0.048}_{-0.039}}$	$1.327^{+0.773}_{-0.000}$	$1.257^{+0.225}_{-0.000}$
k_T thr. ($q = 0.4, p = -1$)	$0.416^{+0.009}_{-0.010}$	$0.372^{+0.008}_{-0.010}$	$0.447^{+0.012}_{-0.014}$	$0.659^{+0.017}_{-0.018}$	$0.443^{+0.024}_{-0.022}$	$0.515^{+0.307}_{-0.000}$	$0.512^{+0.016}_{-0.022}$
k_T thr. ($q = 0.4, p = 0$)	$0.219^{+0.007}_{-0.007}$	$0.197^{+0.007}_{-0.007}$	$0.290^{+0.007}_{-0.009}$	$0.440^{+0.010}_{-0.013}$	$0.337^{+0.011}_{-0.012}$	$0.498^{+0.331}_{-0.000}$	$0.263^{+0.081}_{-0.000}$
k_T thr. ($q = 0.4, p = +1$)	$0.413^{+0.008}_{-0.009}$	$0.393^{+0.009}_{-0.010}$	$0.398^{+0.013}_{-0.015}$	$0.565^{+0.016}_{-0.016}$	$0.416^{+0.020}_{-0.023}$	$0.614^{+0.312}_{-0.000}$	$0.423^{+0.012}_{-0.024}$
k_T thr. ($q = 0.5, p = -1$)	$0.349^{+0.008}_{-0.009}$	$0.270^{+0.008}_{-0.008}$	$0.465^{+0.011}_{-0.015}$	$0.641^{+0.017}_{-0.018}$	$0.486^{+0.017}_{-0.018}$	$0.543^{+0.306}_{-0.000}$	$0.487^{+0.016}_{-0.018}$
k_T thr. ($q = 0.5, p = 0$)	$0.164^{+0.007}_{-0.007}$	$0.183^{+0.006}_{-0.007}$	$0.302^{+0.007}_{-0.007}$	$0.441^{+0.012}_{-0.012}$	$0.394^{+0.010}_{-0.015}$	$0.561^{+0.371}_{-0.000}$	$0.223^{+0.183}_{-0.000}$
k_T thr. ($q = 0.5, p = +1$)	$0.355^{+0.006}_{-0.010}$	$0.305^{+0.008}_{-0.010}$	$0.382^{+0.013}_{-0.014}$	$0.510^{+0.015}_{-0.013}$	$0.406^{+0.012}_{-0.021}$	$0.630^{+0.333}_{-0.000}$	$0.342^{+0.039}_{-0.000}$
k_T thr. ($q = 0.6, p = -1$)	$0.189^{+0.007}_{-0.008}$	$0.145^{+0.007}_{-0.008}$	$0.414^{+0.007}_{-0.008}$	$0.578^{+0.014}_{-0.015}$	$0.449^{+0.008}_{-0.013}$	$0.542^{+0.309}_{-0.000}$	$0.361^{+0.140}_{-0.000}$
k_T thr. ($q = 0.7, p = -1$)	$0.081^{+0.003}_{-0.005}$	$0.134^{+0.004}_{-0.005}$	$0.442^{+0.013}_{-0.013}$	$0.531^{+0.013}_{-0.013}$	$0.444^{+0.020}_{-0.027}$	$0.511^{+0.305}_{-0.000}$	$0.308^{+0.281}_{-0.000}$
k_T thr. ($q = 0.8, k_{\max} = 5, p = -1$)	$0.173^{+0.006}_{-0.006}$	$0.204^{+0.007}_{-0.006}$	$0.421^{+0.010}_{-0.012}$	$0.604^{+0.017}_{-0.016}$	$0.525^{+0.019}_{-0.023}$	$0.467^{+0.491}_{-0.000}$	$0.369^{+0.208}_{-0.000}$
k_T thr. ($q = 0.8, p = -1$)	$0.265^{+0.004}_{-0.006}$	$0.327^{+0.005}_{-0.007}$	$0.479^{+0.013}_{-0.013}$	$0.572^{+0.011}_{-0.012}$	$0.488^{+0.022}_{-0.025}$	$0.528^{+0.333}_{-0.000}$	$0.386^{+0.274}_{-0.000}$
k_T thr. ($q = 0.8, p = 0$)	$0.348^{+0.006}_{-0.007}$	$0.417^{+0.007}_{-0.008}$	$0.498^{+0.014}_{-0.014}$	$0.553^{+0.012}_{-0.011}$	$0.490^{+0.023}_{-0.022}$	$0.668^{+0.277}_{-0.000}$	$0.410^{+0.214}_{-0.000}$
k_T thr. ($q = 0.8, p = +1$)	$0.197^{+0.006}_{-0.008}$	$0.246^{+0.007}_{-0.009}$	$0.274^{+0.012}_{-0.012}$	$0.574^{+0.016}_{-0.013}$	$0.448^{+0.017}_{-0.024}$	$1.150^{+0.000}_{-0.226}$	$0.285^{+0.131}_{-0.000}$
k_T thr. ($q = 0.8, k_{\max} = 10, p = -1$)	$0.368^{+0.007}_{-0.007}$	$0.437^{+0.007}_{-0.007}$	$0.529^{+0.014}_{-0.013}$	$0.676^{+0.013}_{-0.012}$	$0.547^{+0.030}_{-0.027}$	$0.627^{+0.174}_{-0.000}$	$0.472^{+0.359}_{-0.000}$
k_T thr. ($q = 0.9, p = -1$)	$0.508^{+0.010}_{-0.012}$	$0.541^{+0.010}_{-0.012}$	$0.522^{+0.015}_{-0.016}$	$0.655^{+0.012}_{-0.014}$	$0.523^{+0.028}_{-0.030}$	$0.687^{+0.256}_{-0.000}$	$0.553^{+0.147}_{-0.000}$
k_T thr. ($q = 0.9, p = 0$)	$0.495^{+0.011}_{-0.011}$	$0.523^{+0.011}_{-0.012}$	$0.457^{+0.014}_{-0.016}$	$0.595^{+0.012}_{-0.011}$	$0.382^{+0.024}_{-0.025}$	$0.673^{+0.217}_{-0.000}$	$0.446^{+0.193}_{-0.000}$
k_T thr. ($q = 0.9, p = +1$)	$0.437^{+0.010}_{-0.010}$	$0.462^{+0.010}_{-0.011}$	$0.390^{+0.011}_{-0.014}$	$0.720^{+0.016}_{-0.016}$	$0.445^{+0.018}_{-0.023}$	$\mathbf{1.341^{+0.000}_{-0.272}}$	$0.466^{+0.097}_{-0.000}$
Laman ($p = -1$)	$0.797^{+0.013}_{-0.016}$	$0.789^{+0.012}_{-0.015}$	$0.781^{+0.019}_{-0.021}$	$1.190^{+0.026}_{-0.024}$	$0.545^{+0.031}_{-0.036}$	$0.943^{+0.447}_{-0.000}$	$0.892^{+0.152}_{-0.000}$
Laman ($p = +1$)	$0.467^{+0.008}_{-0.008}$	$0.491^{+0.008}_{-0.008}$	$0.398^{+0.012}_{-0.013}$	$0.659^{+0.012}_{-0.013}$	$0.374^{+0.020}_{-0.019}$	$0.568^{+0.273}_{-0.000}$	$0.464^{+0.102}_{-0.000}$
Unique-3 ($p = -1$)	$0.659^{+0.010}_{-0.012}$	$0.607^{+0.010}_{-0.011}$	$0.764^{+0.017}_{-0.017}$	$1.157^{+0.027}_{-0.024}$	$0.742^{+0.023}_{-0.026}$	$0.728^{+0.939}_{-0.000}$	$0.919^{+0.119}_{-0.000}$
Unique-3 ($p = +1$)	$0.454^{+0.008}_{-0.008}$	$0.442^{+0.008}_{-0.008}$	$0.510^{+0.012}_{-0.014}$	$0.736^{+0.015}_{-0.016}$	$0.559^{+0.020}_{-0.022}$	$0.725^{+0.304}_{-0.000}$	$0.610^{+0.083}_{-0.000}$
Unique-4 ($p = -1$)	$0.357^{+0.009}_{-0.010}$	$0.295^{+0.008}_{-0.008}$	$0.530^{+0.012}_{-0.013}$	$0.819^{+0.025}_{-0.023}$	$0.682^{+0.017}_{-0.017}$	$0.580^{+0.746}_{-0.000}$	$0.681^{+0.094}_{-0.000}$
Unique-4 ($p = +1$)	$0.243^{+0.007}_{-0.006}$	$0.215^{+0.006}_{-0.006}$	$0.381^{+0.009}_{-0.011}$	$0.533^{+0.018}_{-0.017}$	$0.515^{+0.012}_{-0.014}$	$0.700^{+0.178}_{-0.000}$	$0.479^{+0.051}_{-0.000}$
Unique-5 ($p = -1$)	$0.266^{+0.009}_{-0.011}$	$0.255^{+0.009}_{-0.011}$	$0.357^{+0.013}_{-0.017}$	$0.581^{+0.018}_{-0.017}$	$0.532^{+0.019}_{-0.022}$	$0.475^{+0.569}_{-0.000}$	$0.350^{+0.209}_{-0.000}$
Unique-5 ($p = 0$)	$0.227^{+0.009}_{-0.011}$	$0.230^{+0.009}_{-0.010}$	$0.286^{+0.012}_{-0.015}$	$0.430^{+0.013}_{-0.014}$	$0.386^{+0.016}_{-0.020}$	$0.553^{+0.287}_{-0.000}$	$0.241^{+0.135}_{-0.000}$
Fully connected	$1.232^{+0.018}_{-0.019}$	$1.232^{+0.018}_{-0.019}$	$1.161^{+0.027}_{-0.027}$	$1.658^{+0.036}_{-0.035}$	$1.194^{+0.037}_{-0.038}$	$1.608^{+0.797}_{-0.000}$	$1.267^{+0.265}_{-0.000}$

F.10 SEPARATION PRESERVED BY CONDITIONING ON n

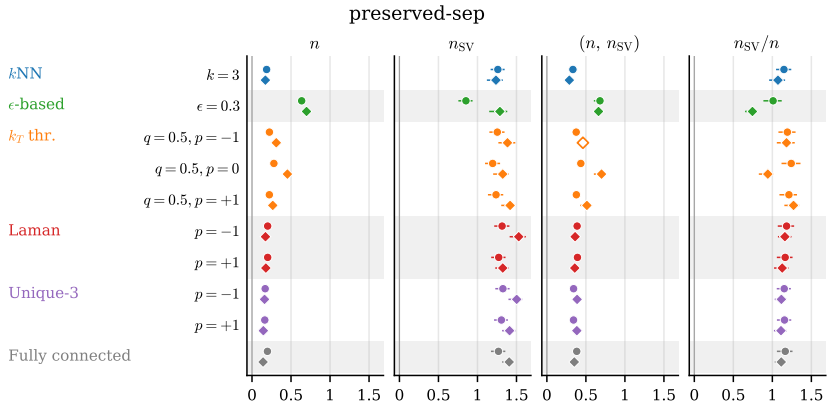


Figure 25: Preserved separation over all nodes under ORC for the default constructions, union aggregation, conditioned on n, n_{sv}, (n, n_{sv}), and n_{sv}/n (Equation 6). Error bars are 95% bootstrap intervals. Conditioning on n_{sv} or n_{sv}/n alone pushes the ratio above one for almost every construction, because flavors matched on n_{sv} differ more in n; Table 19 gives the values at fixed n on every structural subset.

Table 19: Preserved separation under conditioning on n, preserved-sep$_n$ of Equation 6, under ORC for the default constructions, with (a) union and (b) per-jet-mean aggregation. Each entry is the binned separation of Equation 5 over the n bins of Table 4 divided by the pooled separation of the same jets. The subscripts and superscripts give the 95% bootstrap interval, and † marks intervals whose replicates miss the point estimate by more than one standard error (§ C.3). The FC is the reference that finite bin width and the W_1 bias of Equation 8 leave when no separation survives, and a value above one means that pooling over n was hiding separation. Coverage (Equation 16) is 0.99 on all nodes and PV nodes and 0.43 on SV nodes.

(a) Union aggregation

Construction	all nodes	all edges	PV nodes	SV nodes	PV-PV edges	SV-SV edges	PV-SV edges
kNN ($k = 3$)	$0.187^{+0.007}_{-0.012}$	$0.203^{+0.010}_{-0.013}$	$0.283^{+0.015}_{-0.022}$	$0.168^{+0.006}_{-0.016}$	$0.661^{+0.056}_{-0.063}$	$0.308^{+0.017}_{-0.037}$	$0.181^{+0.015}_{-0.028}$
ϵ-based ($\epsilon = 0.3$)	$0.637^{+0.020}_{-0.028}$	$0.725^{+0.044}_{-0.043}$	$0.869^{+0.045}_{-0.058}$	$0.510^{+0.019}_{-0.027}$	$1.179^{+0.135}_{-0.175}$	$0.428^{+0.026}_{-0.045}$	$1.014^{+0.071}_{-0.050}$
k_T thr. ($q = 0.5$, $p = -1$)	$0.223^{+0.008}_{-0.012}$	$0.285^{+0.013}_{-0.016}$	$0.298^{+0.018}_{-0.020}$	$0.245^{+0.012}_{-0.017}$	$1.059^{+0.088}_{-0.126}$	$0.370^{+0.021}_{-0.033}$	$0.674^{+0.045}_{-0.057}$
k_T thr. ($q = 0.5$, $p = 0$)	$0.280^{+0.009}_{-0.016}$	$0.364^{+0.012}_{-0.022}$	$0.340^{+0.019}_{-0.028}$	$0.253^{+0.006}_{-0.021}$	$0.708^{+0.070}_{-0.095}$	$0.409^{+0.019}_{-0.037}$	$0.516^{+0.057}_{-0.070}$
k_T thr. ($q = 0.5$, $p = +1$)	$0.222^{+0.009}_{-0.012}$	$0.252^{+0.009}_{-0.017}$	$0.260^{+0.016}_{-0.023}$	$0.216^{+0.008}_{-0.015}$	$0.424^{+0.026}_{-0.061}$	$0.367^{+0.024}_{-0.033}$	$0.266^{+0.022}_{-0.038}$
Laman ($p = -1$)	$0.201^{+0.007}_{-0.010}$	$0.218^{+0.008}_{-0.011}$	$0.254^{+0.015}_{-0.016}$	$0.215^{+0.007}_{-0.010}$	$0.705^{+0.058}_{-0.089}$	$0.364^{+0.012}_{-0.021}$	$0.321^{+0.018}_{-0.022}$
Laman ($p \in \{0, +1\}$)	$0.201^{+0.006}_{-0.010}$	$0.204^{+0.007}_{-0.011}$	$0.397^{+0.017}_{-0.020}$	$0.269^{+0.010}_{-0.012}$	$0.976^{+0.086}_{-0.113}$	$0.358^{+0.018}_{-0.033}$	$0.209^{+0.019}_{-0.029}$
Unique-3 ($p = -1$)	$0.169^{+0.007}_{-0.009}$	$0.184^{+0.008}_{-0.010}$	$0.229^{+0.011}_{-0.014}$	$0.202^{+0.006}_{-0.009}$	$0.388^{+0.033}_{-0.043}$	$0.403^{+0.014}_{-0.020}$	$0.231^{+0.010}_{-0.018}$
Unique-3 ($p \in \{0, +1\}$)	$0.164^{+0.007}_{-0.009}$	$0.176^{+0.008}_{-0.010}$	$0.220^{+0.013}_{-0.014}$	$0.183^{+0.007}_{-0.009}$	$0.368^{+0.033}_{-0.040}$	$0.288^{+0.014}_{-0.017}$	$0.169^{+0.010}_{-0.019}$
Fully connected	$0.198^{+0.007}_{-0.009}$	$0.236^{+0.010}_{-0.012}$	$0.215^{+0.015}_{-0.014}$	$0.180^{+0.006}_{-0.008}$	$0.329^{+0.045}_{-0.046}$	$0.190^{+0.007}_{-0.012}$	$0.197^{+0.010}_{-0.016}$

(b) Per-jet-mean aggregation

Construction	all nodes	all edges	PV nodes	SV nodes	PV-PV edges	SV-SV edges	PV-SV edges
kNN ($k = 3$)	$0.165^{+0.007}_{-0.011}$	$0.165^{+0.007}_{-0.011}$	$0.301^{+0.013}_{-0.017}$	$0.251^{+0.007}_{-0.013}$	$0.895^{+0.038}_{-0.054}$	$0.423^{+0.020}_{-0.033}$	$0.215^{+0.012}_{-0.029}$
ϵ-based ($\epsilon = 0.3$)	$0.582^{+0.025}_{-0.030}$	$0.493^{+0.015}_{-0.022}$	$0.948^{+0.049}_{-0.051}$	$0.535^{+0.032}_{-0.032}$	$0.943^{+0.077}_{-0.125}$	$0.359^{+0.009\,\dagger}_{-0.045}$	$0.860^{+0.068}_{-0.067}$
k_T thr. ($q = 0.5$, $p = -1$)	$0.194^{+0.007}_{-0.009}$	$0.200^{+0.007}_{-0.008}$	$0.347^{+0.011}_{-0.015}$	$0.384^{+0.015}_{-0.017}$	$1.072^{+0.035}_{-0.043}$	$0.572^{+0.025}_{-0.035}$	$0.500^{+0.030}_{-0.039}$
k_T thr. ($q = 0.5$, $p = 0$)	$0.237^{+0.006}_{-0.011}$	$0.249^{+0.007}_{-0.010}$	$0.407^{+0.014}_{-0.017}$	$0.356^{+0.010}_{-0.020}$	$1.000^{+0.038}_{-0.047}$	$0.597^{+0.023}_{-0.036}$	$0.564^{+0.037}_{-0.050}$
k_T thr. ($q = 0.5$, $p = +1$)	$0.192^{+0.007}_{-0.009}$	$0.185^{+0.007}_{-0.009}$	$0.307^{+0.012}_{-0.015}$	$0.292^{+0.008}_{-0.016}$	$0.568^{+0.023}_{-0.030}$	$0.472^{+0.025}_{-0.029}$	$0.362^{+0.030}_{-0.042}$
Laman ($p = -1$)	$0.179^{+0.005}_{-0.007}$	$0.175^{+0.006}_{-0.008}$	$0.247^{+0.009}_{-0.011}$	$0.251^{+0.009}_{-0.010}$	$0.997^{+0.050}_{-0.062}$	$0.496^{+0.019}_{-0.030}$	$0.218^{+0.011}_{-0.017}$
Laman ($p \in \{0, +1\}$)	$0.183^{+0.005}_{-0.009}$	$0.178^{+0.006}_{-0.008}$	$0.422^{+0.015}_{-0.015}$	$0.328^{+0.011}_{-0.014}$	$1.065^{+0.042}_{-0.048}$	$0.497^{+0.021}_{-0.033}$	$0.272^{+0.019}_{-0.026}$
Unique-3 ($p = -1$)	$0.157^{+0.006}_{-0.008}$	$0.155^{+0.006}_{-0.008}$	$0.220^{+0.007}_{-0.009}$	$0.229^{+0.007}_{-0.009}$	$0.453^{+0.020}_{-0.024}$	$0.498^{+0.017}_{-0.021}$	$0.194^{+0.008}_{-0.014}$
Unique-3 ($p \in \{0, +1\}$)	$0.153^{+0.005}_{-0.008}$	$0.153^{+0.006}_{-0.008}$	$0.218^{+0.008}_{-0.010}$	$0.213^{+0.008}_{-0.008}$	$0.439^{+0.020}_{-0.023}$	$0.363^{+0.015}_{-0.022}$	$0.185^{+0.011}_{-0.014}$
Fully connected	$0.181^{+0.006}_{-0.007}$	$0.181^{+0.006}_{-0.007}$	$0.183^{+0.009}_{-0.011}$	$0.178^{+0.007}_{-0.008}$	$0.222^{+0.013}_{-0.015}$	$0.176^{+0.016}_{-0.019}$	$0.172^{+0.010}_{-0.011}$

G Predictive Value: Full Results

Table 20 summarizes the jet results over the three flavor pairs. The QM9 tables follow in § G.1, the per-pair jet tables in § G.2–§ G.4, the lower learning rate in § G.5, and the two controls in § G.6 and § G.7.

Table 20: Change in test AUROC (Equation 7, $\times 10^{-3}$) from adding curvature features. Brackets give the minimum and maximum over graph constructions (36 for b and c, 16 for τ). Per-construction results with intervals are in § G.2–§ G.4. Summary-statistic models see curvature only through jet-level statistics, while per-track models receive it per track. †The maximum is from the $\epsilon = 0.3$ ϵ-based graph, all other constructions give at most +3.7.

	b vs light	c vs light	τ vs light
Summary-statistic models			
Logistic regression	[+4.6, +10.6]	[+2.8, +6.7]	[+0.7, +10.8]†
Small MLP	[−1.6, +0.6]	[−1.6, +1.3]	[−5.2, +0.5]
Large MLP	[−2.0, +0.9]	[−3.2, +1.7]	[−6.3, −1.3]
Boosted trees	[−0.8, +4.8]	[−3.0, +5.2]	[−1.7, +2.9]
Per-track feature models			
Transformer	[+9.0, +13.7]	[−0.4, +3.1]	[+14.3, +25.9]
GNN, rest	[+6.5, +17.0]	[−1.5, +2.3]	[+2.4, +16.5]
GNN, fully connected	+5.3	+0.2	+0.2

G.1 QM9

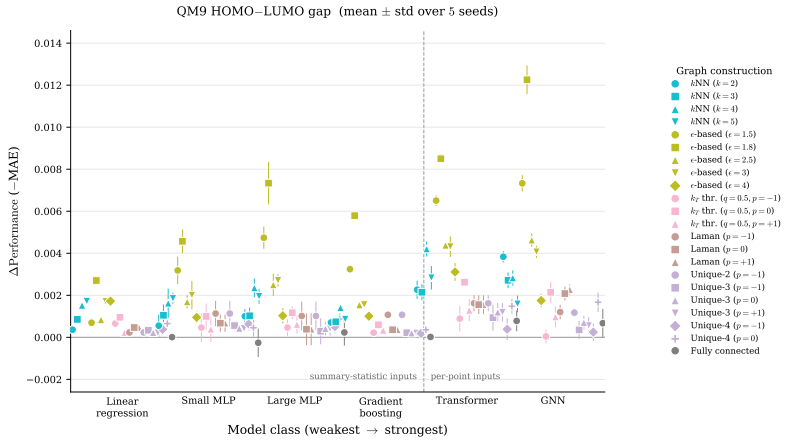


Figure 26: Incremental predictive value of curvature (Equation 7) for HOMO–LUMO gap for 22 graph constructions. Each marker is ΔPerformance, the AUROC with curvature appended to the inputs minus the AUROC without it, for one graph construction (color and marker shape) and one model class, with model classes (with learning rate (10^{-3})) ordered from weakest to strongest along the horizontal axis. Markers show the mean of the per-seed deltas over 5 seeds and error bars their standard deviation across seeds. Error bars combine seed to seed and test set bootstrap standard deviations in quadrature ($\sqrt{(\sigma_{seed})^2 + \sigma_{boot}^2}$). For logistic regression and gradient boosting the seed component is zero, so their bars show bootstrap uncertainty alone.

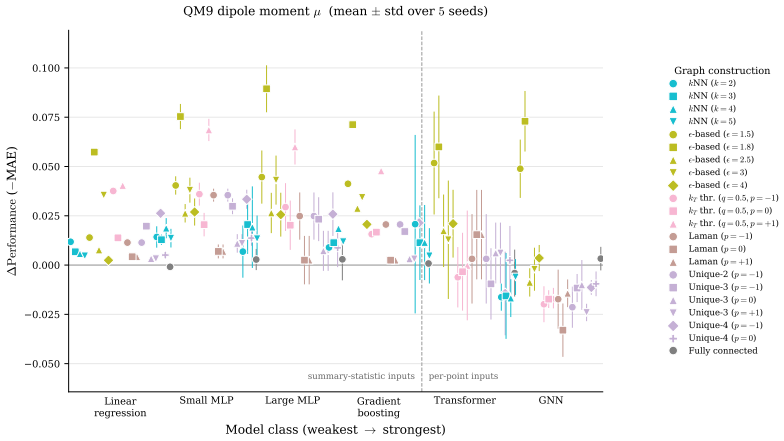


Figure 27: Incremental predictive value of curvature (Equation 7) for dipole moment for 22 graph constructions. Each marker is ΔPerformance, the AUROC with curvature appended to the inputs minus the AUROC without it, for one graph construction (color and marker shape) and one model class, with model classes (with learning rate (10^{-3})) ordered from weakest to strongest along the horizontal axis. Markers show the mean of the per-seed deltas over 5 seeds and error bars their standard deviation across seeds. Error bars combine seed to seed and test set bootstrap standard deviations in quadrature ($\sqrt{(\sigma_{seed})^2 + \sigma_{boot}^2}$). For logistic regression and gradient boosting the seed component is zero, so their bars show bootstrap uncertainty alone.

Table 21: Reduction in QM9 test MAE from adding curvature (Equation 7; positive is better), next to bond recovery F1 and ORC hydrogen-heavy separation from Table 9. Mean over 5 seeds. Bold marks a 95% bootstrap interval excluding zero. Baselines without curvature: transformer 0.0180 (gap), 0.508 (μ) and complete graph GNN 0.0159 Ha, 0.497 D. All constructions, model classes, and intervals are in Table 22–Table 25.

			Gap Δ ($\times 10^{-3}$)		μ Δ ($\times 10^{-2}$)	
Construction	F1	ORC S_{sep}	Transf.	GNN	Transf.	GNN
FC	0.222	0.000	**+0.8**	**+0.7**	−0.41	+0.33
kNN, $k = 2$	0.738	0.208	**+2.3**	**+3.8**	+2.08	**−1.63**
ϵ-based, 1.5 Å	0.874	0.160	**+6.5**	**+7.3**	**+5.18**	**+4.89**
ϵ-based, 1.8 Å	0.908	0.437	**+8.5**	**+12.3**	**+5.99**	**+7.30**
ϵ-based, 2.5 Å	0.546	0.085	**+4.4**	**+4.6**	**+1.74**	**−0.88**
ϵ-based, 3.0 Å	0.412	0.033	**+4.3**	**+4.1**	+1.28	−0.21
ϵ-based, 4.0 Å	0.294	0.008	**+3.1**	**+1.7**	**+2.10**	+0.36
Laman, $p = 1$	0.691	0.221	**+1.6**	**+2.3**	+1.54	**−1.43**
Laman, $p = -1$	0.690	0.125	**+1.6**	**+1.2**	+0.32	**−1.73**
unique-k, $k = 4$, $p = 0$	0.463	0.049	**+1.5**	**+1.7**	+0.24	**−0.95**
unique-k, $k = 4$, $p = -1$	0.463	0.003	+0.4	+0.2	−1.44	**−1.13**

Table 22: Test MAE (lower is better) for QM9 HOMO–LUMO gap for each graph construction and model class, (a) without and (b) with FRC and ORC features. Entries give the mean over 5 training seeds; brackets give the 95% percentile bootstrap confidence interval of that mean, obtained by resampling the seeds with replacement 10,000 times, averaging each resample, and taking the 2.5th and 97.5th percentiles of the resampled means.

(a) Without curvature features

Graph construction	LinReg	Small MLP	Large MLP	Grad. boost.	Transformer	GNN
Fully connected	0.0288 [0.0288, 0.0288]	0.0238 [0.0237, 0.0241]	0.0212 [0.0209, 0.0215]	0.0218 [0.0218, 0.0218]	0.0180 [0.0177, 0.0183]	0.0159 [0.0157, 0.0162]
kNN ($k = 2$)	0.0288 [0.0288, 0.0288]	0.0238 [0.0237, 0.0241]	0.0212 [0.0209, 0.0215]	0.0218 [0.0218, 0.0218]	0.0180 [0.0177, 0.0183]	0.0224 [0.0223, 0.0225]
kNN ($k = 3$)	0.0288 [0.0288, 0.0288]	0.0238 [0.0237, 0.0241]	0.0212 [0.0209, 0.0215]	0.0218 [0.0218, 0.0218]	0.0180 [0.0177, 0.0183]	0.0217 [0.0216, 0.0219]
kNN ($k = 4$)	0.0288 [0.0288, 0.0288]	0.0238 [0.0237, 0.0241]	0.0212 [0.0209, 0.0215]	0.0218 [0.0218, 0.0218]	0.0180 [0.0177, 0.0183]	0.0212 [0.0210, 0.0214]
kNN ($k = 5$)	0.0288 [0.0288, 0.0288]	0.0238 [0.0237, 0.0241]	0.0212 [0.0209, 0.0215]	0.0218 [0.0218, 0.0218]	0.0180 [0.0177, 0.0183]	0.0210 [0.0208, 0.0213]
k_T threshold ($p = -1$)	0.0288 [0.0288, 0.0288]	0.0238 [0.0237, 0.0241]	0.0212 [0.0209, 0.0215]	0.0218 [0.0218, 0.0218]	0.0180 [0.0177, 0.0183]	0.0203 [0.0200, 0.0206]
k_T threshold ($p = +1$)	0.0288 [0.0288, 0.0288]	0.0238 [0.0237, 0.0241]	0.0212 [0.0209, 0.0215]	0.0218 [0.0218, 0.0218]	0.0180 [0.0177, 0.0183]	0.0202 [0.0199, 0.0204]
k_T threshold ($p = 0$)	0.0288 [0.0288, 0.0288]	0.0238 [0.0237, 0.0241]	0.0212 [0.0209, 0.0215]	0.0218 [0.0218, 0.0218]	0.0180 [0.0177, 0.0183]	0.0204 [0.0201, 0.0207]
Laman ($p = -1$)	0.0288 [0.0288, 0.0288]	0.0238 [0.0237, 0.0241]	0.0212 [0.0209, 0.0215]	0.0218 [0.0218, 0.0218]	0.0180 [0.0177, 0.0183]	0.0217 [0.0216, 0.0218]
Laman ($p = +1$)	0.0288 [0.0288, 0.0288]	0.0238 [0.0237, 0.0241]	0.0212 [0.0209, 0.0215]	0.0218 [0.0218, 0.0218]	0.0180 [0.0177, 0.0183]	0.0218 [0.0217, 0.0220]
Laman ($p = 0$)	0.0288 [0.0288, 0.0288]	0.0238 [0.0237, 0.0241]	0.0212 [0.0209, 0.0215]	0.0218 [0.0218, 0.0218]	0.0180 [0.0177, 0.0183]	0.0218 [0.0215, 0.0219]
ϵ-based ($\epsilon = 1.5$)	0.0288 [0.0288, 0.0288]	0.0238 [0.0237, 0.0241]	0.0212 [0.0209, 0.0215]	0.0218 [0.0218, 0.0218]	0.0180 [0.0177, 0.0183]	0.0227 [0.0226, 0.0229]
ϵ-based ($\epsilon = 2.5$)	0.0288 [0.0288, 0.0288]	0.0238 [0.0237, 0.0241]	0.0212 [0.0209, 0.0215]	0.0218 [0.0218, 0.0218]	0.0180 [0.0177, 0.0183]	0.0212 [0.0211, 0.0214]
ϵ-based ($\epsilon = 3$)	0.0288 [0.0288, 0.0288]	0.0238 [0.0237, 0.0241]	0.0212 [0.0209, 0.0215]	0.0218 [0.0218, 0.0218]	0.0180 [0.0177, 0.0183]	0.0202 [0.0200, 0.0205]
ϵ-based ($\epsilon = 4$)	0.0288 [0.0288, 0.0288]	0.0238 [0.0237, 0.0241]	0.0212 [0.0209, 0.0215]	0.0218 [0.0218, 0.0218]	0.0180 [0.0177, 0.0183]	0.0189 [0.0187, 0.0190]
ϵ-based ($\epsilon = 1.8$)	0.0288 [0.0288, 0.0288]	0.0238 [0.0237, 0.0241]	0.0212 [0.0209, 0.0215]	0.0218 [0.0218, 0.0218]	0.0180 [0.0177, 0.0183]	0.0219 [0.0215, 0.0223]
Unique-2 ($p = -1$)	0.0288 [0.0288, 0.0288]	0.0238 [0.0237, 0.0241]	0.0212 [0.0209, 0.0215]	0.0218 [0.0218, 0.0218]	0.0180 [0.0177, 0.0183]	0.0218 [0.0217, 0.0218]
Unique-3 ($p = -1$)	0.0288 [0.0288, 0.0288]	0.0238 [0.0237, 0.0241]	0.0212 [0.0209, 0.0215]	0.0218 [0.0218, 0.0218]	0.0180 [0.0177, 0.0183]	0.0213 [0.0210, 0.0215]
Unique-3 ($p = +1$)	0.0288 [0.0288, 0.0288]	0.0238 [0.0237, 0.0241]	0.0212 [0.0209, 0.0215]	0.0218 [0.0218, 0.0218]	0.0180 [0.0177, 0.0183]	0.0210 [0.0207, 0.0213]
Unique-3 ($p = 0$)	0.0288 [0.0288, 0.0288]	0.0238 [0.0237, 0.0241]	0.0212 [0.0209, 0.0215]	0.0218 [0.0218, 0.0218]	0.0180 [0.0177, 0.0183]	0.0212 [0.0210, 0.0214]
Unique-4 ($p = -1$)	0.0288 [0.0288, 0.0288]	0.0238 [0.0237, 0.0241]	0.0212 [0.0209, 0.0215]	0.0218 [0.0218, 0.0218]	0.0180 [0.0177, 0.0183]	0.0206 [0.0205, 0.0207]
Unique-4 ($p = 0$)	0.0288 [0.0288, 0.0288]	0.0238 [0.0237, 0.0241]	0.0212 [0.0209, 0.0215]	0.0218 [0.0218, 0.0218]	0.0180 [0.0177, 0.0183]	0.0207 [0.0204, 0.0210]

(b) With curvature features

Graph construction	LinReg	Small MLP	Large MLP	Grad. boost.	Transformer	GNN
Fully connected	0.0288 [0.0288, 0.0288]	0.0241 [0.0237, 0.0245]	0.0210 [0.0207, 0.0213]	0.0218 [0.0218, 0.0218]	0.0172 [0.0169, 0.0175]	0.0153 [0.0148, 0.0157]
kNN ($k = 2$)	0.0285 [0.0285, 0.0285]	0.0233 [0.0231, 0.0236]	0.0202 [0.0201, 0.0203]	0.0211 [0.0211, 0.0211]	0.0157 [0.0155, 0.0162]	0.0185 [0.0184, 0.0187]
kNN ($k = 3$)	0.0280 [0.0280, 0.0280]	0.0228 [0.0226, 0.0230]	0.0202 [0.0200, 0.0204]	0.0211 [0.0211, 0.0211]	0.0159 [0.0154, 0.0163]	0.0190 [0.0189, 0.0192]
kNN ($k = 4$)	0.0273 [0.0273, 0.0273]	0.0222 [0.0218, 0.0227]	0.0189 [0.0187, 0.0190]	0.0204 [0.0204, 0.0204]	0.0138 [0.0136, 0.0142]	0.0184 [0.0181, 0.0187]
kNN ($k = 5$)	0.0271 [0.0271, 0.0271]	0.0220 [0.0219, 0.0221]	0.0193 [0.0192, 0.0193]	0.0210 [0.0210, 0.0210]	0.0152 [0.0148, 0.0156]	0.0194 [0.0193, 0.0196]
k_T threshold ($p = -1$)	0.0282 [0.0282, 0.0282]	0.0234 [0.0230, 0.0238]	0.0208 [0.0206, 0.0209]	0.0216 [0.0216, 0.0216]	0.0171 [0.0169, 0.0174]	0.0203 [0.0201, 0.0204]
k_T threshold ($p = +1$)	0.0286 [0.0286, 0.0286]	0.0235 [0.0232, 0.0238]	0.0206 [0.0204, 0.0208]	0.0215 [0.0215, 0.0215]	0.0167 [0.0164, 0.0171]	0.0192 [0.0190, 0.0194]
k_T threshold ($p = 0$)	0.0279 [0.0279, 0.0279]	0.0229 [0.0226, 0.0232]	0.0200 [0.0200, 0.0201]	0.0212 [0.0212, 0.0212]	0.0154 [0.0151, 0.0156]	0.0183 [0.0182, 0.0184]
Laman ($p = -1$)	0.0286 [0.0286, 0.0286]	0.0227 [0.0225, 0.0230]	0.0202 [0.0200, 0.0205]	0.0208 [0.0208, 0.0208]	0.0164 [0.0163, 0.0165]	0.0205 [0.0203, 0.0206]
Laman ($p = +1$)	0.0284 [0.0284, 0.0284]	0.0232 [0.0230, 0.0233]	0.0208 [0.0206, 0.0212]	0.0215 [0.0215, 0.0215]	0.0164 [0.0160, 0.0169]	0.0196 [0.0193, 0.0198]
Laman ($p = 0$)	0.0284 [0.0284, 0.0284]	0.0232 [0.0230, 0.0233]	0.0208 [0.0206, 0.0212]	0.0215 [0.0215, 0.0215]	0.0164 [0.0160, 0.0169]	0.0197 [0.0195, 0.0199]
ϵ-based ($\epsilon = 1.5$)	0.0282 [0.0282, 0.0282]	0.0207 [0.0204, 0.0211]	0.0165 [0.0163, 0.0167]	0.0186 [0.0186, 0.0186]	0.0115 [0.0114, 0.0117]	0.0154 [0.0151, 0.0156]
ϵ-based ($\epsilon = 2.5$)	0.0280 [0.0280, 0.0280]	0.0222 [0.0219, 0.0224]	0.0187 [0.0185, 0.0189]	0.0203 [0.0203, 0.0203]	0.0136 [0.0134, 0.0139]	0.0166 [0.0163, 0.0168]
ϵ-based ($\epsilon = 3$)	0.0271 [0.0271, 0.0271]	0.0218 [0.0215, 0.0222]	0.0185 [0.0184, 0.0186]	0.0203 [0.0203, 0.0203]	0.0137 [0.0135, 0.0140]	0.0162 [0.0160, 0.0163]
ϵ-based ($\epsilon = 4$)	0.0271 [0.0271, 0.0271]	0.0229 [0.0227, 0.0230]	0.0202 [0.0201, 0.0203]	0.0208 [0.0208, 0.0208]	0.0149 [0.0145, 0.0153]	0.0171 [0.0169, 0.0173]
ϵ-based ($\epsilon = 1.8$)	0.0261 [0.0261, 0.0261]	0.0193 [0.0190, 0.0195]	0.0139 [0.0134, 0.0146]	0.0160 [0.0160, 0.0160]	0.0095 [0.0093, 0.0097]	0.0097 [0.0095, 0.0098]
Unique-2 ($p = -1$)	0.0286 [0.0286, 0.0286]	0.0227 [0.0224, 0.0230]	0.0202 [0.0200, 0.0205]	0.0208 [0.0208, 0.0208]	0.0164 [0.0163, 0.0165]	0.0206 [0.0204, 0.0207]
Unique-3 ($p = -1$)	0.0285 [0.0285, 0.0285]	0.0233 [0.0231, 0.0235]	0.0209 [0.0207, 0.0212]	0.0216 [0.0216, 0.0216]	0.0171 [0.0168, 0.0173]	0.0209 [0.0208, 0.0211]
Unique-3 ($p = +1$)	0.0286 [0.0286, 0.0286]	0.0234 [0.0233, 0.0236]	0.0208 [0.0207, 0.0209]	0.0216 [0.0216, 0.0216]	0.0168 [0.0166, 0.0171]	0.0205 [0.0203, 0.0206]
Unique-3 ($p = 0$)	0.0286 [0.0286, 0.0286]	0.0234 [0.0233, 0.0236]	0.0208 [0.0207, 0.0209]	0.0216 [0.0216, 0.0216]	0.0168 [0.0166, 0.0171]	0.0205 [0.0203, 0.0208]
Unique-4 ($p = -1$)	0.0285 [0.0285, 0.0285]	0.0232 [0.0231, 0.0233]	0.0207 [0.0206, 0.0208]	0.0216 [0.0216, 0.0216]	0.0176 [0.0174, 0.0178]	0.0203 [0.0201, 0.0206]
Unique-4 ($p = 0$)	0.0282 [0.0282, 0.0282]	0.0234 [0.0232, 0.0237]	0.0203 [0.0202, 0.0204]	0.0215 [0.0215, 0.0215]	0.0165 [0.0163, 0.0167]	0.0190 [0.0189, 0.0192]

Table 23: Reduction in test MAE from adding Ricci curvature features for QM9 HOMO−LUMO gap, Equation 7, so positive values mean curvature lowered the error, computed per seed between runs that share architecture, hyperparameters, data and seed. Entries give the mean over 5 training seeds; brackets give the 95% percentile bootstrap confidence interval of that mean, obtained by resampling the seeds with replacement 10,000 times, averaging each resample, and taking the 2.5th and 97.5th percentiles of the resampled means. Seeds are resampled jointly for both settings, so the pairing is preserved. Bold marks entries where the 95% interval does not contain zero, indicating a consistent improvement (positive) or degradation (negative) from adding curvature.

Graph construction	LinReg	Small MLP	Large MLP	Grad. boost.	Transformer	GNN
Fully connected	**+0.0000** [+0.0000, +0.0000]	−0.0003 [−0.0007, +0.0003]	+0.0002 [−0.0003, +0.0007]	**+0.0000** [+0.0000, +0.0000]	**+0.0008** [+0.0004, +0.0011]	**+0.0007** [+0.0002, +0.0012]
kNN ($k = 2$)	**+0.0004** [+0.0004, +0.0004]	**+0.0006** [+0.0001, +0.0010]	**+0.0010** [+0.0007, +0.0013]	**+0.0007** [+0.0007, +0.0007]	**+0.0023** [+0.0019, +0.0026]	**+0.0038** [+0.0036, +0.0040]
kNN ($k = 3$)	**+0.0009** [+0.0009, +0.0009]	**+0.0010** [+0.0007, +0.0015]	**+0.0010** [+0.0007, +0.0013]	**+0.0007** [+0.0007, +0.0007]	**+0.0021** [+0.0018, +0.0024]	**+0.0027** [+0.0024, +0.0030]
kNN ($k = 4$)	**+0.0015** [+0.0015, +0.0015]	**+0.0016** [+0.0011, +0.0022]	**+0.0024** [+0.0020, +0.0027]	**+0.0014** [+0.0014, +0.0014]	**+0.0042** [+0.0039, +0.0045]	**+0.0028** [+0.0025, +0.0031]
kNN ($k = 5$)	**+0.0017** [+0.0017, +0.0017]	**+0.0019** [+0.0017, +0.0021]	**+0.0020** [+0.0017, +0.0022]	**+0.0009** [+0.0009, +0.0009]	**+0.0028** [+0.0025, +0.0033]	**+0.0016** [+0.0013, +0.0019]
k_T threshold ($p = -1$)	**+0.0007** [+0.0007, +0.0007]	+0.0005 [−0.0000, +0.0011]	**+0.0005** [+0.0002, +0.0008]	**+0.0002** [+0.0002, +0.0002]	**+0.0009** [+0.0004, +0.0013]	+0.0000 [−0.0002, +0.0003]
k_T threshold ($p = +1$)	**+0.0002** [+0.0002, +0.0002]	+0.0004 [−0.0001, +0.0009]	**+0.0006** [+0.0003, +0.0010]	**+0.0003** [+0.0003, +0.0003]	**+0.0013** [+0.0009, +0.0017]	**+0.0010** [+0.0006, +0.0013]
k_T threshold ($p = 0$)	**+0.0010** [+0.0010, +0.0010]	**+0.0010** [+0.0005, +0.0015]	**+0.0012** [+0.0009, +0.0014]	**+0.0006** [+0.0006, +0.0006]	**+0.0026** [+0.0025, +0.0027]	**+0.0021** [+0.0018, +0.0025]
Laman ($p = -1$)	**+0.0002** [+0.0002, +0.0002]	**+0.0011** [+0.0007, +0.0016]	**+0.0010** [+0.0004, +0.0015]	**+0.0011** [+0.0011, +0.0011]	**+0.0016** [+0.0013, +0.0019]	**+0.0012** [+0.0010, +0.0015]
Laman ($p = +1$)	**+0.0005** [+0.0005, +0.0005]	**+0.0007** [+0.0004, +0.0010]	+0.0004 [−0.0003, +0.0009]	**+0.0004** [+0.0004, +0.0004]	**+0.0016** [+0.0012, +0.0019]	**+0.0023** [+0.0021, +0.0025]
Laman ($p = 0$)	**+0.0005** [+0.0005, +0.0005]	**+0.0007** [+0.0004, +0.0010]	+0.0004 [−0.0003, +0.0009]	**+0.0004** [+0.0004, +0.0004]	**+0.0016** [+0.0012, +0.0019]	**+0.0021** [+0.0018, +0.0023]
ϵ-based ($\epsilon = 1.5$)	**+0.0007** [+0.0007, +0.0007]	**+0.0032** [+0.0027, +0.0037]	**+0.0047** [+0.0044, +0.0052]	**+0.0032** [+0.0032, +0.0032]	**+0.0065** [+0.0063, +0.0067]	**+0.0073** [+0.0071, +0.0077]
ϵ-based ($\epsilon = 2.5$)	**+0.0008** [+0.0008, +0.0008]	**+0.0017** [+0.0014, +0.0019]	**+0.0025** [+0.0021, +0.0029]	**+0.0015** [+0.0015, +0.0015]	**+0.0044** [+0.0042, +0.0045]	**+0.0046** [+0.0044, +0.0049]
ϵ-based ($\epsilon = 3$)	**+0.0017** [+0.0017, +0.0017]	**+0.0020** [+0.0015, +0.0026]	**+0.0027** [+0.0025, +0.0030]	**+0.0016** [+0.0016, +0.0016]	**+0.0043** [+0.0039, +0.0046]	**+0.0041** [+0.0038, +0.0043]
ϵ-based ($\epsilon = 4$)	**+0.0017** [+0.0017, +0.0017]	**+0.0009** [+0.0007, +0.0012]	**+0.0010** [+0.0008, +0.0013]	**+0.0010** [+0.0010, +0.0010]	**+0.0031** [+0.0028, +0.0035]	**+0.0017** [+0.0015, +0.0020]
ϵ-based ($\epsilon = 1.8$)	**+0.0027** [+0.0027, +0.0027]	**+0.0046** [+0.0042, +0.0050]	**+0.0073** [+0.0064, +0.0079]	**+0.0058** [+0.0058, +0.0058]	**+0.0085** [+0.0083, +0.0087]	**+0.0123** [+0.0117, +0.0126]
Unique-2 ($p = -1$)	**+0.0002** [+0.0002, +0.0002]	**+0.0011** [+0.0007, +0.0017]	**+0.0010** [+0.0004, +0.0015]	**+0.0011** [+0.0011, +0.0011]	**+0.0016** [+0.0013, +0.0019]	**+0.0012** [+0.0010, +0.0013]
Unique-3 ($p = -1$)	**+0.0003** [+0.0003, +0.0003]	**+0.0006** [+0.0004, +0.0007]	+0.0003 [−0.0002, +0.0007]	**+0.0002** [+0.0002, +0.0002]	**+0.0009** [+0.0004, +0.0014]	**+0.0004** [+0.0000, +0.0007]
Unique-3 ($p = +1$)	**+0.0002** [+0.0002, +0.0002]	**+0.0004** [+0.0003, +0.0006]	**+0.0004** [+0.0002, +0.0007]	**+0.0002** [+0.0002, +0.0002]	**+0.0012** [+0.0008, +0.0015]	**+0.0006** [+0.0003, +0.0009]
Unique-3 ($p = 0$)	**+0.0002** [+0.0002, +0.0002]	**+0.0004** [+0.0003, +0.0006]	**+0.0004** [+0.0002, +0.0007]	**+0.0002** [+0.0002, +0.0002]	**+0.0012** [+0.0008, +0.0015]	**+0.0007** [+0.0005, +0.0010]
Unique-4 ($p = -1$)	**+0.0004** [+0.0004, +0.0004]	**+0.0006** [+0.0004, +0.0009]	**+0.0005** [+0.0003, +0.0008]	**+0.0002** [+0.0002, +0.0002]	+0.0004 [−0.0000, +0.0008]	+0.0002 [−0.0000, +0.0006]
Unique-4 ($p = 0$)	**+0.0007** [+0.0007, +0.0007]	**+0.0005** [+0.0001, +0.0008]	**+0.0009** [+0.0007, +0.0011]	**+0.0004** [+0.0004, +0.0004]	**+0.0015** [+0.0012, +0.0018]	**+0.0017** [+0.0013, +0.0020]

Table 24: Test MAE (lower is better) for QM9 dipole moment μ for each graph construction and model class, (a) without and (b) with FRC and ORC features. Entries give the mean over 5 training seeds; brackets give the 95% percentile bootstrap confidence interval of that mean, obtained by resampling the seeds with replacement 10,000 times, averaging each resample, and taking the 2.5th and 97.5th percentiles of the resampled means.

(a) Without curvature features

Graph construction	LinReg	Small MLP	Large MLP	Grad. boost.	Transformer	GNN
Fully connected	0.9404 [0.9404, 0.9404]	0.8616 [0.8594, 0.8639]	0.8047 [0.7970, 0.8143]	0.7901 [0.7901, 0.7901]	0.5076 [0.4966, 0.5212]	0.4970 [0.4933, 0.5007]
kNN ($k = 2$)	0.9404 [0.9404, 0.9404]	0.8616 [0.8594, 0.8639]	0.8047 [0.7970, 0.8143]	0.7901 [0.7901, 0.7901]	0.5076 [0.4966, 0.5212]	0.6731 [0.6705, 0.6757]
kNN ($k = 3$)	0.9404 [0.9404, 0.9404]	0.8616 [0.8594, 0.8639]	0.8047 [0.7970, 0.8143]	0.7901 [0.7901, 0.7901]	0.5076 [0.4966, 0.5211]	0.6632 [0.6567, 0.6717]
kNN ($k = 4$)	0.9404 [0.9404, 0.9404]	0.8616 [0.8594, 0.8639]	0.8047 [0.7970, 0.8137]	0.7901 [0.7901, 0.7901]	0.5076 [0.4966, 0.5212]	0.6490 [0.6457, 0.6526]
kNN ($k = 5$)	0.9404 [0.9404, 0.9404]	0.8616 [0.8594, 0.8639]	0.8047 [0.7970, 0.8143]	0.7901 [0.7901, 0.7901]	0.5076 [0.4966, 0.5212]	0.6487 [0.6433, 0.6531]
k_T threshold ($p = -1$)	0.9404 [0.9404, 0.9404]	0.8616 [0.8594, 0.8639]	0.8047 [0.7970, 0.8143]	0.7901 [0.7901, 0.7901]	0.5076 [0.4966, 0.5212]	0.6396 [0.6308, 0.6481]
k_T threshold ($p = +1$)	0.9404 [0.9404, 0.9404]	0.8616 [0.8594, 0.8639]	0.8047 [0.7970, 0.8137]	0.7901 [0.7901, 0.7901]	0.5076 [0.4966, 0.5212]	0.6375 [0.6332, 0.6416]
k_T threshold ($p = 0$)	0.9404 [0.9404, 0.9404]	0.8616 [0.8594, 0.8639]	0.8047 [0.7970, 0.8137]	0.7901 [0.7901, 0.7901]	0.5076 [0.4966, 0.5212]	0.6381 [0.6302, 0.6447]
Laman ($p = -1$)	0.9404 [0.9404, 0.9404]	0.8616 [0.8594, 0.8639]	0.8047 [0.7970, 0.8143]	0.7901 [0.7901, 0.7901]	0.5076 [0.4966, 0.5212]	0.6786 [0.6681, 0.6892]
Laman ($p = +1$)	0.9404 [0.9404, 0.9404]	0.8616 [0.8594, 0.8639]	0.8047 [0.7970, 0.8143]	0.7901 [0.7901, 0.7901]	0.5076 [0.4966, 0.5212]	0.6715 [0.6670, 0.6760]
Laman ($p = 0$)	0.9404 [0.9404, 0.9404]	0.8616 [0.8594, 0.8639]	0.8047 [0.7970, 0.8137]	0.7901 [0.7901, 0.7901]	0.5076 [0.4966, 0.5212]	0.6603 [0.6536, 0.6672]
ϵ-based ($\epsilon = 1.5$)	0.9404 [0.9404, 0.9404]	0.8616 [0.8594, 0.8639]	0.8047 [0.7970, 0.8137]	0.7901 [0.7901, 0.7901]	0.5076 [0.4966, 0.5212]	0.7065 [0.6981, 0.7145]
ϵ-based ($\epsilon = 2.5$)	0.9404 [0.9404, 0.9404]	0.8616 [0.8594, 0.8639]	0.8047 [0.7970, 0.8143]	0.7901 [0.7901, 0.7901]	0.5076 [0.4966, 0.5212]	0.6608 [0.6561, 0.6641]
ϵ-based ($\epsilon = 3$)	0.9404 [0.9404, 0.9404]	0.8616 [0.8594, 0.8639]	0.8047 [0.7970, 0.8143]	0.7901 [0.7901, 0.7901]	0.5076 [0.4966, 0.5212]	0.6378 [0.6282, 0.6475]
ϵ-based ($\epsilon = 4$)	0.9404 [0.9404, 0.9404]	0.8616 [0.8594, 0.8639]	0.8047 [0.7970, 0.8143]	0.7901 [0.7901, 0.7901]	0.5076 [0.4966, 0.5212]	0.5947 [0.5899, 0.5995]
ϵ-based ($\epsilon = 1.8$)	0.9404 [0.9404, 0.9404]	0.8616 [0.8594, 0.8639]	0.8047 [0.7970, 0.8143]	0.7901 [0.7901, 0.7901]	0.5076 [0.4966, 0.5212]	0.6883 [0.6734, 0.7030]
Unique-2 ($p = -1$)	0.9404 [0.9404, 0.9404]	0.8616 [0.8594, 0.8639]	0.8047 [0.7970, 0.8143]	0.7901 [0.7901, 0.7901]	0.5076 [0.4966, 0.5212]	0.6752 [0.6673, 0.6837]
Unique-3 ($p = -1$)	0.9404 [0.9404, 0.9404]	0.8616 [0.8594, 0.8639]	0.8047 [0.7970, 0.8137]	0.7901 [0.7901, 0.7901]	0.5076 [0.4966, 0.5212]	0.6596 [0.6550, 0.6647]
Unique-3 ($p = +1$)	0.9404 [0.9404, 0.9404]	0.8616 [0.8594, 0.8639]	0.8047 [0.7970, 0.8143]	0.7901 [0.7901, 0.7901]	0.5076 [0.4966, 0.5212]	0.6533 [0.6457, 0.6617]
Unique-3 ($p = 0$)	0.9404 [0.9404, 0.9404]	0.8616 [0.8594, 0.8639]	0.8047 [0.7970, 0.8137]	0.7901 [0.7901, 0.7901]	0.5076 [0.4966, 0.5212]	0.6591 [0.6497, 0.6669]
Unique-4 ($p = -1$)	0.9404 [0.9404, 0.9404]	0.8616 [0.8594, 0.8639]	0.8047 [0.7970, 0.8137]	0.7901 [0.7901, 0.7901]	0.5076 [0.4966, 0.5212]	0.6412 [0.6342, 0.6464]
Unique-4 ($p = 0$)	0.9404 [0.9404, 0.9404]	0.8616 [0.8594, 0.8639]	0.8047 [0.7970, 0.8137]	0.7901 [0.7901, 0.7901]	0.5076 [0.4966, 0.5212]	0.6465 [0.6439, 0.6488]

(b) With curvature features

Graph construction	LinReg	Small MLP	Large MLP	Grad. boost.	Transformer	GNN
Fully connected	0.9413 [0.9413, 0.9413]	0.8588 [0.8566, 0.8610]	0.8018 [0.8004, 0.8030]	0.7893 [0.7893, 0.7893]	0.5117 [0.5011, 0.5232]	0.4937 [0.4896, 0.4985]
kNN ($k = 2$)	0.9286 [0.9286, 0.9286]	0.8474 [0.8451, 0.8498]	0.7979 [0.7947, 0.8011]	0.7811 [0.7811, 0.7811]	0.4868 [0.4628, 0.5091]	0.6894 [0.6869, 0.6931]
kNN ($k = 3$)	0.9336 [0.9336, 0.9336]	0.8488 [0.8468, 0.8506]	0.7842 [0.7809, 0.7872]	0.7787 [0.7787, 0.7787]	0.4962 [0.4859, 0.5066]	0.6789 [0.6677, 0.6896]
kNN ($k = 4$)	0.9346 [0.9346, 0.9346]	0.8429 [0.8409, 0.8453]	0.7855 [0.7769, 0.7977]	0.7716 [0.7716, 0.7716]	0.4962 [0.4879, 0.5024]	0.6657 [0.6569, 0.6735]
kNN ($k = 5$)	0.9357 [0.9357, 0.9357]	0.8480 [0.8460, 0.8497]	0.7913 [0.7877, 0.7949]	0.7782 [0.7782, 0.7782]	0.5029 [0.4973, 0.5091]	0.6548 [0.6462, 0.6633]
k_T threshold ($p = -1$)	0.9028 [0.9028, 0.9028]	0.8257 [0.8231, 0.8294]	0.7753 [0.7738, 0.7775]	0.7744 [0.7744, 0.7744]	0.5138 [0.5073, 0.5207]	0.6595 [0.6561, 0.6620]
k_T threshold ($p = +1$)	0.9000 [0.9000, 0.9000]	0.7931 [0.7902, 0.7960]	0.7448 [0.7429, 0.7468]	0.7423 [0.7423, 0.7423]	0.5078 [0.4947, 0.5216]	0.6542 [0.6522, 0.6564]
k_T threshold ($p = 0$)	0.9266 [0.9266, 0.9266]	0.8411 [0.8384, 0.8433]	0.7845 [0.7809, 0.7881]	0.7734 [0.7734, 0.7734]	0.5110 [0.4980, 0.5223]	0.6554 [0.6501, 0.6596]
Laman ($p = -1$)	0.9290 [0.9290, 0.9290]	0.8262 [0.8250, 0.8275]	0.7798 [0.7754, 0.7835]	0.7695 [0.7695, 0.7695]	0.5044 [0.4881, 0.5165]	0.6959 [0.6926, 0.6993]
Laman ($p = +1$)	0.9361 [0.9361, 0.9361]	0.8547 [0.8537, 0.8556]	0.8022 [0.7977, 0.8064]	0.7876 [0.7876, 0.7876]	0.4922 [0.4786, 0.5011]	0.6858 [0.6814, 0.6895]
Laman ($p = 0$)	0.9361 [0.9361, 0.9361]	0.8547 [0.8537, 0.8556]	0.8022 [0.7977, 0.8064]	0.7876 [0.7876, 0.7876]	0.4922 [0.4786, 0.5013]	0.6933 [0.6884, 0.6987]
ϵ-based ($\epsilon = 1.5$)	0.9264 [0.9264, 0.9264]	0.8213 [0.8195, 0.8237]	0.7601 [0.7574, 0.7623]	0.7488 [0.7488, 0.7488]	0.4558 [0.4378, 0.4692]	0.6576 [0.6536, 0.6625]
ϵ-based ($\epsilon = 2.5$)	0.9327 [0.9327, 0.9327]	0.8354 [0.8330, 0.8378]	0.7783 [0.7752, 0.7818]	0.7613 [0.7613, 0.7613]	0.4902 [0.4812, 0.4977]	0.6696 [0.6658, 0.6719]
ϵ-based ($\epsilon = 3$)	0.9049 [0.9049, 0.9049]	0.8236 [0.8197, 0.8263]	0.7616 [0.7580, 0.7654]	0.7557 [0.7557, 0.7557]	0.4948 [0.4758, 0.5066]	0.6399 [0.6367, 0.6431]
ϵ-based ($\epsilon = 4$)	0.9379 [0.9379, 0.9379]	0.8347 [0.8305, 0.8378]	0.7791 [0.7769, 0.7810]	0.7694 [0.7694, 0.7694]	0.4866 [0.4763, 0.4960]	0.5912 [0.5872, 0.5939]
ϵ-based ($\epsilon = 1.8$)	0.8831 [0.8831, 0.8831]	0.7863 [0.7835, 0.7896]	0.7153 [0.7140, 0.7171]	0.7188 [0.7188, 0.7188]	0.4477 [0.4393, 0.4585]	0.6154 [0.6112, 0.6205]
Unique-2 ($p = -1$)	0.9290 [0.9290, 0.9290]	0.8262 [0.8250, 0.8275]	0.7798 [0.7754, 0.7835]	0.7695 [0.7695, 0.7695]	0.5044 [0.4881, 0.5165]	0.6966 [0.6943, 0.6982]
Unique-3 ($p = -1$)	0.9206 [0.9206, 0.9206]	0.8318 [0.8291, 0.8354]	0.7814 [0.7800, 0.7827]	0.7731 [0.7731, 0.7731]	0.5171 [0.5043, 0.5285]	0.6713 [0.6682, 0.6748]
Unique-3 ($p = +1$)	0.9371 [0.9371, 0.9371]	0.8507 [0.8490, 0.8519]	0.7975 [0.7961, 0.7990]	0.7869 [0.7869, 0.7869]	0.5015 [0.4948, 0.5080]	0.6773 [0.6698, 0.6844]
Unique-3 ($p = 0$)	0.9371 [0.9371, 0.9371]	0.8507 [0.8491, 0.8519]	0.7975 [0.7961, 0.7990]	0.7869 [0.7869, 0.7869]	0.5015 [0.4950, 0.5079]	0.6691 [0.6586, 0.6789]
Unique-4 ($p = -1$)	0.9141 [0.9141, 0.9141]	0.8282 [0.8261, 0.8304]	0.7789 [0.7781, 0.7798]	0.7687 [0.7687, 0.7687]	0.5220 [0.5102, 0.5303]	0.6525 [0.6483, 0.6560]
Unique-4 ($p = 0$)	0.9353 [0.9353, 0.9353]	0.8480 [0.8466, 0.8492]	0.7960 [0.7942, 0.7975]	0.7793 [0.7793, 0.7793]	0.5052 [0.4993, 0.5096]	0.6561 [0.6522, 0.6605]

Table 25: Reduction in test MAE from adding Ricci curvature features for QM9 dipole moment μ, Equation 7, so positive values mean curvature lowered the error, computed per seed between runs that share architecture, hyperparameters, data and seed. Entries give the mean over 5 training seeds. Brackets give the 95% percentile bootstrap confidence interval of that mean, obtained by resampling the seeds with replacement 10,000 times, averaging each resample, and taking the 2.5th and 97.5th percentiles of the resampled means. Seeds are resampled jointly for both settings, so the pairing is preserved. Bold marks entries where the 95% interval does not contain zero, indicating a consistent improvement (positive) or degradation (negative) from adding curvature.

Graph construction	LinReg	Small MLP	Large MLP	Grad. boost.	Transformer	GNN
Fully connected	**−0.0009** [−0.0009, −0.0009]	+0.0028 [−0.0013, +0.0071]	+0.0029 [−0.0051, +0.0116]	**+0.0008** [+0.0008, +0.0008]	−0.0041 [−0.0130, +0.0052]	+0.0033 [−0.0014, +0.0079]
kNN ($k = 2$)	**+0.0118** [+0.0118, +0.0118]	**+0.0142** [+0.0101, +0.0184]	+0.0068 [−0.0026, +0.0181]	**+0.0089** [+0.0089, +0.0089]	+0.0208 [−0.0124, +0.0562]	**−0.0163** [−0.0224, −0.0122]
kNN ($k = 3$)	**+0.0067** [+0.0067, +0.0067]	**+0.0129** [+0.0109, +0.0148]	**+0.0205** [+0.0098, +0.0334]	**+0.0114** [+0.0114, +0.0114]	+0.0114 [−0.0042, +0.0246]	−0.0156 [−0.0321, +0.0015]
kNN ($k = 4$)	**+0.0058** [+0.0058, +0.0058]	**+0.0187** [+0.0147, +0.0227]	**+0.0192** [+0.0027, +0.0337]	**+0.0185** [+0.0185, +0.0185]	+0.0114 [−0.0034, +0.0262]	**−0.0168** [−0.0233, −0.0084]
kNN ($k = 5$)	**+0.0047** [+0.0047, +0.0047]	**+0.0136** [+0.0100, +0.0173]	**+0.0134** [+0.0043, +0.0224]	**+0.0119** [+0.0119, +0.0119]	+0.0047 [−0.0065, +0.0159]	−0.0060 [−0.0133, +0.0012]
k_T threshold ($p = -1$)	**+0.0376** [+0.0376, +0.0376]	**+0.0360** [+0.0310, +0.0400]	**+0.0294** [+0.0205, +0.0398]	**+0.0157** [+0.0157, +0.0157]	−0.0062 [−0.0177, +0.0054]	**−0.0199** [−0.0271, −0.0126]
k_T threshold ($p = +1$)	**+0.0403** [+0.0403, +0.0403]	**+0.0685** [+0.0643, +0.0728]	**+0.0599** [+0.0534, +0.0669]	**+0.0478** [+0.0478, +0.0478]	−0.0002 [−0.0216, +0.0212]	**−0.0168** [−0.0210, −0.0129]
k_T threshold ($p = 0$)	**+0.0138** [+0.0138, +0.0138]	**+0.0205** [+0.0163, +0.0255]	**+0.0202** [+0.0118, +0.0305]	**+0.0167** [+0.0167, +0.0167]	−0.0034 [−0.0179, +0.0124]	**−0.0173** [−0.0209, −0.0139]
Laman ($p = -1$)	**+0.0114** [+0.0114, +0.0114]	**+0.0355** [+0.0328, +0.0379]	**+0.0249** [+0.0158, +0.0341]	**+0.0206** [+0.0206, +0.0206]	+0.0032 [−0.0121, +0.0215]	**−0.0173** [−0.0291, −0.0058]
Laman ($p = +1$)	**+0.0042** [+0.0042, +0.0042]	**+0.0069** [+0.0044, +0.0100]	+0.0025 [−0.0073, +0.0123]	**+0.0025** [+0.0025, +0.0025]	+0.0154 [−0.0019, +0.0333]	**−0.0143** [−0.0203, −0.0091]
Laman ($p = 0$)	**+0.0042** [+0.0042, +0.0042]	**+0.0069** [+0.0044, +0.0100]	+0.0025 [−0.0073, +0.0123]	**+0.0025** [+0.0025, +0.0025]	+0.0154 [−0.0019, +0.0333]	**−0.0330** [−0.0431, −0.0230]
ϵ-based ($\epsilon = 1.5$)	**+0.0139** [+0.0139, +0.0139]	**+0.0404** [+0.0369, +0.0442]	**+0.0446** [+0.0354, +0.0558]	**+0.0412** [+0.0412, +0.0412]	**+0.0518** [+0.0329, +0.0734]	**+0.0489** [+0.0366, +0.0591]
ϵ-based ($\epsilon = 2.5$)	**+0.0076** [+0.0076, +0.0076]	**+0.0263** [+0.0230, +0.0300]	**+0.0264** [+0.0186, +0.0343]	**+0.0287** [+0.0287, +0.0287]	**+0.0174** [+0.0029, +0.0319]	**−0.0088** [−0.0149, −0.0033]
ϵ-based ($\epsilon = 3$)	**+0.0355** [+0.0355, +0.0355]	**+0.0380** [+0.0343, +0.0435]	**+0.0431** [+0.0337, +0.0525]	**+0.0343** [+0.0343, +0.0343]	+0.0128 [−0.0093, +0.0377]	−0.0021 [−0.0111, +0.0058]
ϵ-based ($\epsilon = 4$)	**+0.0024** [+0.0024, +0.0024]	**+0.0269** [+0.0223, +0.0329]	**+0.0256** [+0.0171, +0.0341]	**+0.0207** [+0.0207, +0.0207]	**+0.0210** [+0.0085, +0.0356]	+0.0036 [−0.0022, +0.0081]
ϵ-based ($\epsilon = 1.8$)	**+0.0573** [+0.0573, +0.0573]	**+0.0753** [+0.0703, +0.0801]	**+0.0894** [+0.0805, +0.0992]	**+0.0712** [+0.0712, +0.0712]	**+0.0599** [+0.0398, +0.0816]	**+0.0730** [+0.0599, +0.0838]
Unique-2 ($p = -1$)	**+0.0114** [+0.0114, +0.0114]	**+0.0355** [+0.0328, +0.0379]	**+0.0249** [+0.0158, +0.0343]	**+0.0206** [+0.0206, +0.0206]	+0.0032 [−0.0121, +0.0215]	**−0.0214** [−0.0295, −0.0132]
Unique-3 ($p = -1$)	**+0.0197** [+0.0197, +0.0197]	**+0.0298** [+0.0268, +0.0331]	**+0.0233** [+0.0149, +0.0323]	**+0.0170** [+0.0170, +0.0170]	−0.0095 [−0.0222, +0.0058]	**−0.0116** [−0.0162, −0.0055]
Unique-3 ($p = +1$)	**+0.0032** [+0.0032, +0.0032]	**+0.0110** [+0.0074, +0.0149]	**+0.0072** [+0.0004, +0.0163]	**+0.0031** [+0.0031, +0.0031]	+0.0061 [−0.0056, +0.0173]	**−0.0240** [−0.0273, −0.0204]
Unique-3 ($p = 0$)	**+0.0032** [+0.0032, +0.0032]	**+0.0110** [+0.0075, +0.0145]	**+0.0072** [+0.0004, +0.0163]	**+0.0031** [+0.0031, +0.0031]	+0.0061 [−0.0073, +0.0173]	−0.0101 [−0.0190, +0.0008]
Unique-4 ($p = -1$)	**+0.0263** [+0.0263, +0.0263]	**+0.0334** [+0.0297, +0.0370]	**+0.0258** [+0.0180, +0.0346]	**+0.0214** [+0.0214, +0.0214]	−0.0144 [−0.0303, +0.0030]	**−0.0113** [−0.0142, −0.0086]
Unique-4 ($p = 0$)	**+0.0051** [+0.0051, +0.0051]	**+0.0137** [+0.0105, +0.0166]	**+0.0087** [+0.0015, +0.0163]	**+0.0108** [+0.0108, +0.0108]	+0.0024 [−0.0113, +0.0161]	**−0.0095** [−0.0145, −0.0045]

G.2 JETSET: b VS LIGHT

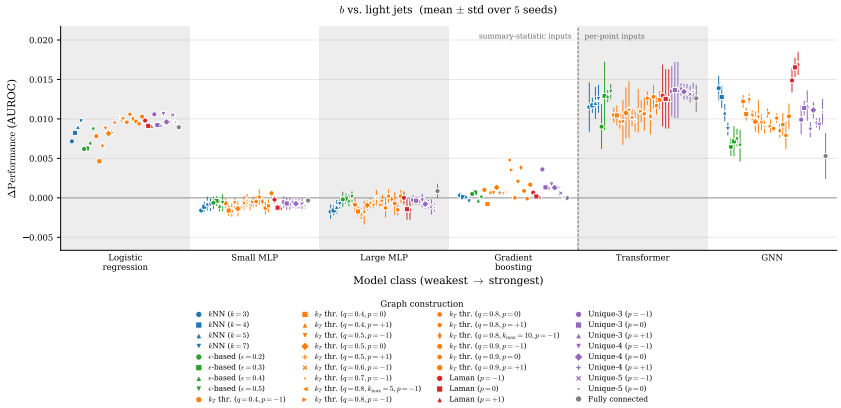


Figure 28: Incremental predictive value of curvature (Equation 7) for b vs light jets for 36 graph constructions. Each marker is ΔPerformance, the AUROC with curvature appended to the inputs minus the AUROC without it, for one graph construction (color and marker shape) and one model class, with model classes (with learning rate (10^{-3})) ordered from weakest to strongest along the horizontal axis. Markers show the mean of the per-seed deltas over 5 seeds and error bars their standard deviation across seeds. Error bars combine seed to seed and test set bootstrap standard deviations in quadrature ($\sqrt{(\sigma_{seed})^2 + \sigma_{boot}^2}$). For logistic regression and gradient boosting the seed component is zero, so their bars show bootstrap uncertainty alone.

Table 26: Test AUROC for b vs light jets for all 36 JetSet constructions of Table 1 (learning rate 10^{-3}). (a) Baselines without curvature features for the five models that never see the graph,so do not depend on the construction. (b) For each construction, the GNN baseline without curvature features and every model with both FRC and ORC node features computed on that construction. Rows are grouped by family, and * marks the default setting of each family; the sweep constructions come from a separate training run whose transformer baseline differs from (a) by 0.0001. Entries give the mean over 5 training seeds. Brackets give the 95% percentile bootstrap confidence interval of that mean, obtained by resampling the seeds with replacement 10,000 times, averaging each resample, and taking the 2.5th and 97.5th percentiles of the resampled means.

(a) Without curvature features: the five models that never see the graph

Logistic Regression	Small MLP	Large MLP	Gradient Boosting	Transformer
0.6589 [0.6589, 0.6589]	0.6888 [0.6886, 0.6891]	0.6902 [0.6899, 0.6905]	0.6916 [0.6916, 0.6916]	0.7119 [0.7091, 0.7149]

(b) Per construction: GNN without curvature features, and all models with them

	Without curvature	With curvature features					
Graph construction	GNN	Logistic Regression	Small MLP	Large MLP	Gradient Boosting	Transformer	GNN
Fully connected*	0.7202 [0.7180, 0.7215]	0.6678 [0.6678, 0.6678]	0.6884 [0.6883, 0.6885]	0.6911 [0.6908, 0.6916]	0.6916 [0.6916, 0.6916]	0.7245 [0.7220, 0.7263]	0.7255 [0.7248, 0.7262]
kNN ($k = 3$)*	0.7068 [0.7061, 0.7075]	0.6661 [0.6661, 0.6661]	0.6872 [0.6868, 0.6876]	0.6885 [0.6879, 0.6891]	0.6920 [0.6920, 0.6920]	0.7234 [0.7227, 0.7239]	0.7207 [0.7199, 0.7213]
kNN ($k = 4$)	0.7099 [0.7094, 0.7104]	0.6671 [0.6671, 0.6671]	0.6877 [0.6874, 0.6879]	0.6886 [0.6883, 0.6888]	0.6917 [0.6917, 0.6917]	0.7236 [0.7224, 0.7245]	0.7227 [0.7219, 0.7233]
kNN ($k = 5$)	0.7122 [0.7119, 0.7125]	0.6679 [0.6679, 0.6679]	0.6881 [0.6876, 0.6885]	0.6891 [0.6883, 0.6897]	0.6918 [0.6918, 0.6918]	0.7239 [0.7232, 0.7244]	0.7230 [0.7222, 0.7235]
kNN ($k = 7$)	0.7163 [0.7153, 0.7173]	0.6686 [0.6686, 0.6686]	0.6879 [0.6874, 0.6883]	0.6894 [0.6889, 0.6899]	0.6912 [0.6912, 0.6912]	0.7244 [0.7238, 0.7250]	0.7251 [0.7245, 0.7255]
ϵ-based ($\epsilon = 0.2$)	0.7145 [0.7138, 0.7153]	0.6651 [0.6651, 0.6651]	0.6882 [0.6876, 0.6887]	0.6900 [0.6894, 0.6907]	0.6921 [0.6921, 0.6921]	0.7209 [0.7185, 0.7238]	0.7210 [0.7203, 0.7216]
ϵ-based ($\epsilon = 0.3$)*	0.7154 [0.7140, 0.7168]	0.6652 [0.6652, 0.6652]	0.6884 [0.6881, 0.6887]	0.6901 [0.6894, 0.6906]	0.6923 [0.6923, 0.6923]	0.7248 [0.7238, 0.7260]	0.7226 [0.7220, 0.7230]
ϵ-based ($\epsilon = 0.4$)	0.7176 [0.7164, 0.7185]	0.6659 [0.6659, 0.6659]	0.6878 [0.6874, 0.6881]	0.6899 [0.6896, 0.6903]	0.6913 [0.6913, 0.6913]	0.7247 [0.7228, 0.7264]	0.7252 [0.7249, 0.7256]
ϵ-based ($\epsilon = 0.5$)	0.7187 [0.7168, 0.7202]	0.6676 [0.6676, 0.6676]	0.6881 [0.6874, 0.6888]	0.6903 [0.6896, 0.6910]	0.6917 [0.6917, 0.6917]	0.7253 [0.7245, 0.7261]	0.7253 [0.7246, 0.7260]
k_T threshold ($q = 0.4$, $p = -1$)	0.7076 [0.7073, 0.7080]	0.6667 [0.6667, 0.6667]	0.6881 [0.6878, 0.6884]	0.6894 [0.6889, 0.6899]	0.6927 [0.6927, 0.6927]	0.7223 [0.7209, 0.7233]	0.7198 [0.7194, 0.7203]
k_T threshold ($q = 0.4$, $p = 0$)	0.7077 [0.7070, 0.7082]	0.6635 [0.6635, 0.6635]	0.6872 [0.6869, 0.6874]	0.6885 [0.6882, 0.6888]	0.6909 [0.6909, 0.6909]	0.7223 [0.7215, 0.7233]	0.7184 [0.7181, 0.7186]
k_T threshold ($q = 0.4$, $p = +1$)	0.7050 [0.7044, 0.7058]	0.6656 [0.6656, 0.6656]	0.6872 [0.6868, 0.6876]	0.6882 [0.6878, 0.6887]	0.6924 [0.6924, 0.6924]	0.7216 [0.7208, 0.7222]	0.7176 [0.7172, 0.7182]
k_T threshold ($q = 0.5$, $p = -1$)*	0.7105 [0.7097, 0.7113]	0.6677 [0.6677, 0.6677]	0.6881 [0.6878, 0.6884]	0.6883 [0.6874, 0.6892]	0.6922 [0.6922, 0.6922]	0.7214 [0.7205, 0.7221]	0.7209 [0.7202, 0.7213]
k_T threshold ($q = 0.5$, $p = 0$)*	0.7113 [0.7102, 0.7123]	0.6671 [0.6671, 0.6671]	0.6874 [0.6868, 0.6881]	0.6893 [0.6891, 0.6896]	0.6930 [0.6930, 0.6930]	0.7226 [0.7221, 0.7232]	0.7210 [0.7205, 0.7215]
k_T threshold ($q = 0.5$, $p = +1$)*	0.7085 [0.7069, 0.7097]	0.6672 [0.6672, 0.6672]	0.6875 [0.6874, 0.6876]	0.6896 [0.6892, 0.6899]	0.6923 [0.6923, 0.6923]	0.7230 [0.7225, 0.7237]	0.7188 [0.7184, 0.7192]
k_T threshold ($q = 0.6$, $p = -1$)	0.7131 [0.7123, 0.7139]	0.6684 [0.6684, 0.6684]	0.6881 [0.6877, 0.6886]	0.6894 [0.6889, 0.6899]	0.6923 [0.6923, 0.6923]	0.7220 [0.7216, 0.7224]	0.7227 [0.7222, 0.7231]
k_T threshold ($q = 0.7$, $p = -1$)	0.7150 [0.7143, 0.7156]	0.6692 [0.6692, 0.6692]	0.6885 [0.6878, 0.6891]	0.6902 [0.6895, 0.6908]	0.6924 [0.6924, 0.6924]	0.7216 [0.7211, 0.7221]	0.7240 [0.7231, 0.7249]
k_T threshold ($q = 0.8$, $k_{\max} = 5$, $p = -1$)	0.7146 [0.7143, 0.7149]	0.6687 [0.6687, 0.6687]	0.6879 [0.6874, 0.6884]	0.6894 [0.6890, 0.6898]	0.6964 [0.6964, 0.6964]	0.7229 [0.7212, 0.7243]	0.7253 [0.7250, 0.7256]
k_T threshold ($q = 0.8$, $p = -1$)	0.7157 [0.7150, 0.7166]	0.6690 [0.6690, 0.6690]	0.6883 [0.6879, 0.6886]	0.6898 [0.6895, 0.6902]	0.6952 [0.6952, 0.6952]	0.7228 [0.7224, 0.7232]	0.7248 [0.7241, 0.7253]
k_T threshold ($q = 0.8$, $p = 0$)	0.7146 [0.7142, 0.7151]	0.6685 [0.6685, 0.6685]	0.6883 [0.6880, 0.6886]	0.6889 [0.6882, 0.6897]	0.6916 [0.6916, 0.6916]	0.7225 [0.7218, 0.7232]	0.7234 [0.7232, 0.7237]
k_T threshold ($q = 0.8$, $p = +1$)	0.7131 [0.7125, 0.7136]	0.6695 [0.6695, 0.6695]	0.6889 [0.6884, 0.6895]	0.6904 [0.6899, 0.6909]	0.6937 [0.6937, 0.6937]	0.7245 [0.7238, 0.7251]	0.7232 [0.7225, 0.7239]
k_T threshold ($q = 0.8$, $k_{\max} = 10$, $p = -1$)	0.7164 [0.7162, 0.7166]	0.6689 [0.6689, 0.6689]	0.6883 [0.6880, 0.6886]	0.6901 [0.6896, 0.6905]	0.6955 [0.6955, 0.6955]	0.7222 [0.7214, 0.7229]	0.7249 [0.7242, 0.7256]
k_T threshold ($q = 0.9$, $p = -1$)	0.7164 [0.7154, 0.7171]	0.6686 [0.6686, 0.6686]	0.6875 [0.6870, 0.6879]	0.6895 [0.6885, 0.6904]	0.6925 [0.6925, 0.6925]	0.7247 [0.7244, 0.7249]	0.7257 [0.7253, 0.7260]
k_T threshold ($q = 0.9$, $p = 0$)	0.7155 [0.7148, 0.7161]	0.6683 [0.6683, 0.6683]	0.6878 [0.6874, 0.6881]	0.6887 [0.6883, 0.6891]	0.6915 [0.6915, 0.6915]	0.7235 [0.7230, 0.7240]	0.7234 [0.7224, 0.7241]
k_T threshold ($q = 0.9$, $p = +1$)	0.7138 [0.7125, 0.7150]	0.6692 [0.6692, 0.6692]	0.6894 [0.6889, 0.6898]	0.6904 [0.6900, 0.6908]	0.6933 [0.6933, 0.6933]	0.7242 [0.7236, 0.7251]	0.7241 [0.7232, 0.7250]
Laman ($p = -1$)*	0.7069 [0.7057, 0.7080]	0.6687 [0.6687, 0.6687]	0.6886 [0.6883, 0.6888]	0.6902 [0.6898, 0.6908]	0.6923 [0.6923, 0.6923]	0.7248 [0.7243, 0.7254]	0.7217 [0.7211, 0.7225]
Laman ($p = 0$)*	0.7044 [0.7039, 0.7049]	0.6680 [0.6680, 0.6680]	0.6875 [0.6875, 0.6876]	0.6888 [0.6880, 0.6897]	0.6918 [0.6918, 0.6918]	0.7244 [0.7235, 0.7250]	0.7210 [0.7203, 0.7216]
Laman ($p = +1$)*	0.7041 [0.7031, 0.7050]	0.6680 [0.6680, 0.6680]	0.6875 [0.6875, 0.6876]	0.6888 [0.6880, 0.6897]	0.6918 [0.6918, 0.6918]	0.7244 [0.7235, 0.7250]	0.7211 [0.7205, 0.7216]
Unique-3 ($p = -1$)*	0.7132 [0.7124, 0.7141]	0.6695 [0.6695, 0.6695]	0.6882 [0.6877, 0.6887]	0.6899 [0.6894, 0.6905]	0.6953 [0.6953, 0.6953]	0.7253 [0.7244, 0.7265]	0.7231 [0.7222, 0.7241]
Unique-3 ($p = 0$)*	0.7108 [0.7099, 0.7119]	0.6681 [0.6681, 0.6681]	0.6881 [0.6877, 0.6884]	0.6899 [0.6895, 0.6903]	0.6930 [0.6930, 0.6930]	0.7255 [0.7243, 0.7267]	0.7221 [0.7218, 0.7226]
Unique-3 ($p = +1$)*	0.7105 [0.7097, 0.7115]	0.6681 [0.6681, 0.6681]	0.6881 [0.6877, 0.6883]	0.6899 [0.6894, 0.6903]	0.6930 [0.6930, 0.6930]	0.7255 [0.7243, 0.7267]	0.7222 [0.7213, 0.7228]
Unique-4 ($p = -1$)	0.7167 [0.7161, 0.7172]	0.6695 [0.6695, 0.6695]	0.6881 [0.6879, 0.6883]	0.6899 [0.6894, 0.6905]	0.6933 [0.6933, 0.6933]	0.7256 [0.7246, 0.7266]	0.7253 [0.7249, 0.7258]
Unique-4 ($p = 0$)	0.7137 [0.7131, 0.7142]	0.6685 [0.6685, 0.6685]	0.6880 [0.6878, 0.6883]	0.6894 [0.6889, 0.6899]	0.6929 [0.6929, 0.6929]	0.7253 [0.7244, 0.7260]	0.7248 [0.7245, 0.7251]
Unique-4 ($p = +1$)	0.7147 [0.7142, 0.7152]	0.6685 [0.6685, 0.6685]	0.6880 [0.6878, 0.6883]	0.6894 [0.6889, 0.6899]	0.6929 [0.6929, 0.6929]	0.7253 [0.7244, 0.7260]	0.7242 [0.7234, 0.7248]
Unique-5 ($p = -1$)	0.7167 [0.7163, 0.7171]	0.6694 [0.6694, 0.6694]	0.6880 [0.6878, 0.6882]	0.6891 [0.6886, 0.6896]	0.6922 [0.6922, 0.6922]	0.7250 [0.7243, 0.7256]	0.7261 [0.7259, 0.7263]
Unique-5 ($p = 0$)	0.7149 [0.7141, 0.7158]	0.6685 [0.6685, 0.6685]	0.6878 [0.6874, 0.6882]	0.6890 [0.6888, 0.6892]	0.6917 [0.6917, 0.6917]	0.7253 [0.7246, 0.7260]	0.7258 [0.7253, 0.7264]

Table 27: Change in test AUROC from adding Ricci curvature features for b vs light jets (learning rate 10^{-3}; Equation 7) for all 36 JetSet constructions of Table 1, computed per seed between runs that share architecture, hyperparameters, data, and seed. Rows are grouped by family, and * marks the default setting of each family. Entries give the mean over 5 training seeds; brackets give the 95% percentile bootstrap confidence interval of that mean, obtained by resampling the seeds with replacement 10,000 times, averaging each resample, and taking the 2.5th and 97.5th percentiles of the resampled means. Seeds are resampled jointly for both settings, so the pairing is preserved. Bold marks entries whose 95% interval excludes zero, a consistent gain (positive) or loss (negative) from adding curvature.

Graph construction	Logistic Regression	Small MLP	Large MLP	Gradient Boosting	Transformer	GNN
Fully connected*	**+0.0089** [+0.0089, +0.0089]	**−0.0003** [−0.0007, −0.0000]	**+0.0009** [+0.0003, +0.0017]	+0.0000 [+0.0000, +0.0000]	**+0.0126** [+0.0111, +0.0139]	**+0.0053** [+0.0035, +0.0079]
kNN ($k = 3$)*	**+0.0072** [+0.0072, +0.0072]	**−0.0016** [−0.0019, −0.0013]	**−0.0018** [−0.0023, −0.0009]	**+0.0003** [+0.0003, +0.0003]	**+0.0115** [+0.0088, +0.0136]	**+0.0139** [+0.0126, +0.0151]
kNN ($k = 4$)	**+0.0082** [+0.0082, +0.0082]	**−0.0011** [−0.0016, −0.0006]	**−0.0016** [−0.0021, −0.0012]	**+0.0001** [+0.0001, +0.0001]	**+0.0117** [+0.0109, +0.0124]	**+0.0128** [+0.0116, +0.0138]
kNN ($k = 5$)	**+0.0090** [+0.0090, +0.0090]	**−0.0007** [−0.0015, −0.0001]	**−0.0012** [−0.0018, −0.0004]	**+0.0002** [+0.0002, +0.0002]	**+0.0120** [+0.0102, +0.0135]	**+0.0108** [+0.0098, +0.0115]
kNN ($k = 7$)	**+0.0097** [+0.0097, +0.0097]	**−0.0009** [−0.0016, −0.0003]	**−0.0008** [−0.0012, −0.0004]	**−0.0004** [−0.0004, −0.0004]	**+0.0125** [+0.0113, +0.0141]	**+0.0087** [+0.0080, +0.0093]
ϵ-based ($\epsilon = 0.2$)	**+0.0062** [+0.0062, +0.0062]	**−0.0006** [−0.0012, −0.0000]	−0.0002 [−0.0009, +0.0006]	**+0.0005** [+0.0005, +0.0005]	**+0.0090** [+0.0068, +0.0112]	**+0.0065** [+0.0056, +0.0073]
ϵ-based ($\epsilon = 0.3$)*	**+0.0063** [+0.0063, +0.0063]	−0.0004 [−0.0008, +0.0001]	−0.0001 [−0.0006, +0.0004]	**+0.0007** [+0.0007, +0.0007]	**+0.0129** [+0.0091, +0.0160]	**+0.0072** [+0.0057, +0.0086]
ϵ-based ($\epsilon = 0.4$)	**+0.0070** [+0.0070, +0.0070]	**−0.0010** [−0.0017, −0.0005]	−0.0003 [−0.0009, +0.0003]	**−0.0003** [−0.0003, −0.0003]	**+0.0129** [+0.0123, +0.0134]	**+0.0076** [+0.0069, +0.0086]
ϵ-based ($\epsilon = 0.5$)	**+0.0087** [+0.0087, +0.0087]	−0.0006 [−0.0015, +0.0002]	+0.0001 [−0.0005, +0.0006]	**+0.0000** [+0.0000, +0.0000]	**+0.0134** [+0.0127, +0.0143]	**+0.0066** [+0.0052, +0.0084]
k_T threshold ($q = 0.4$, $p = -1$)	**+0.0078** [+0.0078, +0.0078]	**−0.0007** [−0.0012, −0.0002]	**−0.0009** [−0.0015, −0.0002]	**+0.0010** [+0.0010, +0.0010]	**+0.0105** [+0.0098, +0.0113]	**+0.0122** [+0.0116, +0.0129]
k_T threshold ($q = 0.4$, $p = 0$)	**+0.0046** [+0.0046, +0.0046]	**−0.0016** [−0.0022, −0.0012]	**−0.0017** [−0.0022, −0.0012]	**−0.0008** [−0.0008, −0.0008]	**+0.0104** [+0.0096, +0.0116]	**+0.0106** [+0.0099, +0.0113]
k_T threshold ($q = 0.4$, $p = +1$)	**+0.0067** [+0.0067, +0.0067]	**−0.0016** [−0.0023, −0.0010]	**−0.0020** [−0.0025, −0.0015]	**+0.0008** [+0.0008, +0.0008]	**+0.0098** [+0.0089, +0.0104]	**+0.0126** [+0.0121, +0.0131]
k_T threshold ($q = 0.5$, $p = -1$)*	**+0.0088** [+0.0088, +0.0088]	**−0.0007** [−0.0012, −0.0003]	**−0.0019** [−0.0031, −0.0008]	**+0.0006** [+0.0006, +0.0006]	**+0.0096** [+0.0071, +0.0116]	**+0.0104** [+0.0099, +0.0110]
k_T threshold ($q = 0.5$, $p = 0$)*	**+0.0082** [+0.0082, +0.0082]	**−0.0014** [−0.0021, −0.0006]	**−0.0009** [−0.0013, −0.0004]	**+0.0013** [+0.0013, +0.0013]	**+0.0108** [+0.0083, +0.0133]	**+0.0097** [+0.0086, +0.0107]
k_T threshold ($q = 0.5$, $p = +1$)*	**+0.0083** [+0.0083, +0.0083]	**−0.0013** [−0.0015, −0.0010]	**−0.0007** [−0.0011, −0.0001]	**+0.0006** [+0.0006, +0.0006]	**+0.0112** [+0.0083, +0.0139]	**+0.0103** [+0.0087, +0.0122]
k_T threshold ($q = 0.6$, $p = -1$)	**+0.0095** [+0.0095, +0.0095]	−0.0006 [−0.0013, +0.0000]	**−0.0008** [−0.0013, −0.0006]	**+0.0006** [+0.0006, +0.0006]	**+0.0102** [+0.0094, +0.0111]	**+0.0096** [+0.0087, +0.0103]
k_T threshold ($q = 0.7$, $p = -1$)	**+0.0103** [+0.0103, +0.0103]	−0.0003 [−0.0009, +0.0003]	−0.0000 [−0.0008, +0.0008]	**+0.0007** [+0.0007, +0.0007]	**+0.0098** [+0.0087, +0.0108]	**+0.0090** [+0.0084, +0.0096]
k_T threshold ($q = 0.8$, $k_{\max} = 5$, $p = -1$)	**+0.0098** [+0.0098, +0.0098]	**−0.0009** [−0.0014, −0.0003]	**−0.0008** [−0.0011, −0.0006]	**+0.0048** [+0.0048, +0.0048]	**+0.0110** [+0.0091, +0.0132]	**+0.0107** [+0.0102, +0.0111]
k_T threshold ($q = 0.8$, $p = -1$)	**+0.0101** [+0.0101, +0.0101]	**−0.0005** [−0.0011, −0.0000]	−0.0004 [−0.0010, +0.0002]	**+0.0035** [+0.0035, +0.0035]	**+0.0110** [+0.0099, +0.0121]	**+0.0091** [+0.0084, +0.0097]
k_T threshold ($q = 0.8$, $p = 0$)	**+0.0096** [+0.0096, +0.0096]	**−0.0005** [−0.0011, −0.0000]	**−0.0013** [−0.0022, −0.0004]	**−0.0000** [−0.0000, −0.0000]	**+0.0106** [+0.0090, +0.0123]	**+0.0088** [+0.0085, +0.0091]
k_T threshold ($q = 0.8$, $p = +1$)	**+0.0106** [+0.0106, +0.0106]	+0.0001 [−0.0007, +0.0009]	+0.0002 [−0.0005, +0.0008]	**+0.0021** [+0.0021, +0.0021]	**+0.0126** [+0.0121, +0.0134]	**+0.0101** [+0.0096, +0.0105]
k_T threshold ($q = 0.8$, $k_{\max} = 10$, $p = -1$)	**+0.0100** [+0.0100, +0.0100]	**−0.0005** [−0.0010, −0.0000]	−0.0001 [−0.0005, +0.0002]	**+0.0038** [+0.0038, +0.0038]	**+0.0103** [+0.0087, +0.0122]	**+0.0085** [+0.0077, +0.0091]
k_T threshold ($q = 0.9$, $p = -1$)	**+0.0097** [+0.0097, +0.0097]	**−0.0013** [−0.0019, −0.0007]	−0.0007 [−0.0019, +0.0003]	**+0.0009** [+0.0009, +0.0009]	**+0.0128** [+0.0116, +0.0140]	**+0.0093** [+0.0081, +0.0104]
k_T threshold ($q = 0.9$, $p = 0$)	**+0.0094** [+0.0094, +0.0094]	**−0.0010** [−0.0017, −0.0004]	**−0.0015** [−0.0021, −0.0010]	**−0.0001** [−0.0001, −0.0001]	**+0.0117** [+0.0108, +0.0126]	**+0.0079** [+0.0063, +0.0092]
k_T threshold ($q = 0.9$, $p = +1$)	**+0.0103** [+0.0103, +0.0103]	**+0.0006** [+0.0002, +0.0009]	+0.0002 [−0.0004, +0.0009]	**+0.0017** [+0.0017, +0.0017]	**+0.0124** [+0.0116, +0.0134]	**+0.0103** [+0.0091, +0.0116]
Laman ($p = -1$)*	**+0.0098** [+0.0098, +0.0098]	−0.0002 [−0.0006, +0.0001]	+0.0000 [−0.0005, +0.0007]	**+0.0007** [+0.0007, +0.0007]	**+0.0130** [+0.0098, +0.0157]	**+0.0149** [+0.0136, +0.0160]
Laman ($p = 0$)*	**+0.0091** [+0.0091, +0.0091]	**−0.0012** [−0.0016, −0.0010]	**−0.0014** [−0.0024, −0.0004]	**+0.0002** [+0.0002, +0.0002]	**+0.0125** [+0.0099, +0.0155]	**+0.0165** [+0.0157, +0.0175]
Laman ($p = +1$)*	**+0.0091** [+0.0091, +0.0091]	**−0.0012** [−0.0016, −0.0010]	**−0.0014** [−0.0024, −0.0004]	**+0.0002** [+0.0002, +0.0002]	**+0.0125** [+0.0097, +0.0155]	**+0.0170** [+0.0160, +0.0183]
Unique-3 ($p = -1$)*	**+0.0106** [+0.0106, +0.0106]	−0.0005 [−0.0011, +0.0000]	−0.0003 [−0.0007, +0.0002]	**+0.0036** [+0.0036, +0.0036]	**+0.0135** [+0.0108, +0.0158]	**+0.0099** [+0.0082, +0.0112]
Unique-3 ($p = 0$)*	**+0.0092** [+0.0092, +0.0092]	**−0.0007** [−0.0014, −0.0002]	−0.0004 [−0.0008, +0.0000]	**+0.0013** [+0.0013, +0.0013]	**+0.0137** [+0.0110, +0.0165]	**+0.0114** [+0.0106, +0.0120]
Unique-3 ($p = +1$)*	**+0.0092** [+0.0092, +0.0092]	**−0.0007** [−0.0014, −0.0002]	−0.0004 [−0.0008, +0.0000]	**+0.0013** [+0.0013, +0.0013]	**+0.0137** [+0.0110, +0.0165]	**+0.0117** [+0.0102, +0.0132]
Unique-4 ($p = -1$)	**+0.0106** [+0.0106, +0.0106]	**−0.0007** [−0.0012, −0.0004]	−0.0003 [−0.0010, +0.0003]	**+0.0017** [+0.0017, +0.0017]	**+0.0137** [+0.0132, +0.0143]	**+0.0086** [+0.0080, +0.0095]
Unique-4 ($p = 0$)	**+0.0096** [+0.0096, +0.0096]	**−0.0007** [−0.0012, −0.0003]	**−0.0008** [−0.0012, −0.0004]	**+0.0013** [+0.0013, +0.0013]	**+0.0135** [+0.0126, +0.0141]	**+0.0111** [+0.0103, +0.0119]
Unique-4 ($p = +1$)	**+0.0096** [+0.0096, +0.0096]	**−0.0007** [−0.0012, −0.0003]	**−0.0008** [−0.0012, −0.0004]	**+0.0013** [+0.0013, +0.0013]	**+0.0135** [+0.0126, +0.0141]	**+0.0095** [+0.0091, +0.0100]
Unique-5 ($p = -1$)	**+0.0105** [+0.0105, +0.0105]	**−0.0007** [−0.0014, −0.0003]	**−0.0012** [−0.0019, −0.0004]	**+0.0006** [+0.0006, +0.0006]	**+0.0131** [+0.0122, +0.0138]	**+0.0094** [+0.0088, +0.0100]
Unique-5 ($p = 0$)	**+0.0096** [+0.0096, +0.0096]	**−0.0010** [−0.0013, −0.0007]	**−0.0012** [−0.0016, −0.0008]	**+0.0001** [+0.0001, +0.0001]	**+0.0135** [+0.0126, +0.0143]	**+0.0109** [+0.0098, +0.0123]

G.3 JetSet: c vs light

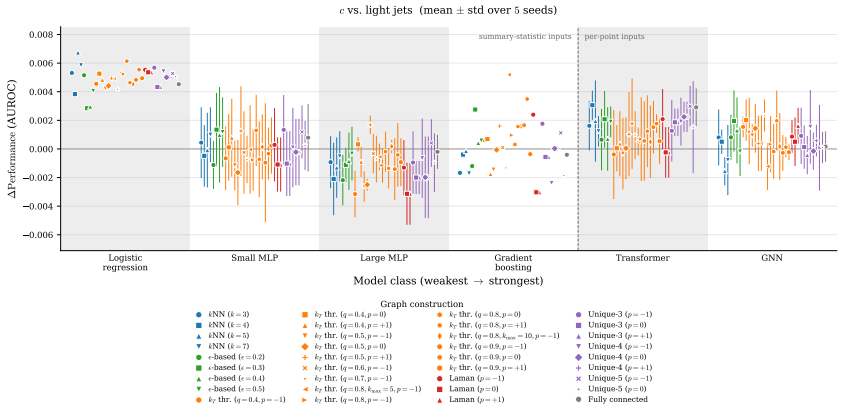


Figure 29: Incremental predictive value of curvature (Equation 7) for c vs light jets for 36 graph constructions. Each marker is ΔPerformance, the AUROC with curvature appended to the inputs minus the AUROC without it, for one graph construction (color and marker shape) and one model class, with model classes (with learning rate (10^{-3})) ordered from weakest to strongest along the horizontal axis. Markers show the mean of the per-seed deltas over 5 seeds and error bars their standard deviation across seeds. Error bars combine seed to seed and test set bootstrap standard deviations in quadrature ($\sqrt{(\sigma_{seed})^2 + \sigma_{boot}^2}$). For logistic regression and gradient boosting the seed component is zero, so their bars show bootstrap uncertainty alone.

Table 28: Test AUROC for c vs light jets for all 36 JetSet constructions of Table 1 (learning rate 10^{-3}). (a) Baselines without curvature features for the five models that never see the graph, which therefore do not depend on the construction. (b) For each construction, the GNN baseline without curvature features and every model with FRC and ORC features computed on that construction. Rows are grouped by family, and * marks the default setting of each family; the sweep constructions come from a separate training run whose transformer baseline differs from (a) by 0.0001. Entries give the mean over 5 training seeds; brackets give the 95% percentile bootstrap confidence interval of that mean, obtained by resampling the seeds with replacement 10,000 times, averaging each resample, and taking the 2.5th and 97.5th percentiles of the resampled means.

(a) Without curvature features: the five models that never see the graph

Logistic Regression	Small MLP	Large MLP	Gradient Boosting	Transformer
0.6003 [0.6003, 0.6003]	0.6184 [0.6159, 0.6202]	0.6192 [0.6187, 0.6197]	0.6117 [0.6117, 0.6117]	0.6175 [0.6159, 0.6192]

(b) Per construction: GNN without curvature features, and all models with them

	Without curvature	With curvature features					
Graph construction	GNN	Logistic Regression	Small MLP	Large MLP	Gradient Boosting	Transformer	GNN
Fully connected*	0.6195 [0.6184, 0.6205]	0.6049 [0.6049, 0.6049]	0.6192 [0.6182, 0.6201]	0.6190 [0.6181, 0.6200]	0.6113 [0.6113, 0.6113]	0.6204 [0.6188, 0.6222]	0.6196 [0.6180, 0.6212]
kNN ($k = 3$)*	0.6171 [0.6165, 0.6182]	0.6057 [0.6057, 0.6057]	0.6188 [0.6182, 0.6195]	0.6182 [0.6168, 0.6196]	0.6100 [0.6100, 0.6100]	0.6191 [0.6187, 0.6196]	0.6179 [0.6161, 0.6193]
kNN ($k = 4$)	0.6185 [0.6178, 0.6194]	0.6042 [0.6042, 0.6042]	0.6179 [0.6172, 0.6186]	0.6171 [0.6155, 0.6191]	0.6113 [0.6113, 0.6113]	0.6206 [0.6194, 0.6222]	0.6190 [0.6179, 0.6202]
kNN ($k = 5$)	0.6195 [0.6188, 0.6203]	0.6071 [0.6071, 0.6071]	0.6183 [0.6178, 0.6188]	0.6179 [0.6165, 0.6194]	0.6115 [0.6115, 0.6115]	0.6195 [0.6184, 0.6207]	0.6180 [0.6171, 0.6187]
kNN ($k = 7$)	0.6197 [0.6191, 0.6203]	0.6062 [0.6062, 0.6062]	0.6194 [0.6180, 0.6205]	0.6186 [0.6174, 0.6198]	0.6099 [0.6099, 0.6099]	0.6188 [0.6170, 0.6206]	0.6189 [0.6166, 0.6209]
ϵ-based ($\epsilon = 0.2$)	0.6146 [0.6137, 0.6161]	0.6055 [0.6055, 0.6055]	0.6173 [0.6161, 0.6185]	0.6170 [0.6153, 0.6185]	0.6105 [0.6105, 0.6105]	0.6182 [0.6168, 0.6194]	0.6154 [0.6137, 0.6168]
ϵ-based ($\epsilon = 0.3$)*	0.6170 [0.6157, 0.6184]	0.6032 [0.6032, 0.6032]	0.6198 [0.6190, 0.6206]	0.6180 [0.6168, 0.6193]	0.6144 [0.6144, 0.6144]	0.6196 [0.6185, 0.6207]	0.6190 [0.6183, 0.6196]
ϵ-based ($\epsilon = 0.4$)	0.6181 [0.6163, 0.6195]	0.6033 [0.6033, 0.6033]	0.6194 [0.6187, 0.6200]	0.6183 [0.6166, 0.6197]	0.6121 [0.6121, 0.6121]	0.6183 [0.6172, 0.6194]	0.6194 [0.6187, 0.6199]
ϵ-based ($\epsilon = 0.5$)	0.6197 [0.6184, 0.6210]	0.6044 [0.6044, 0.6044]	0.6196 [0.6187, 0.6205]	0.6188 [0.6174, 0.6201]	0.6122 [0.6122, 0.6122]	0.6195 [0.6179, 0.6212]	0.6195 [0.6190, 0.6201]
k_T threshold ($q = 0.4$, $p = -1$)	0.6158 [0.6151, 0.6167]	0.6049 [0.6049, 0.6049]	0.6177 [0.6167, 0.6188]	0.6160 [0.6144, 0.6173]	0.6123 [0.6123, 0.6123]	0.6172 [0.6156, 0.6187]	0.6174 [0.6166, 0.6183]
k_T threshold ($q = 0.4$, $p = 0$)	0.6155 [0.6151, 0.6161]	0.6056 [0.6056, 0.6056]	0.6185 [0.6178, 0.6192]	0.6195 [0.6189, 0.6200]	0.6123 [0.6123, 0.6123]	0.6176 [0.6166, 0.6186]	0.6176 [0.6167, 0.6188]
k_T threshold ($q = 0.4$, $p = +1$)	0.6164 [0.6152, 0.6176]	0.6052 [0.6052, 0.6052]	0.6192 [0.6181, 0.6205]	0.6183 [0.6176, 0.6188]	0.6099 [0.6099, 0.6099]	0.6179 [0.6155, 0.6202]	0.6176 [0.6161, 0.6192]
k_T threshold ($q = 0.5$, $p = -1$)*	0.6174 [0.6169, 0.6179]	0.6046 [0.6046, 0.6046]	0.6173 [0.6162, 0.6184]	0.6165 [0.6152, 0.6178]	0.6102 [0.6102, 0.6102]	0.6172 [0.6164, 0.6180]	0.6183 [0.6175, 0.6192]
k_T threshold ($q = 0.5$, $p = 0$)*	0.6168 [0.6155, 0.6181]	0.6048 [0.6048, 0.6048]	0.6168 [0.6158, 0.6179]	0.6167 [0.6160, 0.6176]	0.6116 [0.6116, 0.6116]	0.6176 [0.6168, 0.6183]	0.6182 [0.6172, 0.6195]
k_T threshold ($q = 0.5$, $p = +1$)*	0.6169 [0.6158, 0.6182]	0.6053 [0.6053, 0.6053]	0.6197 [0.6188, 0.6205]	0.6208 [0.6201, 0.6218]	0.6133 [0.6133, 0.6133]	0.6186 [0.6169, 0.6206]	0.6173 [0.6154, 0.6188]
k_T threshold ($q = 0.6$, $p = -1$)	0.6186 [0.6178, 0.6193]	0.6053 [0.6053, 0.6053]	0.6182 [0.6172, 0.6195]	0.6188 [0.6178, 0.6195]	0.6118 [0.6118, 0.6118]	0.6193 [0.6186, 0.6199]	0.6190 [0.6168, 0.6210]
k_T threshold ($q = 0.7$, $p = -1$)	0.6171 [0.6158, 0.6181]	0.6045 [0.6045, 0.6045]	0.6177 [0.6164, 0.6191]	0.6185 [0.6172, 0.6200]	0.6103 [0.6103, 0.6103]	0.6185 [0.6177, 0.6196]	0.6194 [0.6187, 0.6205]
k_T threshold ($q = 0.8$, $k_{\max} = 5$, $p = -1$)	0.6197 [0.6184, 0.6209]	0.6057 [0.6057, 0.6057]	0.6184 [0.6180, 0.6188]	0.6186 [0.6177, 0.6194]	0.6169 [0.6169, 0.6169]	0.6181 [0.6164, 0.6197]	0.6185 [0.6166, 0.6203]
k_T threshold ($q = 0.8$, $p = -1$)	0.6184 [0.6169, 0.6198]	0.6056 [0.6056, 0.6056]	0.6176 [0.6170, 0.6182]	0.6180 [0.6176, 0.6186]	0.6126 [0.6126, 0.6126]	0.6182 [0.6167, 0.6197]	0.6187 [0.6181, 0.6193]
k_T threshold ($q = 0.8$, $p = 0$)	0.6185 [0.6174, 0.6201]	0.6065 [0.6065, 0.6065]	0.6197 [0.6186, 0.6208]	0.6190 [0.6178, 0.6200]	0.6120 [0.6120, 0.6120]	0.6181 [0.6169, 0.6193]	0.6183 [0.6162, 0.6201]
k_T threshold ($q = 0.8$, $p = +1$)	0.6181 [0.6171, 0.6192]	0.6050 [0.6050, 0.6050]	0.6177 [0.6168, 0.6187]	0.6178 [0.6166, 0.6192]	0.6132 [0.6132, 0.6132]	0.6188 [0.6176, 0.6202]	0.6201 [0.6183, 0.6217]
k_T threshold ($q = 0.8$, $k_{\max} = 10$, $p = -1$)	0.6192 [0.6181, 0.6202]	0.6049 [0.6049, 0.6049]	0.6185 [0.6171, 0.6200]	0.6193 [0.6188, 0.6200]	0.6132 [0.6132, 0.6132]	0.6181 [0.6162, 0.6201]	0.6193 [0.6181, 0.6204]
k_T threshold ($q = 0.9$, $p = -1$)	0.6200 [0.6190, 0.6209]	0.6052 [0.6052, 0.6052]	0.6174 [0.6157, 0.6188]	0.6177 [0.6164, 0.6188]	0.6133 [0.6133, 0.6133]	0.6191 [0.6182, 0.6202]	0.6197 [0.6192, 0.6205]
k_T threshold ($q = 0.9$, $p = 0$)	0.6190 [0.6184, 0.6199]	0.6059 [0.6059, 0.6059]	0.6181 [0.6170, 0.6189]	0.6187 [0.6174, 0.6202]	0.6151 [0.6151, 0.6151]	0.6193 [0.6176, 0.6213]	0.6191 [0.6188, 0.6194]
k_T threshold ($q = 0.9$, $p = +1$)	0.6199 [0.6193, 0.6207]	0.6053 [0.6053, 0.6053]	0.6186 [0.6178, 0.6195]	0.6182 [0.6170, 0.6192]	0.6113 [0.6113, 0.6113]	0.6181 [0.6167, 0.6195]	0.6197 [0.6190, 0.6204]
Laman ($p = -1$)*	0.6174 [0.6166, 0.6182]	0.6059 [0.6059, 0.6059]	0.6187 [0.6174, 0.6200]	0.6179 [0.6161, 0.6191]	0.6141 [0.6141, 0.6141]	0.6196 [0.6183, 0.6213]	0.6183 [0.6175, 0.6191]
Laman ($p = 0$)*	0.6167 [0.6156, 0.6182]	0.6057 [0.6057, 0.6057]	0.6173 [0.6162, 0.6187]	0.6160 [0.6141, 0.6179]	0.6086 [0.6086, 0.6086]	0.6173 [0.6158, 0.6182]	0.6172 [0.6158, 0.6183]
Laman ($p = +1$)*	0.6162 [0.6151, 0.6174]	0.6057 [0.6057, 0.6057]	0.6173 [0.6162, 0.6187]	0.6160 [0.6141, 0.6179]	0.6086 [0.6086, 0.6086]	0.6173 [0.6158, 0.6182]	0.6178 [0.6170, 0.6185]
Unique-3 ($p = -1$)*	0.6172 [0.6164, 0.6180]	0.6060 [0.6060, 0.6060]	0.6197 [0.6188, 0.6207]	0.6182 [0.6164, 0.6194]	0.6134 [0.6134, 0.6134]	0.6188 [0.6177, 0.6202]	0.6181 [0.6161, 0.6196]
Unique-3 ($p = 0$)*	0.6181 [0.6171, 0.6194]	0.6047 [0.6047, 0.6047]	0.6174 [0.6168, 0.6180]	0.6172 [0.6164, 0.6179]	0.6111 [0.6111, 0.6111]	0.6194 [0.6180, 0.6208]	0.6183 [0.6176, 0.6190]
Unique-3 ($p = +1$)*	0.6180 [0.6173, 0.6189]	0.6047 [0.6047, 0.6047]	0.6174 [0.6168, 0.6181]	0.6172 [0.6164, 0.6179]	0.6111 [0.6111, 0.6111]	0.6194 [0.6180, 0.6208]	0.6176 [0.6163, 0.6186]
Unique-4 ($p = -1$)	0.6188 [0.6170, 0.6206]	0.6058 [0.6058, 0.6058]	0.6185 [0.6176, 0.6194]	0.6185 [0.6161, 0.6202]	0.6093 [0.6093, 0.6093]	0.6195 [0.6180, 0.6213]	0.6204 [0.6194, 0.6213]
Unique-4 ($p = 0$)	0.6193 [0.6182, 0.6205]	0.6054 [0.6054, 0.6054]	0.6182 [0.6172, 0.6193]	0.6172 [0.6147, 0.6188]	0.6117 [0.6117, 0.6117]	0.6198 [0.6184, 0.6216]	0.6192 [0.6172, 0.6210]
Unique-4 ($p = +1$)	0.6196 [0.6185, 0.6205]	0.6054 [0.6054, 0.6054]	0.6182 [0.6172, 0.6193]	0.6172 [0.6147, 0.6188]	0.6117 [0.6117, 0.6117]	0.6198 [0.6184, 0.6216]	0.6201 [0.6190, 0.6213]
Unique-5 ($p = -1$)	0.6196 [0.6189, 0.6202]	0.6056 [0.6056, 0.6056]	0.6196 [0.6187, 0.6205]	0.6196 [0.6182, 0.6205]	0.6128 [0.6128, 0.6128]	0.6205 [0.6188, 0.6223]	0.6197 [0.6173, 0.6222]
Unique-5 ($p = 0$)	0.6196 [0.6182, 0.6211]	0.6054 [0.6054, 0.6054]	0.6187 [0.6173, 0.6198]	0.6182 [0.6168, 0.6196]	0.6098 [0.6098, 0.6098]	0.6191 [0.6175, 0.6206]	0.6197 [0.6191, 0.6205]

Table 29: Change in test AUROC from adding Ricci curvature features for c vs light jets (learning rate 10^{-3}; Equation 7) for all 36 JetSet constructions of Table 1, computed per seed between runs that share architecture, hyperparameters, data, and seed. Rows are grouped by family, and * marks the default setting of each family. Entries give the mean over 5 training seeds; brackets give the 95% percentile bootstrap confidence interval of that mean, obtained by resampling the seeds with replacement 10,000 times, averaging each resample, and taking the 2.5th and 97.5th percentiles of the resampled means. Seeds are resampled jointly for both settings, so the pairing is preserved. Bold marks entries whose 95% interval excludes zero, a consistent gain (positive) or loss (negative) from adding curvature.

Graph construction	Logistic Regression	Small MLP	Large MLP	Gradient Boosting	Transformer	GNN
Fully connected*	**+0.0045** [+0.0045, +0.0045]	+0.0008 [−0.0006, +0.0029]	−0.0002 [−0.0011, +0.0007]	**−0.0004** [−0.0004, −0.0004]	**+0.0029** [+0.0019, +0.0040]	+0.0002 [−0.0008, +0.0007]
kNN (k = 3)*	**+0.0053** [+0.0053, +0.0053]	+0.0004 [−0.0013, +0.0024]	−0.0009 [−0.0022, +0.0005]	**−0.0017** [−0.0017, −0.0017]	**+0.0016** [+0.0003, +0.0030]	+0.0008 [−0.0008, +0.0023]
kNN (k = 4)	**+0.0038** [+0.0038, +0.0038]	−0.0005 [−0.0021, +0.0013]	−0.0021 [−0.0038, +0.0002]	**−0.0004** [−0.0004, −0.0004]	**+0.0031** [+0.0022, +0.0039]	+0.0005 [−0.0000, +0.0012]
kNN (k = 5)	**+0.0067** [+0.0067, +0.0067]	−0.0001 [−0.0020, +0.0021]	−0.0013 [−0.0027, +0.0006]	**−0.0002** [−0.0002, −0.0002]	+0.0019 [−0.0000, +0.0043]	**−0.0015** [−0.0023, −0.0007]
kNN (k = 7)	**+0.0058** [+0.0058, +0.0058]	+0.0010 [−0.0006, +0.0024]	−0.0005 [−0.0019, +0.0008]	**−0.0017** [−0.0017, −0.0017]	**+0.0012** [+0.0003, +0.0020]	−0.0008 [−0.0029, +0.0008]
ϵ-based (ϵ = 0.2)	**+0.0052** [+0.0052, +0.0052]	−0.0011 [−0.0022, +0.0003]	**−0.0022** [−0.0037, −0.0010]	**−0.0012** [−0.0012, −0.0012]	+0.0006 [−0.0010, +0.0025]	+0.0008 [−0.0004, +0.0020]
ϵ-based (ϵ = 0.3)*	**+0.0028** [+0.0028, +0.0028]	+0.0013 [−0.0003, +0.0035]	**−0.0011** [−0.0021, −0.0002]	**+0.0028** [+0.0028, +0.0028]	**+0.0021** [+0.0005, +0.0036]	**+0.0020** [+0.0003, +0.0035]
ϵ-based (ϵ = 0.4)	**+0.0030** [+0.0030, +0.0030]	+0.0010 [−0.0013, +0.0037]	−0.0009 [−0.0026, +0.0006]	**+0.0004** [+0.0004, +0.0004]	+0.0007 [−0.0010, +0.0024]	+0.0013 [−0.0007, +0.0028]
ϵ-based (ϵ = 0.5)	**+0.0041** [+0.0041, +0.0041]	+0.0012 [−0.0003, +0.0033]	−0.0003 [−0.0017, +0.0013]	**+0.0006** [+0.0006, +0.0006]	**+0.0019** [+0.0011, +0.0028]	−0.0002 [−0.0015, +0.0011]
k_T threshold (q = 0.4, p = −1)	**+0.0046** [+0.0046, +0.0046]	−0.0007 [−0.0023, +0.0012]	**−0.0032** [−0.0045, −0.0021]	**+0.0006** [+0.0006, +0.0006]	−0.0004 [−0.0027, +0.0021]	**+0.0015** [+0.0010, +0.0021]
k_T threshold (q = 0.4, p = 0)	**+0.0053** [+0.0053, +0.0053]	+0.0001 [−0.0013, +0.0022]	+0.0003 [−0.0001, +0.0007]	**+0.0007** [+0.0007, +0.0007]	+0.0000 [−0.0012, +0.0013]	**+0.0020** [+0.0008, +0.0032]
k_T threshold (q = 0.4, p = +1)	**+0.0048** [+0.0048, +0.0048]	+0.0008 [−0.0008, +0.0032]	**−0.0009** [−0.0018, −0.0001]	**−0.0018** [−0.0018, −0.0018]	+0.0004 [−0.0014, +0.0021]	+0.0012 [−0.0009, +0.0026]
k_T threshold (q = 0.5, p = −1)*	**+0.0042** [+0.0042, +0.0042]	−0.0011 [−0.0020, +0.0004]	**−0.0026** [−0.0035, −0.0018]	**−0.0014** [−0.0014, −0.0014]	−0.0003 [−0.0024, +0.0015]	**+0.0009** [+0.0004, +0.0016]
k_T threshold (q = 0.5, p = 0)*	**+0.0044** [+0.0044, +0.0044]	−0.0016 [−0.0030, +0.0004]	**−0.0025** [−0.0028, −0.0021]	**−0.0001** [−0.0001, −0.0001]	+0.0000 [−0.0016, +0.0017]	**+0.0014** [+0.0001, +0.0028]
k_T threshold (q = 0.5, p = +1)*	**+0.0049** [+0.0049, +0.0049]	+0.0013 [−0.0009, +0.0040]	**+0.0017** [+0.0012, +0.0022]	**+0.0016** [+0.0016, +0.0016]	+0.0010 [−0.0013, +0.0033]	+0.0004 [−0.0011, +0.0019]
k_T threshold (q = 0.6, p = −1)	**+0.0049** [+0.0049, +0.0049]	−0.0002 [−0.0021, +0.0023]	−0.0003 [−0.0015, +0.0006]	**+0.0001** [+0.0001, +0.0001]	+0.0018 [−0.0008, +0.0040]	+0.0004 [−0.0012, +0.0019]
k_T threshold (q = 0.7, p = −1)	**+0.0042** [+0.0042, +0.0042]	−0.0007 [−0.0018, +0.0007]	−0.0007 [−0.0020, +0.0006]	**−0.0013** [−0.0013, −0.0013]	+0.0010 [−0.0003, +0.0022]	**+0.0023** [+0.0015, +0.0032]
k_T threshold (q = 0.8, $k_{\max}$ = 5, p = −1)	**+0.0053** [+0.0053, +0.0053]	−0.0000 [−0.0022, +0.0027]	−0.0006 [−0.0013, +0.0001]	**+0.0052** [+0.0052, +0.0052]	+0.0005 [−0.0003, +0.0012]	−0.0012 [−0.0024, +0.0000]
k_T threshold (q = 0.8, p = −1)	**+0.0052** [+0.0052, +0.0052]	−0.0008 [−0.0023, +0.0011]	**−0.0011** [−0.0020, −0.0001]	**+0.0010** [+0.0010, +0.0010]	+0.0007 [−0.0017, +0.0027]	+0.0003 [−0.0010, +0.0020]
k_T threshold (q = 0.8, p = 0)	**+0.0061** [+0.0061, +0.0061]	+0.0013 [−0.0004, +0.0034]	−0.0002 [−0.0013, +0.0009]	**+0.0003** [+0.0003, +0.0003]	+0.0006 [−0.0014, +0.0029]	−0.0002 [−0.0018, +0.0011]
k_T threshold (q = 0.8, p = +1)	**+0.0046** [+0.0046, +0.0046]	−0.0007 [−0.0026, +0.0017]	**−0.0014** [−0.0025, −0.0004]	**+0.0016** [+0.0016, +0.0016]	+0.0012 [−0.0010, +0.0030]	**+0.0020** [+0.0005, +0.0037]
k_T threshold (q = 0.8, $k_{\max}$ = 10, p = −1)	**+0.0045** [+0.0045, +0.0045]	+0.0001 [−0.0015, +0.0022]	+0.0002 [−0.0002, +0.0006]	**+0.0016** [+0.0016, +0.0016]	+0.0005 [−0.0017, +0.0021]	+0.0002 [−0.0019, +0.0017]
k_T threshold (q = 0.9, p = −1)	**+0.0048** [+0.0048, +0.0048]	−0.0010 [−0.0043, +0.0022]	−0.0014 [−0.0033, +0.0001]	**+0.0017** [+0.0017, +0.0017]	+0.0015 [−0.0001, +0.0026]	−0.0003 [−0.0014, +0.0005]
k_T threshold (q = 0.9, p = 0)	**+0.0055** [+0.0055, +0.0055]	−0.0003 [−0.0017, +0.0012]	−0.0004 [−0.0019, +0.0016]	**+0.0035** [+0.0035, +0.0035]	**+0.0018** [+0.0014, +0.0023]	+0.0001 [−0.0004, +0.0006]
k_T threshold (q = 0.9, p = +1)	**+0.0049** [+0.0049, +0.0049]	+0.0002 [−0.0010, +0.0020]	−0.0009 [−0.0024, +0.0003]	**−0.0004** [−0.0004, −0.0004]	+0.0005 [−0.0007, +0.0018]	−0.0002 [−0.0007, +0.0002]
Laman (p = −1)*	**+0.0055** [+0.0055, +0.0055]	+0.0003 [−0.0018, +0.0022]	−0.0013 [−0.0028, +0.0000]	**+0.0024** [+0.0024, +0.0024]	**+0.0021** [+0.0003, +0.0034]	+0.0009 [−0.0000, +0.0017]
Laman (p = 0)*	**+0.0054** [+0.0054, +0.0054]	−0.0011 [−0.0024, +0.0005]	**−0.0031** [−0.0048, −0.0016]	**−0.0030** [−0.0030, −0.0030]	−0.0002 [−0.0014, +0.0013]	+0.0005 [−0.0008, +0.0018]
Laman (p = +1)*	**+0.0054** [+0.0054, +0.0054]	−0.0011 [−0.0024, +0.0005]	**−0.0031** [−0.0048, −0.0016]	**−0.0030** [−0.0030, −0.0030]	−0.0002 [−0.0014, +0.0013]	**+0.0016** [+0.0009, +0.0027]
Unique-3 (p = −1)*	**+0.0057** [+0.0057, +0.0057]	+0.0013 [−0.0005, +0.0033]	−0.0010 [−0.0027, +0.0006]	**+0.0018** [+0.0018, +0.0018]	+0.0013 [−0.0004, +0.0030]	+0.0009 [−0.0007, +0.0022]
Unique-3 (p = 0)*	**+0.0043** [+0.0043, +0.0043]	−0.0010 [−0.0022, +0.0010]	**−0.0020** [−0.0027, −0.0010]	**−0.0006** [−0.0006, −0.0006]	**+0.0019** [+0.0011, +0.0026]	+0.0001 [−0.0008, +0.0014]
Unique-3 (p = +1)*	**+0.0043** [+0.0043, +0.0043]	−0.0010 [−0.0022, +0.0010]	**−0.0020** [−0.0027, −0.0010]	**−0.0006** [−0.0006, −0.0006]	**+0.0019** [+0.0011, +0.0026]	−0.0004 [−0.0015, +0.0006]
Unique-4 (p = −1)	**+0.0054** [+0.0054, +0.0054]	+0.0001 [−0.0022, +0.0022]	−0.0006 [−0.0029, +0.0013]	**−0.0024** [−0.0024, −0.0024]	**+0.0019** [+0.0004, +0.0030]	+0.0015 [−0.0001, +0.0039]
Unique-4 (p = 0)	**+0.0050** [+0.0050, +0.0050]	−0.0002 [−0.0019, +0.0017]	−0.0020 [−0.0043, +0.0001]	**+0.0000** [+0.0000, +0.0000]	**+0.0022** [+0.0014, +0.0031]	−0.0001 [−0.0013, +0.0006]
Unique-4 (p = +1)	**+0.0050** [+0.0050, +0.0050]	−0.0002 [−0.0019, +0.0018]	−0.0020 [−0.0044, +0.0001]	**+0.0000** [+0.0000, +0.0000]	**+0.0022** [+0.0014, +0.0031]	+0.0006 [−0.0004, +0.0013]
Unique-5 (p = −1)	**+0.0053** [+0.0053, +0.0053]	**+0.0012** [+0.0001, +0.0029]	+0.0004 [−0.0008, +0.0017]	**+0.0011** [+0.0011, +0.0011]	**+0.0030** [+0.0019, +0.0041]	+0.0001 [−0.0023, +0.0023]
Unique-5 (p = 0)	**+0.0051** [+0.0051, +0.0051]	+0.0002 [−0.0012, +0.0016]	−0.0010 [−0.0023, +0.0008]	**−0.0019** [−0.0019, −0.0019]	+0.0015 [−0.0011, +0.0039]	+0.0001 [−0.0008, +0.0009]

G.4 JetSet: τ vs light

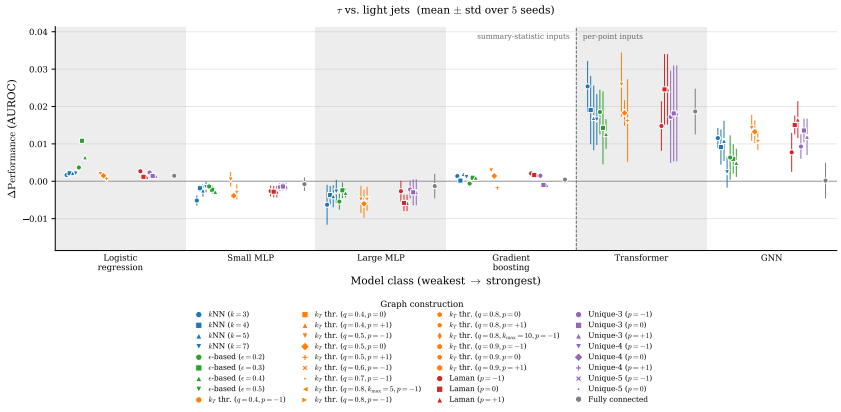


Figure 30: Incremental predictive value of curvature (Equation 7) for τ vs light jets for 36 graph constructions. Each marker is ΔPerformance, the AUROC with curvature appended to the inputs minus the AUROC without it, for one graph construction (color and marker shape) and one model class, with model classes (with learning rate (10^{-3})) ordered from weakest to strongest along the horizontal axis. Markers show the mean of the per-seed deltas over 5 seeds and error bars their standard deviation across seeds. Error bars combine seed to seed and test set bootstrap standard deviations in quadrature ($\sqrt{(\sigma_{seed})^2 + \sigma_{boot}^2}$). For logistic regression and gradient boosting the seed component is zero, so their bars show bootstrap uncertainty alone.

Table 30: Test AUROC for τ vs light jets for the 12 default JetSet constructions and the four sweep constructions available at submission (kNN at $k = 4, 5, 7$ and the ϵ-based construction at $\epsilon = 0.2$), learning rate 10^{-3}. (a) Baselines without curvature features for the five models that never see the graph, which therefore do not depend on the construction. (b) For each construction, the GNN baseline without curvature features and every model with FRC and ORC features computed on that construction. Rows are grouped by family, and * marks the default setting of each family; the sweep constructions come from a separate training run whose transformer baseline is 0.8505 [0.8479, 0.8533]. Entries give the mean over 5 training seeds; brackets give the 95% percentile bootstrap confidence interval of that mean, obtained by resampling the seeds with replacement 10,000 times, averaging each resample, and taking the 2.5th and 97.5th percentiles of the resampled means.

(a) Without curvature features: the five models that never see the graph

Logistic Regression	Small MLP	Large MLP	Gradient Boosting	Transformer
0.7957 [0.7957, 0.7957]	0.8635 [0.8627, 0.8642]	0.8643 [0.8639, 0.8646]	0.8649 [0.8649, 0.8649]	0.8490 [0.8467, 0.8515]

(b) Per construction: GNN without curvature features, and all models with them

	Without curvature	With curvature features					
Graph construction	GNN	Logistic Regression	Small MLP	Large MLP	Gradient Boosting	Transformer	GNN
Fully connected*	0.8788 [0.8769, 0.8803]	0.7971 [0.7971, 0.7971]	0.8627 [0.8617, 0.8635]	0.8630 [0.8605, 0.8653]	0.8654 [0.8654, 0.8654]	0.8677 [0.8619, 0.8717]	0.8790 [0.8766, 0.8812]
kNN ($k = 3$)*	0.8581 [0.8565, 0.8597]	0.7973 [0.7973, 0.7973]	0.8584 [0.8576, 0.8591]	0.8580 [0.8540, 0.8615]	0.8663 [0.8663, 0.8663]	0.8744 [0.8685, 0.8799]	0.8697 [0.8672, 0.8726]
kNN ($k = 4$)	0.8656 [0.8631, 0.8680]	0.7978 [0.7978, 0.7978]	0.8617 [0.8606, 0.8626]	0.8606 [0.8586, 0.8617]	0.8651 [0.8651, 0.8651]	0.8695 [0.8620, 0.8771]	0.8747 [0.8729, 0.8759]
kNN ($k = 5$)	0.8661 [0.8627, 0.8694]	0.7979 [0.7979, 0.7979]	0.8610 [0.8594, 0.8626]	0.8603 [0.8580, 0.8624]	0.8669 [0.8669, 0.8669]	0.8674 [0.8616, 0.8738]	0.8770 [0.8739, 0.8801]
kNN ($k = 7$)	0.8755 [0.8740, 0.8774]	0.7977 [0.7977, 0.7977]	0.8620 [0.8600, 0.8632]	0.8615 [0.8595, 0.8639]	0.8660 [0.8660, 0.8660]	0.8670 [0.8622, 0.8723]	0.8779 [0.8750, 0.8801]
ϵ-based ($\epsilon = 0.2$)	0.8622 [0.8585, 0.8653]	0.7993 [0.7993, 0.7993]	0.8621 [0.8616, 0.8626]	0.8588 [0.8570, 0.8606]	0.8643 [0.8643, 0.8643]	0.8690 [0.8651, 0.8737]	0.8686 [0.8661, 0.8709]
ϵ-based ($\epsilon = 0.3$)*	0.8711 [0.8681, 0.8740]	0.8065 [0.8065, 0.8065]	0.8613 [0.8602, 0.8625]	0.8618 [0.8606, 0.8633]	0.8659 [0.8659, 0.8659]	0.8633 [0.8584, 0.8695]	0.8770 [0.8750, 0.8793]
k_T threshold ($q = 0.5$, $p = -1$)*	0.8579 [0.8563, 0.8597]	0.7976 [0.7976, 0.7976]	0.8640 [0.8630, 0.8652]	0.8594 [0.8567, 0.8621]	0.8678 [0.8678, 0.8678]	0.8749 [0.8698, 0.8790]	0.8722 [0.8691, 0.8745]
k_T threshold ($q = 0.5$, $p = 0$)*	0.8572 [0.8559, 0.8584]	0.7972 [0.7972, 0.7972]	0.8596 [0.8592, 0.8601]	0.8582 [0.8549, 0.8603]	0.8664 [0.8664, 0.8664]	0.8673 [0.8632, 0.8714]	0.8704 [0.8689, 0.8714]
k_T threshold ($q = 0.5$, $p = +1$)*	0.8570 [0.8557, 0.8584]	0.7964 [0.7964, 0.7964]	0.8607 [0.8597, 0.8618]	0.8596 [0.8568, 0.8620]	0.8632 [0.8632, 0.8632]	0.8652 [0.8576, 0.8723]	0.8679 [0.8663, 0.8693]
Laman ($p = -1$)*	0.8637 [0.8599, 0.8675]	0.7983 [0.7983, 0.7983]	0.8609 [0.8594, 0.8621]	0.8616 [0.8594, 0.8636]	0.8670 [0.8670, 0.8670]	0.8638 [0.8600, 0.8682]	0.8714 [0.8679, 0.8746]
Laman ($p = 0$)*	0.8558 [0.8544, 0.8572]	0.7969 [0.7969, 0.7969]	0.8607 [0.8595, 0.8619]	0.8585 [0.8569, 0.8595]	0.8666 [0.8666, 0.8666]	0.8736 [0.8676, 0.8790]	0.8709 [0.8694, 0.8723]
Laman ($p = +1$)*	0.8554 [0.8528, 0.8574]	0.7969 [0.7969, 0.7969]	0.8607 [0.8595, 0.8619]	0.8585 [0.8569, 0.8595]	0.8666 [0.8666, 0.8666]	0.8736 [0.8674, 0.8790]	0.8718 [0.8699, 0.8736]
Unique-3 ($p = -1$)*	0.8623 [0.8604, 0.8642]	0.7980 [0.7980, 0.7980]	0.8620 [0.8613, 0.8627]	0.8621 [0.8602, 0.8641]	0.8664 [0.8664, 0.8664]	0.8662 [0.8562, 0.8754]	0.8716 [0.8700, 0.8733]
Unique-3 ($p = 0$)*	0.8591 [0.8564, 0.8608]	0.7971 [0.7971, 0.7971]	0.8621 [0.8618, 0.8624]	0.8613 [0.8589, 0.8637]	0.8639 [0.8639, 0.8639]	0.8671 [0.8564, 0.8767]	0.8727 [0.8718, 0.8734]
Unique-3 ($p = +1$)*	0.8603 [0.8581, 0.8626]	0.7971 [0.7971, 0.7971]	0.8621 [0.8618, 0.8624]	0.8613 [0.8589, 0.8637]	0.8639 [0.8639, 0.8639]	0.8671 [0.8564, 0.8767]	0.8722 [0.8695, 0.8754]

Table 31: Change in test AUROC from adding Ricci curvature features for τ vs light jets (learning rate 10^{-3}; Equation 7) for the 12 default JetSet constructions and the four sweep constructions available at submission, computed per seed between runs that share architecture, hyperparameters, data, and seed. Rows are grouped by family, and * marks the default setting of each family. Entries give the mean over 5 training seeds; brackets give the 95% percentile bootstrap confidence interval of that mean, obtained by resampling the seeds with replacement 10,000 times, averaging each resample, and taking the 2.5th and 97.5th percentiles of the resampled means. Seeds are resampled jointly for both settings, so the pairing is preserved. Bold marks entries whose 95% interval excludes zero, a consistent gain (positive) or loss (negative) from adding curvature.

Graph construction	Logistic Regression	Small MLP	Large MLP	Gradient Boosting	Transformer	GNN
Fully connected*	**+0.0015** [+0.0015, +0.0015]	−0.0008 [−0.0022, +0.0006]	−0.0013 [−0.0041, +0.0011]	**+0.0004** [+0.0004, +0.0004]	**+0.0187** [+0.0141, +0.0233]	+0.0002 [−0.0035, +0.0040]
kNN ($k = 3$)*	**+0.0017** [+0.0017, +0.0017]	**−0.0052** [−0.0061, −0.0040]	**−0.0063** [−0.0105, −0.0025]	**+0.0014** [+0.0014, +0.0014]	**+0.0254** [+0.0201, +0.0303]	**+0.0115** [+0.0091, +0.0135]
kNN ($k = 4$)	**+0.0021** [+0.0021, +0.0021]	**−0.0019** [−0.0025, −0.0011]	**−0.0037** [−0.0059, −0.0024]	**+0.0002** [+0.0002, +0.0002]	**+0.0190** [+0.0124, +0.0267]	**+0.0091** [+0.0055, +0.0128]
kNN ($k = 5$)	**+0.0022** [+0.0022, +0.0022]	**−0.0025** [−0.0038, −0.0012]	**−0.0040** [−0.0065, −0.0019]	**+0.0019** [+0.0019, +0.0019]	**+0.0170** [+0.0109, +0.0242]	**+0.0108** [+0.0063, +0.0145]
kNN ($k = 7$)	**+0.0021** [+0.0021, +0.0021]	**−0.0015** [−0.0028, −0.0007]	**−0.0027** [−0.0051, −0.0001]	**+0.0011** [+0.0011, +0.0011]	**+0.0165** [+0.0119, +0.0225]	+0.0024 [−0.0008, +0.0056]
ϵ-based ($\epsilon = 0.2$)	**+0.0037** [+0.0037, +0.0037]	**−0.0014** [−0.0020, −0.0007]	**−0.0054** [−0.0070, −0.0039]	**−0.0006** [−0.0006, −0.0006]	**+0.0185** [+0.0143, +0.0231]	**+0.0064** [+0.0018, +0.0108]
ϵ-based ($\epsilon = 0.3$)*	**+0.0108** [+0.0108, +0.0108]	**−0.0023** [−0.0029, −0.0013]	**−0.0024** [−0.0039, −0.0008]	**+0.0009** [+0.0009, +0.0009]	**+0.0143** [+0.0071, +0.0226]	**+0.0060** [+0.0028, +0.0090]
k_T threshold ($q = 0.5$, $p = -1$)*	**+0.0019** [+0.0019, +0.0019]	+0.0005 [−0.0010, +0.0020]	**−0.0049** [−0.0077, −0.0020]	**+0.0029** [+0.0029, +0.0029]	**+0.0259** [+0.0182, +0.0310]	**+0.0143** [+0.0113, +0.0166]
k_T threshold ($q = 0.5$, $p = 0$)*	**+0.0015** [+0.0015, +0.0015]	**−0.0039** [−0.0046, −0.0030]	**−0.0060** [−0.0093, −0.0037]	**+0.0014** [+0.0014, +0.0014]	**+0.0183** [+0.0155, +0.0208]	**+0.0132** [+0.0108, +0.0153]
k_T threshold ($q = 0.5$, $p = +1$)*	**+0.0007** [+0.0007, +0.0007]	**−0.0028** [−0.0044, −0.0013]	**−0.0047** [−0.0074, −0.0024]	**−0.0017** [−0.0017, −0.0017]	**+0.0162** [+0.0075, +0.0246]	**+0.0108** [+0.0090, +0.0130]
Laman ($p = -1$)*	**+0.0027** [+0.0027, +0.0027]	**−0.0026** [−0.0036, −0.0015]	**−0.0027** [−0.0049, −0.0006]	**+0.0021** [+0.0021, +0.0021]	**+0.0148** [+0.0105, +0.0206]	**+0.0077** [+0.0041, +0.0121]
Laman ($p = 0$)*	**+0.0012** [+0.0012, +0.0012]	**−0.0028** [−0.0040, −0.0016]	**−0.0058** [−0.0078, −0.0045]	**+0.0017** [+0.0017, +0.0017]	**+0.0246** [+0.0173, +0.0320]	**+0.0150** [+0.0131, +0.0168]
Laman ($p = +1$)*	**+0.0012** [+0.0012, +0.0012]	**−0.0028** [−0.0040, −0.0016]	**−0.0058** [−0.0078, −0.0045]	**+0.0017** [+0.0017, +0.0017]	**+0.0246** [+0.0171, +0.0320]	**+0.0165** [+0.0127, +0.0201]
Unique-3 ($p = -1$)*	**+0.0023** [+0.0023, +0.0023]	**−0.0016** [−0.0024, −0.0007]	**−0.0022** [−0.0040, −0.0003]	**+0.0015** [+0.0015, +0.0015]	**+0.0172** [+0.0078, +0.0275]	**+0.0093** [+0.0063, +0.0111]
Unique-3 ($p = 0$)*	**+0.0014** [+0.0014, +0.0014]	**−0.0014** [−0.0022, −0.0006]	**−0.0030** [−0.0056, −0.0003]	**−0.0010** [−0.0010, −0.0010]	**+0.0181** [+0.0087, +0.0283]	**+0.0136** [+0.0112, +0.0165]
Unique-3 ($p = +1$)*	**+0.0014** [+0.0014, +0.0014]	**−0.0014** [−0.0022, −0.0006]	**−0.0030** [−0.0056, −0.0003]	**−0.0010** [−0.0010, −0.0010]	**+0.0181** [+0.0087, +0.0283]	**+0.0119** [+0.0080, +0.0156]

G.5 b VS LIGHT AND c VS LIGHT WITH LEARNING RATE OF 3×10^{-4}

The per-track models are retrained at learning rate 3×10^{-4}, the rate of the degree and track count controls, for b vs light (Table 32, Table 33) and c vs light (Table 34, Table 35).

Table 32: Test AUROC for b vs light jets for each graph construction and model class (learning rate 3×10^{-4}), (a) without and (b) with FRC and ORC features. Entries give the mean over 5 training seeds; brackets give the 95% percentile bootstrap confidence interval of that mean, obtained by resampling the seeds with replacement 10,000 times, averaging each resample, and taking the 2.5th and 97.5th percentiles of the resampled means.

(a) Without curvature features

Graph construction	Transformer	GNN
Fully connected	0.7026 [0.7014, 0.7036]	0.7153 [0.7134, 0.7166]
kNN ($k = 3$)	0.7026 [0.7014, 0.7036]	0.6973 [0.6971, 0.6975]
k_T threshold ($p = -1$)	0.7026 [0.7014, 0.7036]	0.7014 [0.6981, 0.7036]
Laman ($p = -1$)	0.7026 [0.7014, 0.7036]	0.6999 [0.6989, 0.7010]
ϵ-based ($\epsilon = 0.3$)	0.7026 [0.7014, 0.7036]	0.7078 [0.7055, 0.7101]
Unique-3 ($p = -1$)	0.7026 [0.7014, 0.7036]	0.7064 [0.7055, 0.7072]

(b) With curvature features

Graph construction	Transformer	GNN
Fully connected	0.7191 [0.7186, 0.7197]	0.7215 [0.7212, 0.7218]
kNN ($k = 3$)	0.7175 [0.7161, 0.7186]	0.7186 [0.7183, 0.7188]
k_T threshold ($p = -1$)	0.7172 [0.7158, 0.7185]	0.7186 [0.7182, 0.7189]
Laman ($p = -1$)	0.7182 [0.7157, 0.7206]	0.7197 [0.7192, 0.7202]
ϵ-based ($\epsilon = 0.3$)	0.7213 [0.7203, 0.7224]	0.7195 [0.7189, 0.7198]
Unique-3 ($p = -1$)	0.7204 [0.7183, 0.7222]	0.7209 [0.7205, 0.7212]

Table 33: Change in test AUROC from adding Ricci curvature features for b vs light jets (learning rate 3×10^{-4}), computed per seed between runs that share architecture, hyperparameters, data, and seed. Entries give the mean over 5 training seeds; brackets give the 95% percentile bootstrap confidence interval of that mean, obtained by resampling the seeds with replacement 10,000 times, averaging each resample, and taking the 2.5th and 97.5th percentiles of the resampled means. Seeds are resampled jointly for both settings, so the pairing is preserved. Bold entries have an interval that excludes zero.

Graph construction	Transformer	GNN
Fully connected	**+0.0165** [+0.0157, +0.0173]	**+0.0062** [+0.0047, +0.0080]
kNN ($k = 3$)	**+0.0148** [+0.0134, +0.0168]	**+0.0213** [+0.0210, +0.0216]
k_T threshold ($p = -1$)	**+0.0146** [+0.0132, +0.0162]	**+0.0172** [+0.0150, +0.0203]
Laman ($p = -1$)	**+0.0156** [+0.0142, +0.0173]	**+0.0198** [+0.0185, +0.0211]
ϵ-based ($\epsilon = 0.3$)	**+0.0187** [+0.0170, +0.0203]	**+0.0116** [+0.0096, +0.0138]
Unique-3 ($p = -1$)	**+0.0177** [+0.0168, +0.0187]	**+0.0145** [+0.0134, +0.0157]

Table 34: Test AUROC for c vs light jets for each graph construction and model class (learning rate 3×10^{-4}), (a) without and (b) with FRC and ORC features. Entries give the mean over 5 training seeds; brackets give the 95% percentile bootstrap confidence interval of that mean, obtained by resampling the seeds with replacement 10,000 times, averaging each resample, and taking the 2.5th and 97.5th percentiles of the resampled means.

(a) Without curvature features

Graph construction	Transformer	GNN
Fully connected	0.6160 [0.6150, 0.6169]	0.6161 [0.6155, 0.6165]
kNN ($k = 3$)	0.6160 [0.6150, 0.6169]	0.6148 [0.6141, 0.6156]
k_T threshold ($p = -1$)	0.6160 [0.6150, 0.6169]	0.6144 [0.6137, 0.6150]
Laman ($p = -1$)	0.6160 [0.6150, 0.6169]	0.6143 [0.6136, 0.6148]
ϵ-based ($\epsilon = 0.3$)	0.6160 [0.6150, 0.6169]	0.6146 [0.6142, 0.6151]
Unique-3 ($p = -1$)	0.6160 [0.6150, 0.6169]	0.6149 [0.6142, 0.6154]

(b) With curvature features

Graph construction	Transformer	GNN
Fully connected	0.6192 [0.6179, 0.6204]	0.6173 [0.6165, 0.6180]
kNN ($k = 3$)	0.6209 [0.6195, 0.6224]	0.6171 [0.6164, 0.6177]
k_T threshold ($p = -1$)	0.6173 [0.6161, 0.6184]	0.6175 [0.6168, 0.6182]
Laman ($p = -1$)	0.6194 [0.6179, 0.6209]	0.6178 [0.6173, 0.6184]
ϵ-based ($\epsilon = 0.3$)	0.6202 [0.6189, 0.6215]	0.6175 [0.6168, 0.6182]
Unique-3 ($p = -1$)	0.6189 [0.6162, 0.6214]	0.6168 [0.6155, 0.6176]

Table 35: Change in test AUROC from adding Ricci curvature features for c vs light jets (learning rate 3×10^{-4}), computed per seed between runs that share architecture, hyperparameters, data, and seed. Entries give the mean over 5 training seeds; brackets give the 95% percentile bootstrap confidence interval of that mean, obtained by resampling the seeds with replacement 10,000 times, averaging each resample, and taking the 2.5th and 97.5th percentiles of the resampled means. Seeds are resampled jointly for both settings, so the pairing is preserved. Bold entries have an interval that excludes zero.

Graph construction	Transformer	GNN
Fully connected	**+0.0032** [+0.0017, +0.0050]	**+0.0012** [+0.0004, +0.0017]
kNN ($k = 3$)	**+0.0049** [+0.0035, +0.0063]	**+0.0023** [+0.0020, +0.0025]
k_T threshold ($p = -1$)	+0.0013 [−0.0004, +0.0022]	**+0.0032** [+0.0026, +0.0036]
Laman ($p = -1$)	**+0.0034** [+0.0010, +0.0055]	**+0.0036** [+0.0030, +0.0041]
ϵ-based ($\epsilon = 0.3$)	**+0.0042** [+0.0032, +0.0054]	**+0.0029** [+0.0017, +0.0040]
Unique-3 ($p = -1$)	**+0.0029** [+0.0002, +0.0056]	**+0.0018** [+0.0003, +0.0029]

G.6 Degree control

Each node's degree under the construction replaces curvature as the only extra channel, for b vs light at learning rate 3×10^{-4} (§ 5.3). Table 36 gives the AUROC with and without degree and Table 37 the change.

Table 36: Test AUROC for b vs light jets at learning rate 3×10^{-4}, (a) without and (b) with each node's degree under the construction as the only extra channel. Entries give the mean over 5 training seeds; brackets give the 95% percentile bootstrap confidence interval of that mean, obtained by resampling the seeds with replacement 10,000 times, averaging each resample, and taking the 2.5th and 97.5th percentiles of the resampled means. Seeds are resampled jointly for both settings, so the pairing is preserved. Bold entries have an interval that excludes zero. This was done with cuda device 1.

(a) Without degree features

Graph construction	Transformer	GNN
kNN ($k = 3$)	0.7039 [0.7025, 0.7056]	0.6986 [0.6979, 0.6992]
k_T threshold ($p = -1$)	0.7039 [0.7025, 0.7056]	0.7036 [0.7028, 0.7042]
k_T threshold ($p = +1$)	0.7039 [0.7025, 0.7056]	0.7028 [0.7016, 0.7038]
k_T threshold ($p = 0$)	0.7039 [0.7025, 0.7056]	0.7036 [0.7023, 0.7048]
Laman ($p = -1$)	0.7039 [0.7025, 0.7056]	0.7010 [0.7002, 0.7018]
Laman ($p = +1$)	0.7039 [0.7025, 0.7056]	0.6968 [0.6959, 0.6975]
Laman ($p = 0$)	0.7039 [0.7025, 0.7056]	0.6960 [0.6953, 0.6968]
ϵ-based ($\epsilon = 0.3$)	0.7039 [0.7025, 0.7056]	0.7082 [0.7055, 0.7105]

(b) With degree features

Graph construction	Transformer	GNN
kNN ($k = 3$)	0.7129 [0.7115, 0.7143]	0.7133 [0.7130, 0.7136]
k_T threshold ($p = -1$)	0.7206 [0.7199, 0.7214]	0.7195 [0.7191, 0.7199]
k_T threshold ($p = +1$)	0.7211 [0.7201, 0.7220]	0.7176 [0.7174, 0.7179]
k_T threshold ($p = 0$)	0.7196 [0.7187, 0.7206]	0.7195 [0.7192, 0.7197]
Laman ($p = -1$)	0.7215 [0.7209, 0.7220]	0.7180 [0.7177, 0.7184]
Laman ($p = +1$)	0.7197 [0.7190, 0.7204]	0.7172 [0.7170, 0.7175]
Laman ($p = 0$)	0.7197 [0.7190, 0.7204]	0.7170 [0.7165, 0.7175]
ϵ-based ($\epsilon = 0.3$)	0.7213 [0.7206, 0.7223]	0.7201 [0.7199, 0.7204]

Table 37: Test AUROC for b vs light jets at learning rate 3×10^{-4}, (a) without and (b) with each node's degree under the construction as the only extra channel. Entries give the mean over 5 training seeds; brackets give the 95% percentile bootstrap confidence interval of that mean, obtained by resampling the seeds with replacement 10,000 times, averaging each resample, and taking the 2.5th and 97.5th percentiles of the resampled means. Seeds are resampled jointly for both settings, so the pairing is preserved. Bold entries have an interval that excludes zero.

Graph construction	Transformer	GNN
kNN ($k = 3$)	**+0.0090** [+0.0074, +0.0111]	**+0.0146** [+0.0142, +0.0151]
k_T threshold ($p = -1$)	**+0.0167** [+0.0150, +0.0185]	**+0.0159** [+0.0150, +0.0171]
k_T threshold ($p = +1$)	**+0.0172** [+0.0151, +0.0192]	**+0.0147** [+0.0136, +0.0160]
k_T threshold ($p = 0$)	**+0.0157** [+0.0144, +0.0172]	**+0.0158** [+0.0144, +0.0173]
Laman ($p = -1$)	**+0.0176** [+0.0158, +0.0191]	**+0.0170** [+0.0162, +0.0176]
Laman ($p = +1$)	**+0.0158** [+0.0139, +0.0175]	**+0.0205** [+0.0197, +0.0212]
Laman ($p = 0$)	**+0.0158** [+0.0140, +0.0175]	**+0.0210** [+0.0204, +0.0217]

G.7 Track count control

This control adds the track count n to the GNN's baseline inputs and then adds ORC, FRC, or both, at the learning rate of 0.001.

Table 38: Track count control: GNN test AUROC with baseline features and the track count n, before adding curvature. Brackets: 95% seed-bootstrap intervals (Appendix G).

Graph construction	b vs light	c vs light	τ vs light
Fully connected	0.7252 [0.7249, 0.7256]	0.6207 [0.6195, 0.6220]	0.8801 [0.8795, 0.8810]
kNN ($k = 3$)	0.7223 [0.7219, 0.7226]	0.6207 [0.6200, 0.6215]	0.8762 [0.8750, 0.8775]
k_T threshold ($p = -1$)	0.7221 [0.7215, 0.7227]	0.6197 [0.6194, 0.6203]	0.8776 [0.8767, 0.8788]
k_T threshold ($p = +1$)	0.7217 [0.7212, 0.7222]	0.6210 [0.6201, 0.6219]	0.8772 [0.8742, 0.8788]
k_T threshold ($p = 0$)	0.7221 [0.7219, 0.7223]	0.6197 [0.6191, 0.6202]	0.8728 [0.8702, 0.8758]
Laman ($p = -1$)	0.7208 [0.7195, 0.7217]	0.6196 [0.6184, 0.6207]	0.8777 [0.8765, 0.8787]
Laman ($p = +1$)	0.7215 [0.7213, 0.7217]	0.6199 [0.6191, 0.6208]	0.8766 [0.8754, 0.8778]
ϵ-based ($\epsilon = 0.3$)	0.7233 [0.7229, 0.7236]	0.6192 [0.6180, 0.6203]	0.8699 [0.8688, 0.8709]
Unique-3 ($p = -1$)	0.7224 [0.7222, 0.7227]	0.6197 [0.6187, 0.6203]	0.8794 [0.8768, 0.8817]
Unique-3 ($p = +1$)	0.7224 [0.7215, 0.7230]	0.6198 [0.6184, 0.6207]	0.8785 [0.8771, 0.8797]

Table 39: Change in GNN test AUROC for b vs light from adding curvature on top of baseline features and the track count n (Table 38). Brackets: 95% seed-bootstrap intervals; bold: interval excludes zero.

Graph construction	+ORC	+FRC	FRC+ORC
Fully connected	**+0.0008** [+0.0003, +0.0013]	**+0.0005** [+0.0000, +0.0011]	−0.0002 [−0.0013, +0.0010]
kNN ($k = 3$)	−0.0003 [−0.0012, +0.0007]	+0.0002 [−0.0004, +0.0008]	−0.0006 [−0.0012, +0.0000]
k_T threshold ($p = -1$)	**−0.0012** [−0.0018, −0.0006]	−0.0004 [−0.0010, +0.0002]	**−0.0023** [−0.0042, −0.0007]
k_T threshold ($p = +1$)	−0.0005 [−0.0013, +0.0003]	−0.0007 [−0.0016, +0.0003]	**−0.0013** [−0.0021, −0.0005]
k_T threshold ($p = 0$)	−0.0003 [−0.0007, +0.0001]	**−0.0012** [−0.0018, −0.0005]	**−0.0012** [−0.0017, −0.0007]
Laman ($p = -1$)	+0.0003 [−0.0002, +0.0012]	+0.0006 [−0.0002, +0.0017]	−0.0001 [−0.0019, +0.0016]
Laman ($p = +1$)	**−0.0006** [−0.0014, −0.0000]	−0.0005 [−0.0013, +0.0001]	−0.0012 [−0.0028, +0.0000]
ϵ-based ($\epsilon = 0.3$)	**−0.0006** [−0.0008, −0.0004]	**+0.0006** [+0.0001, +0.0012]	−0.0004 [−0.0014, +0.0006]
Unique-3 ($p = -1$)	−0.0004 [−0.0009, +0.0001]	+0.0001 [−0.0005, +0.0007]	−0.0006 [−0.0012, +0.0002]
Unique-3 ($p = +1$)	+0.0005 [−0.0002, +0.0015]	−0.0008 [−0.0023, +0.0011]	−0.0008 [−0.0023, +0.0007]

Table 40: As Table 39, for c vs light.

Graph construction	+ORC	+FRC	FRC+ORC
Fully connected	−0.0008 [−0.0028, +0.0010]	−0.0014 [−0.0037, +0.0005]	−0.0017 [−0.0040, +0.0003]
kNN ($k = 3$)	**−0.0027** [−0.0041, −0.0013]	−0.0000 [−0.0014, +0.0010]	**−0.0013** [−0.0022, −0.0005]
k_T threshold ($p = -1$)	**−0.0023** [−0.0034, −0.0012]	**−0.0035** [−0.0047, −0.0027]	**−0.0025** [−0.0035, −0.0011]
k_T threshold ($p = +1$)	**−0.0018** [−0.0038, −0.0004]	−0.0013 [−0.0033, +0.0000]	−0.0011 [−0.0024, +0.0004]
k_T threshold ($p = 0$)	**−0.0019** [−0.0038, −0.0003]	−0.0008 [−0.0025, +0.0006]	**−0.0017** [−0.0024, −0.0009]
Laman ($p = -1$)	−0.0012 [−0.0030, +0.0011]	−0.0016 [−0.0037, +0.0005]	−0.0005 [−0.0021, +0.0011]
Laman ($p = +1$)	**−0.0021** [−0.0039, −0.0004]	**−0.0035** [−0.0054, −0.0018]	**−0.0032** [−0.0046, −0.0019]
ϵ-based ($\epsilon = 0.3$)	−0.0009 [−0.0022, +0.0005]	−0.0006 [−0.0018, +0.0007]	+0.0000 [−0.0021, +0.0022]
Unique-3 ($p = -1$)	−0.0013 [−0.0034, +0.0007]	−0.0000 [−0.0011, +0.0009]	−0.0001 [−0.0010, +0.0008]
Unique-3 ($p = +1$)	−0.0008 [−0.0019, +0.0002]	−0.0008 [−0.0031, +0.0013]	**−0.0016** [−0.0032, −0.0002]

Table 41: As Table 39, for τ vs light.

Graph construction	+ORC	+FRC	FRC+ORC
Fully connected	−0.0002 [−0.0020, +0.0016]	−0.0001 [−0.0035, +0.0031]	**−0.0042** [−0.0079, −0.0012]
kNN ($k = 3$)	−0.0015 [−0.0039, +0.0003]	−0.0028 [−0.0061, +0.0003]	**−0.0055** [−0.0079, −0.0033]
k_T threshold ($p = -1$)	**−0.0049** [−0.0058, −0.0042]	−0.0014 [−0.0038, +0.0008]	**−0.0055** [−0.0097, −0.0012]
k_T threshold ($p = +1$)	**−0.0062** [−0.0083, −0.0045]	**−0.0029** [−0.0044, −0.0007]	**−0.0084** [−0.0128, −0.0037]
k_T threshold ($p = 0$)	−0.0017 [−0.0049, +0.0020]	+0.0008 [−0.0017, +0.0032]	−0.0045 [−0.0104, +0.0013]
Laman ($p = -1$)	**−0.0059** [−0.0095, −0.0025]	−0.0008 [−0.0033, +0.0020]	**−0.0122** [−0.0162, −0.0083]
Laman ($p = +1$)	**−0.0041** [−0.0064, −0.0021]	**−0.0022** [−0.0037, −0.0005]	**−0.0105** [−0.0154, −0.0059]
ϵ-based ($\epsilon = 0.3$)	**−0.0039** [−0.0067, −0.0013]	+0.0030 [−0.0001, +0.0054]	**−0.0025** [−0.0047, −0.0006]
Unique-3 ($p = -1$)	−0.0022 [−0.0045, +0.0002]	**−0.0014** [−0.0024, −0.0005]	**−0.0081** [−0.0110, −0.0049]
Unique-3 ($p = +1$)	**−0.0039** [−0.0065, −0.0020]	**−0.0018** [−0.0029, −0.0009]	**−0.0078** [−0.0092, −0.0062]

G.8 Does separation predict the gain?

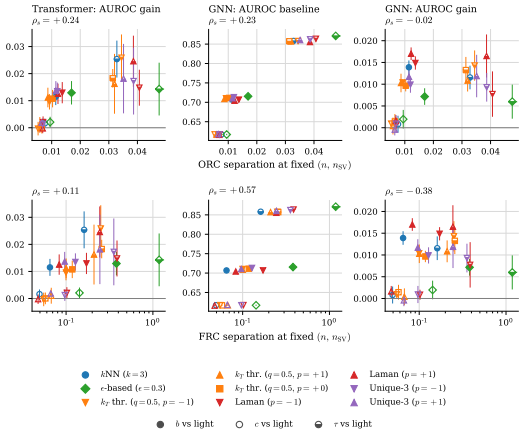


Figure 31: Separation at fixed (n, n_{sv}) against model performance. All nodes separation of b, c, and τ from light jets under ORC (top) and FRC (bottom) against the transformer's gain from curvature, the GNN's baseline AUROC, and the GNN's gain, for the nine sparse default constructions (color and marker) and three pairs (fill). Error bars are σ of § 4.2. ρ_s is the Spearman correlation of within-pair ranks pooled over the three pairs, with a 95% bootstrap interval over constructions (2000 resamples). FRC axis is in log scale since ϵ-based construction separates more than any others. The transformer baseline was excluded here since it does not depend on the construction.

Discussion. Fixing (n, n_{sv}) asks whether curvature carries label information beyond both counts, independent of any model; fixing n, which the models know, measures what curvature could add about displacement. Neither predicts the gain: at fixed n, $\rho_s = +0.04$ (ORC) and -0.28 (FRC) for the GNN and $+0.14$, -0.12 for the transformer. At fixed (n, n_{sv}), -0.02, -0.38 and $+0.24$,

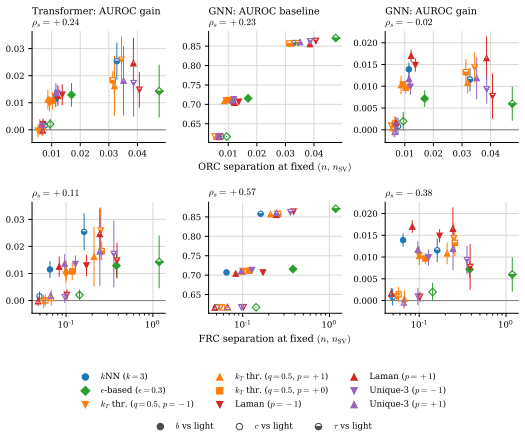


Figure 32: As Figure 32, conditioned on n alone. Since the models know n but not n_{sv}, this is the conditioning under which separation measures what curvature could add to them.

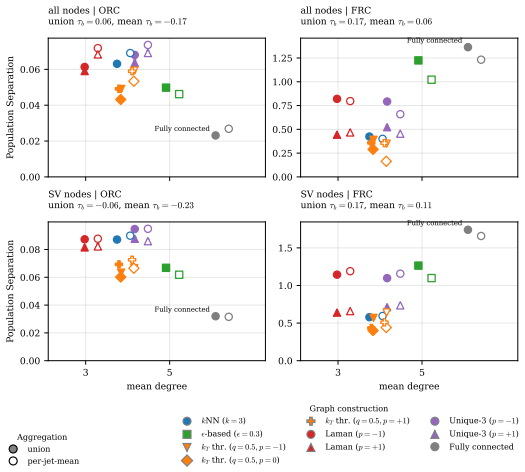


Figure 33: Unconditioned separation against mean degree for the nine sparse constructions on on JetSet. Panel: ORC (left) and FRC (right) with on all nodes (top) for both and SV nodes (bottom) for both. Filled marker: union aggregation and open markers are per jet mean. Each panel title gives Kendall's τ_b between mean degree and separation over the nine constructions. Under ORC the association is weak ($\tau_b = 0.06$ union, -0.17 per-jet mean on all nodes), so the ORC ranking of constructions is not a ranking by degree; under FRC it is somewhat stronger (0.17 and 0.06), as expected for a curvature that scales with degree. Values are from Table 15–Table 18; the τ_b values are those stated in § G.8.

+0.11, consistent with the degree control. Bootstrap intervals over constructions include zero for all eight gain correlations except FRC at fixed (n, n_{sv}) for the GNN ($[-0.68, -0.08]$), which a within-pair permutation test does not confirm ($p = 0.06$). FRC separation at fixed (n, n_{sv}) does track the GNN baseline ($\rho_s = +0.57$, $[+0.16, +0.81]$, permutation $p = 0.002$; $+0.38$ at fixed n and ORC $+0.23$, $+0.01$). This does not come from the ϵ-based construction, which tops both axes: without it $\rho_s = +0.55$ ($[+0.07, +0.81]$, $p = 0.009$). With nine constructions per pair the interval is wide, and mean degree does not explain it (Kendall's τ_b between mean degree and separation 0.06 for ORC, 0.17 for FRC). Before any training, the diagnostic still picks the ϵ-based construction, which is also the sparse construction with the best GNN baseline.