A Additional Proofs on Stability and Expressivity

A.1 Stability Proofs

Edge Filtrations versus Vertex Filtrations Our results are structured to address filtrations built by a function on the *edges* of a graph $G = (V, E)$, $g\colon E \to \mathbb{R}$. This matches our notions of discrete curvature, which are also defined edge-wise. g gives an explicit ordering on E and thus an induced ordering on V given by:

$$v \leq v' \iff \sum_{e \in E_v} g(e) \leq \sum_{e' \in E_{v'}} g(e')$$

where E_x is the set of edges incident to $x \in V$. However, one can also define a filtration over *vertices* with a scalar valued function $f\colon V \to \mathbb{R}$, and induce an ordering on edges. f can also attain a only finite number of values , call them $b_1, b_2, \ldots$ on the graph. Thus we can also compute a filtration $\emptyset \subseteq G_0 \subseteq G_1 \ldots \subseteq G_{k-1} \subseteq G_k = G$, where each $G_i := (V_i, E_i)$, with $V_i := \{v \in V \mid f(v) \leq b_i\}$ and $E_i := \{e \in E \mid \max_{v \in e} f(v) \leq b_i\}$. Similarly, f gives an explicit ordering on V, and induces an ordering on E, given by:

$$e \leq e' \iff \max_{v \in e} f(v) \leq \max_{v' \in e'} f(v')$$

The key idea here is that *either choice gives rise to an ordering of both edges, and vertices* that are used to calculate persistent homology of the graph. This means that the arguments for Theorem 1 and Theorem 5 also bound the bottle-neck distance for persistence diagrams generated using vertex filtrations.

Theorem 1. *Given graphs $F = (V_F, E_F)$ and $G = (V_G, E_G)$ with filtration functions f, g, and corresponding persistence diagrams D_f, D_g, we have $d_B(D_f, D_g) \leq \max\{\mathrm{dis}(f, g), \mathrm{dis}(g, f)\}$, where $\mathrm{dis}(f, g) := |\max_{x \in E_F} f(x) - \min_{y \in E_G} g(y)|$ and vice versa for $\mathrm{dis}(g, f)$.*

Proof. Considering the calculation of persistence diagrams based on scalar-valued filtrations functions, every point in the persistence diagram D_f can be written as a tuple of the form $(f(e_F), f(e'_F))$, with $e_F, e'_F \in E_F$; the sample applies for D_g. The inner distance between such tuples that occur in the bottleneck distance calculation can thus be written as

$$\|(f(e_F), f(e'_F)) - (g(e_G), g(e'_G))\|_\infty. \tag{6}$$

The maximum distance that can be achieved using this expression is determined by the maximum variation of the functions, expressed via $\mathrm{dis}(f, g)$ and $\mathrm{dis}(g, f)$, respectively. □

Graph perturbations. Here we explicitly specify a common framework used in the proofs for stability of curvature functions. As mentioned in the main text, we consider perturbations to *unweighted*, *connected* graphs $G = (V, E)$, with $|V| = n$ and $|E| = m$. In the case of *edge addition*, let $i*$ and $j*$ be arbitrary vertices that we wish to connect with a *new* edge, forming our new graph $G' = (V, E')$ where $E' = E \cup (i*, j*)$ such that $|E'| = m + 1$. For *edge deletion*, we similarly let $(i*, j*) \in E$ be the edge we delete such that $E' \subset E$ and $|E'| = m - 1$. Moreover, we only consider edges $(i*, j*)$ that leave G' connected.

A.1.1 Forman–Ricci Curvature

Theorem 2. *If G' is the graph generated by **edge addition**, then the updated Forman curvature κ'_{FR} for pre-existing edges $(i, j) \in E$ can be bounded by $\kappa_{FR}(i, j) - 1 \leq \kappa'_{FR}(i, j) \leq \kappa_{FR}(i, j) + 2$. If G' is the graph generated by **edge deletion**, then the updated Forman curvature κ'_{FR} for pre-existing edges $(i, j) \in E$ can be bounded by $\kappa_{FR}(i, j) - 2 \leq \kappa'_{FR}(i, j) \leq \kappa_{FR}(i, j) + 1$.*

Proof. We first handle the case of **edge addition**. By definition $\kappa_{\mathrm{FR}}(i, j)$ depends *only* on the degrees of the source and target $(i, j) \in E$ and the number of triangles formed using (i, j), $|\#_{\Delta_{ij}}| = |N(i) \cup N(j)|$, where $N(i), N(j)$ are the set of neighbouring nodes for i, j respectively.This is a local computation– all relevant information can be computed in the subgraph generated by $N(i) \cup N(j)$. Thus, in order to understand stability of κ_{FR}, we need to understand how $N(i)$ and $N(j)$ change under graph perturbations. For our new graph G', the only affected edges lie in the set:

$$E_{ij} := \{(u,v) \in E | u,v \in N(i) \cup N(j)\}$$

For the new edge $(i*, j*)$, we can directly compute $\kappa_{\text{FR}}(i*, j*)$ based on the original structure of the graph. However, in terms of stability we are interested in the member of E_{ij}, which can be split into two cases: one of the nodes is $i*$ or $j*$ or neither is. *Case 1*: WLOG assume the edge is of the form $(i*, v) \in E_{ij}$. Clearly, $d'_i = d_i + 1$. As for $|\#'_{\Delta_{iv}}|$, this can maximally be increased by 1 in the case that $x \in S(i) \cap S(j)$, else the triangle count stays the same. *Case 2*: consider $(u,v) \in E_{ij}$ where $u, v \in V \setminus \{i*, j*\}$. In this case, there is no change to the degree nor the number $|\#'_{\Delta_{uv}}|$. Clearly then, if $(i*, j*)$ forms a new triangle, our curvature can increase by 2, and if no triangle is formed the curvature can decrease by 1 in response to the increased degree. Thus we can bound $\kappa'_{\text{FR}}(i,j) := 4 - d'_i - d'_j + 3|\#'_{\Delta_{ij}}|$ as follows:

$$\kappa_{\text{FR}}(i,j) - 1 \leq \kappa'_{\text{FR}}(i,j) \leq \kappa_{\text{FR}}(i,j) + 2$$

The case of **edge deletion** can be handled similarly. Again, we need only consider the edges in E_{ij}, as defined in the proof above, and can make the same case argument. *Case 1*: WLOG assume the edge is of the form $(i*, v) \in E_{ij}$. Clearly, $d'_i = d_i - 1$. As for $|\#'_{\Delta_{iv}}|$, this can maximally be decreased by 1. *Case 2*: Degree and number of triangles do not change in response to the perturbation. Thus the following bounds hold for κ'_{FR}:

$$\kappa_{\text{FR}}(i,j) - 2 \leq \kappa'_{\text{FR}}(i,j) \leq \kappa_{\text{FR}}(i,j) + 1$$

□

A.1.2 Ollivier–Ricci Curvature

The definition of κ_{OR} establishes a relationship between the graph metric d_G, the Wasserstein distance W_1, the probability distributions μ_i, μ_j at nodes i, j and the curvature. Given that we are considering unweighted, and connected graphs we know that (V, d_G) is a well-defined metric space and therefore W_1 (as defined in [54]) defines the L_1 transportation distance between two probability measures μ_i, μ_j with respect to the metric d_G. This is relevant for a much larger class of graph metrics than just the standard choice of the shortest path distance. We use results from [54] and the metric properties of W_1 and d_G on graphs to bound the potential changes in κ_{OR} following an edge perturbation.

Lemma 1. *Consider the triple $\mathcal{G} = (G, d_G, \mu)$. Let δ_i denote the Dirac measure at node i and $J(i)$:= $W_1(\delta_i, \mu_i)$ the corresponding jump probability in the graph G. The Ollivier–Ricci curvature $\kappa_{OR}(i,j)$ satisfies the following Bonnet-Myers inspired upper bound:*

$$\kappa_{OR}(i,j) \leq \frac{J(i) + J(j)}{d_G(i,j)} \tag{7}$$

Proof. Rearranging the original definition for OR curvature gives:

$$W_1(\mu_i, \mu_j) = d_G(i,j)(1 - \kappa_{\text{OR}}(i,j))$$

By definition of the W_1, we have $d_G(i,j) = W_1(\delta_i, \delta_j)$. Using this and the fact that W_1 satisfies the triangle inequality property, we can construct the desired upper bound on κ_{OR}:

$$\begin{aligned} d_G(i,j) &\leq W_1(\delta_i, \mu_i) + W_1(\mu_i, \mu_j) + W_1(\delta_j, \mu_j) \\ d_G(i,j) &\leq J(i) + d_G(i,j)(1 - \kappa_{\text{OR}}(i,j)) + J(j) \\ d_G(i,j)(1 - (1 - \kappa_{\text{OR}}(i,j))) &\leq J(i) + J(j) \\ \kappa_{\text{OR}}(i,j) &\leq \frac{J(i) + J(j)}{d_G(i,j)} \end{aligned}$$

□

Theorem 3. *Given a perturbation (either **edge addition** or **edge deletion**) producing $\mathcal{G}'$, the Ollivier–Ricci $\kappa'_{OR}(i,j)$ of a pair (i,j) can be bounded via*

$$1 - \frac{1}{d_{G'}(i,j)}\left[2W'_{\max} + W'_1(\mu_i, \mu_j)\right] \leq \kappa'_{OR}(i,j) \leq \frac{J'(i) + J'(j)}{d_{G'}(i,j)}, \tag{4}$$

where $J'(v) := W'_1(\delta_v, \mu'_v)$ refers to the new jump probabilities and $W'_{\max} := \max_{x \in V} W'_1(\mu_x, \mu'_x)$ denotes the maximal reaction to the perturbation (measured using the updated Wasserstein distance).

Proof. We first prove the *upper bound*. Given that G' is still connected, and both W_1' and $d_{G'}$ still satisfy the metric axioms, this result follows directly from Lemma Lemma 1. For proving the *lower bound*, recall from Section 3.1 that $\mathcal{G}' = (G', d_{G'}, \mu')$ specifies the behaviour of the new graph metric $d_{G'}$ and the and the updated probability measure μ' in response to the perturbation. Moreover, this defines a new Wasserstein distance W_1' and we will show that the maximum reaction (as evaluated by W_1') to the perturbation $W'_{\max} := \max_{x \in V} W_1'(\mu_x', \mu_x)$ can be used to express a general lower bound for OR curvature in the event of a perturbation. As per Eq. (2), we can define our curvature following the perturbation as:

$$\kappa'_{\text{OR}}(i,j) = 1 - \frac{1}{d_{G'}(i,j)} W_1'(\mu_i', \mu_j') \tag{8}$$

Once again, we can make use of the metric properties of W_1', to establish the lower bound as

$$\begin{aligned}\kappa'_{\text{OR}}(i,j) &\geq 1 - \frac{1}{d_{G'}(i,j)} \left[W_1'(\mu_i, \mu_i') + W_1'(\mu_j, \mu_j') + W_1'(\mu_i, \mu_j)\right] \\ &\geq \frac{1}{d_{G'}(i,j)} \left[2W'_{\max} + W_1'(\mu_i, \mu_j)\right].\end{aligned}$$

$\square$

A.1.3 Resistance Curvature

The resistance distance, intuitively, measures how well connected two nodes are in a graph. It is defined in [19] as:

$$R_{ij} := (\mathbf{e}_i - \mathbf{e}_j)^\intercal Q^\dagger (\mathbf{e}_i - \mathbf{e}_j) \tag{9}$$

Here Q is the normalized laplacian (weighted degrees on the diagonal, see [19]), $Q^\dagger$ the Moore-Penrose inverse, and $\mathbf{e}_i$ is i^{th} unit vector. This is the main feature that will be studied to understand the stability of the curvature measure, and can be computed for any two nodes in a connected component of a graph.

A brief aside regarding the practice of inverting edge weights: The common practice when computing effective resistance is to invert the edge weights of a graph in order to get a resistance. Given the spirit of resistance from circuit theory, we know that a high resistance should make it difficult for current to pass between nodes. Analogously when thinking about our graph as a markov chain, this would correspond to a low transition probability. So, if we think about our edge weights as coming from some kernel where higher similarity results in a higher edge weight, then we should definitely invert our edge weights to get to resistance. However, in the case that our edge weights represent the cost of travelling between nodes, then this is a suitable proxy for resistance in which case inverting the nodes is unnecessary. In order to achieve the theoretical properties of curvature with well known examples described in [19], we *do not* invert the edge weights in our experiments. Which means that the curvature itself interprets the edge weights themselves as a cost/resistance; I think is an important point to specify especially given the similarity to markov chains and the borrowed terminology from circuit theory.

Recalling the equations for node resistance curvature and resistance curvature, i.e. Eq. (3), it becomes clear that the main task is to understand how the resistance distance changes in response to perturbations. The results below from [45], are crucial for our proofs. Let $C(i,j)$ be the commute time between nodes $i, j \in V$. It is important to note that these results depend on the *normalized Laplacian*, defined in [45] as $N = D^{\frac{1}{2}} A D^{\frac{1}{2}}$, with eigenvalues λ_i, ordered such that $\lambda_1 \geq \lambda_2 \geq ....$ Here, D is the diagonal matrix with inverse degrees and A the adjancecy matrix. Also, as is consistent with the rest of the paper, assume our graph has n nodes and m edges, and d_i is the degree at node $i \in V$.

Proposition 1. *For a graph G, let $N = D^{\frac{1}{2}} A D^{\frac{1}{2}}$ be the* normalized Laplacian *with eigen values $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Then, the commute time in G between nodes i, j is subject to the following bounds:*

$$m(\frac{1}{d_s} + \frac{1}{d_t)}) \leq C(i,j) \leq \frac{2m}{1 - \lambda_2}(\frac{1}{d_s)} + \frac{1}{d_t}) \tag{10}$$

Proposition 2. *Consider the unweighted graph G, where each edge represents a* unit resistance*, i.e we consider each edge in the graph to be artificially weighted with value 1. Then the following equality holds for the commute time between nodes i, j:*

$$C(i,j) = 2mR_{ij} \tag{11}$$

Proposition 3. *If G' arises from a graph G by adding a new edge, then the commute time $C'(i,j)$ between any two nodes in G' is bounded by:*

$$C'(i,j) \leq (1+\frac{1}{m})C(i,j) \tag{12}$$

For proofs of these propositions, we refer the reader to [45]. These results create a direct connection between commute times and resistance distance, and gives insight into how commute time reacts under edge addition, and we use them directly to generate our bounds for resistance curvature.

Theorem 4. *If G' is the graph generated by **edge addition**, then $\kappa_R' \geq \kappa_R$,with the following bound:*

$$|\kappa_R'(i,j) - \kappa_R(i,j)| \leq \frac{\Delta_{add}(d_i + d_j)}{R_{ij} - \Delta_{add}}, \tag{5}$$

where $\Delta_{add} := \max_{i,j \in V}\left(R_{ij} - \frac{1}{2}(\frac{1}{d_i+1} + \frac{1}{d_j+1})\right)$.

Proof. Let R'_{ij} be the resistance distance in G'. Likewise, let $C(i,j)$ be the commute distance in G between nodes i,j and $C'(i,j)$ be the commute time in G'. Then 12 and 11 ensure that R'_{ij} is bounded above, by the original resistance distance in G:

$$2(m+1)R'_{ij} \leq 2m(1+\frac{1}{m})R_{ij}$$
$$R'_{ij} \leq R_{ij}$$

This follows our intuition of resistance distance very well: with the addition of an edge nodes can only get more connected. 10 also gives a nice lower bound:

$$(m+1)(\frac{1}{d'_i} + \frac{1}{d'_j}) \leq C'(i,j)$$
$$\frac{1}{2}(\frac{1}{d'_i} + \frac{1}{d'_j}) \leq R'_{ij}$$

In the case that we are adding a single edge, it is often the case that $d'_x = d_x$. However, the nodes that are connected by the new edge, $(i*, j*) \in E' \setminus E$, increase such that $d'_{i*} = d_{i*} + 1$ and $d'_{j*} = d_{j*} + 1$. Thus, the following lower bound holds in general for R'_{ij} and we can remain agnostic to the precise location of the new edge:

$$\frac{1}{2}(\frac{1}{d_i+1} + \frac{1}{d_j+1}) \leq R'_{ij} \leq R_{ij} \tag{13}$$

And likewise, after adding p edges:

$$\frac{1}{2}(\frac{1}{d_i+p} + \frac{1}{d_j+p}) \leq R^p_{ij} \leq R_{ij}$$

So the bounds of our 'perturbed' resistance distance R'_{ij} are determined by the initial network structure (R_{ij}) and the number connections each specific vertex has. Naturally, certain node pairs will be more strongly affected by the addition of an edge. We can define the maximum reaction to perturbation across pairs as follows:

$$\Delta_{add} := \max_{i,j \in V}\left(R_{ij} - \frac{1}{2}(\frac{1}{d_i+1} + \frac{1}{d_j+1})\right) \tag{14}$$

This can be used to bound node resistance curvature. In an unweighted graph, we have

$$p_i = 1 - \frac{1}{2}\sum_{j\sim i} R_{ij}$$

$$p'_i = 1 - \frac{1}{2}\sum_{j\sim i} R'_{ij}$$

For G and G' respectively. Given that resistance can only increase, p_i is clearly an lower bound for p'_i. Certainly a lower bound occurs when when the resistance between each one of i's neighbors maximally decreases. Thus we get the following inequality:

$$p_i \le p'_i \le p_i + \frac{d_i}{2}\Delta_{add}$$

Finally this gives the desired bound on κ'_{R}:

$$\kappa_{\mathrm{R}}(i,j) \le \kappa'_{\mathrm{R}}(i,j) \le \kappa_{\mathrm{R}}(i,j) + \frac{\Delta_{add}(d_i + d_j)}{R_{ij} - \Delta_{add}}$$

□

Theorem 7. *If G' is the graph generated by **edge deletion**, then $\kappa'_R \le \kappa_R$, bounded by:*

$$|\kappa'_R(i,j) - \kappa_R(i,j)| \le \frac{1}{R_{ij} + \Delta_{del}}\Big[\frac{2}{R_{ij}}(2R_{ij} + \Delta_{del})(p_i + p_j) - \Delta_{del}(d_i + d_j)\Big],$$

where $\Delta_{del} = \frac{2}{1-\lambda_2} - \min_{i,j\in V}(R_{ij})$ and λ_2 is the second largest eigenvalue of N.

Proof. Now we can beg the question of how effective resistance changes when we remove an edge. By inverting our initial argument in above proof of 4, we know that after removing an edge our resistance distance can only increase. Formally, $R_{ij} \le R'_{ij}$. For the upper bound, we can once again make an argument using 10, this time relying on the other half of the inequality. Here we need to also mention the normalized Laplacian N' for G', with eigenvalues $\lambda'_1 \ge \lambda'_2 \ge ... \ge \lambda'_n$.

$$C'(i,j) \le \frac{2(m-1)}{1-\lambda'_2}\left(\frac{1}{d'_i} + \frac{1}{d'_j}\right)$$

$$R'_{ij} \le \frac{1}{1-\lambda'_2}\left(\frac{1}{d'_i} + \frac{1}{d'_j}\right)$$

Again, we know that only the two unique vertices $(i*, j*)$ that shared an edge will have affected degrees, s.t $d'_{i*} = d_{i*} - 1$ and $d'_{j*} = d_{j*} - 1$. Moreover, from [28], we know that $\lambda_2 \ge \lambda'_2$. So we can bound the R'_{ij} as follows:

$$R_{ij} \le R'_{ij} \le \frac{2}{1-\lambda_2} \tag{15}$$

In fact, this applies to any number of edge deletions, as long as G' stays connected. Again, we can define a maximum possible change in resistance distance across the graph:

$$\Delta_{del} = \max_{i,j\in V}\left(\frac{2}{1-\lambda_2} - R_{ij}\right) = \frac{2}{1-\lambda_2} - \min_{i,j\in V}(R_{ij}) \tag{16}$$

This leads to the following bounds on node and edge curvature, and completes the proof:

$$p_i - \frac{d_i}{2}\Delta_{del} \le p'_i \le p_i$$

$$(1-\lambda_2)\big[p_i + p_j - \frac{\Delta_{del}}{2}(d_i + d_j)\big] \le \kappa'_{\mathrm{R}}(i,j) \le \kappa_{\mathrm{R}}(i,j)$$

$$\kappa_{\mathrm{R}}(i,j) - \frac{1}{R_{ij} + \Delta_{del}}\big[\frac{2}{R_{ij}}(2R_{ij} + \Delta_{del})(p_i + p_j) - \Delta_{del}(d_i + d_j)\big] \le \kappa'_{\mathrm{R}}(i,j) \le \kappa_{\mathrm{R}}(i,j)$$

□

A.2 Expressivity Proofs

Theorem 5. *Given two graphs $F = (V_F, E_F)$ and $G = (V_G, E_G)$ with scalar-valued filtration functions f, g, and their respective persistence diagrams D_f, D_g, we have $d_B(D_f, D_g) \geq \inf_{\eta\colon E_F \to E_G} \sup_{x \in E_F} |f(x) - g(\eta(x))|$, where η ranges over all maps from E_F to E_G.*

Proof. Considering the calculation of persistence diagrams based on scalar-valued filtrations functions, every point in the persistence diagram D_f can be written as a tuple of the form $(f(e_F), f(e'_F))$, with $e_F, e'_F \in E_F$; the sample applies for D_g. The inner distance between such tuples that occur in the bottleneck distance calculation can thus be written as

$$\|(f(e_F), f(e'_F)) - (g(e_G), g(e'_G))\|_\infty, \tag{17}$$

which we can rewrite to $\max_{C\colon E_F \to E_G}\{f(x) - g(C(x))\}$ for a general map C induced by the bijection of the bottleneck distance. Not every map is induced by a bijection, though. Hence, if we maximise over *arbitrary* maps between the edge sets, we are guaranteed to never exceed the bottleneck distance. □

B Additional Proofs for Distinguishing Strongly Regular Graphs

Theorem 6 (Expressivity of curvature notions)**.** *Both Forman–Ricci curvature and Resistance curvature* cannot *distinguish distance-regular graphs with the same intersection array, whereas Ollivier–Ricci curvature* can *distinguish the Rook and Shrikhande graphs, which are strongly-regular graphs with the same intersection array.*

Proof. We first show the part of the statement relating to the *Forman–Ricci curvature*. Given a distance-regular graph G with N vertices and intersection array $\{b_0, b_1, \ldots, b_{D-1}; c_1, c_2, \ldots, c_D\}$. Let i, j be adjacent nodes in G. For a regular graph, we have $d_i = d_j = b_0$, where b_0 is a constant. The number of triangles between two adjacent nodes i and j in G is given by $a_1 = b_0 - b_1 - c_1$ [18]. The Forman curvature of i, j is thus

$$\kappa_{\mathrm{FR}}(i, j) := 4 - 2b_0 + 3|b_0 - b_1 - c_1|. \tag{18}$$

Given two strongly-regular graphs with the same intersection array, i.e. the same values of b_0, b_1 and c_1, the Forman curvature yields the same value for all pairs of adjacent nodes and cannot distinguish them. For the *resistance curvature*, the claim follows as an immediate Corollary of Theorem A [6] and described in Koolen et al. [40]. Given the resistance between two nodes depends only on the intersection array and the number of nodes in the graph, then the resistance curvature cannot distinguish two strongly-regular graphs.

The expressivity of Ollivier–Ricci curvature is strictly better, and it turns out that there are graphs with the same intersection array that we can distinguish, namely the Rook graph and the Shrikhande graph. Both graphs have the same intersection array $\{6, 3; 1, 2\}$ but differ in their first hop peripheral subgraphs [22]. It is known that 2-WL cannot distinguish these graphs. Ollivier–Ricci curvature, however, is sensitive to these differences in peripheral subgraphs with the edge curvatures for the Rook graph being: $[0.2, 0.2, 0.33, 0.33, 0.33, 0.2, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.2, 0.2, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.2, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33]$, and for the Shrikhande graph they are $[0, 0, 0.27, 0.27, 0.1, 0, 0.27, 0.27, 0.1, 0, 0.27, 0.1, 0.27, 0, 0.27, 0.1, 0.27, 0, 0.1, 0.27, 0.27, 0.1, 0.27, 0.27, 0.17, 0.17, 0.17, 0.17, 0.17, 0.17, 0.17, 0.17, 0.17, 0.17, 0.17, 0.17, 0.17, 0.17, 0.17, 0.17, 0.17, 0.17]$, demonstrating that OR curvature can distinguish these graphs—*unlike Resistance curvature, Forman–Ricci curvature and the* 2*-WL test.* □

C Additional Stability Analysis

Given the bounds on curvature established in Section 3.1, we explore how curvature changes experimentally by analysing edge perturbations on Erdős–Rényi graphs. In particular, we provide statistics that quantify the maximal change in curvature for random graphs with varying connectivity parameters in response to edge additions and deletions. The experiment fixes the number of nodes in the ER graphs ($n = 100$), and generates a sample of 50 graphs for the selected values of p. For each graph in the sample, we measure the curvature κ of all edges and calculate the standard deviation σ_κ of this distribution. We then perturb the original graph by edge addition/deletion and calculate the new curvature κ'. The following tables present the *worst case* deviations in curvature, which we define as $\Delta\kappa = |\kappa - \kappa'|$, in units of σ_κ; in other words the maximal value of $\Delta\kappa/\sigma_\kappa$ over all sample graphs and their edges.

	Edge Addition: Maximal Change (↓) in Curvature for ER Graphs ($\Delta\kappa/\sigma_\kappa$)								
Curvature	p=0.1	p=0.2	p=0.3	p=0.4	p=0.5	p=0.6	p=0.7	p=0.8	p=0.9
κ_{FR}	0.582	0.419	0.334	0.296	0.253	0.246	0.232	0.25	0.315
κ_{OR}	1.545	1.157	0.613	0.465	0.396	0.366	0.368	0.399	0.512
κ_{R}	0.689	0.417	0.296	0.251	0.221	0.232	0.227	0.243	0.321

	Edge Deletion: Maximal Change (↓) in Curvature for ER Graphs ($\Delta\kappa/\sigma_\kappa$)								
Curvature	p=0.1	p=0.2	p=0.3	p=0.4	p=0.5	p=0.6	p=0.7	p=0.8	p=0.9
κ_{FR}	0.609	0.408	0.334	0.291	0.255	0.239	0.245	0.243	0.319
κ_{OR}	1.397	1.27	0.623	0.479	0.394	0.365	0.347	0.402	0.482
κ_{R}	0.75	0.431	0.336	0.248	0.229	0.218	0.225	0.242	0.312

D Additional Commentary on Counting Substructures

We find that the difference in perspective between the selected curvature notions is underscored by their respective performance when counting substructures. *Forman curvature* is an inherently local measure by definition, depending only on 3-cycles between adjacent nodes and their degrees. *Ollivier–Ricci curvature*, when used with a uniform measure, can bound the number of triangles within a locally finite graph [36] through its relation with the Watts–Strogatz clustering coefficient [65]. It can also be shown that quadrangles and pentagons influence the OR curvature, further enhancing the expressivity of this type of curvature [36].

This is the most global perspective one can achieve using κ_{OR} with uniform probability measures, since polygons with more than five edges do not impact the curvature valuation.

However, by changing the probability measure used by κ_{OR}, we can shift the focus towards even larger substructures. For example, the nth power of the transition matrix provides information on the number of n-paths and can therefore provide substructure information for cycles of size n [42]. *Resistance curvature*, by contrast, is biased towards the largest substructures. Due to the 'global' nature of the resistance distance metric, κ_{R} assigns cycles of size ≥ 5 a positive curvature. Moreover, in a locally finite graph, one cannot use κ_{R} to establish a non-trivial bound on the number of triangles (consider creating an infinite cycle between two nodes).

E Probability Measure for Ollivier–Ricci Curvature and Counting Substructures

The Ollivier–Ricci curvature is of particular interest because of its flexibility. While the predominant probability measure μ used by the community is *uniform* for each node, i.e. each of the node's neighbours is chosen with probability being proportional to the degree of the node. We experimented with different probability measures, one being based on expanding μ to the two-hop neighbourhood of a vertex, the other one being based on random walk probabilities. Specifically, for a node x and a positive integer m, we calculate μ_{RW} as

$$\mu_{\mathrm{RW}}(y) := \sum_{k \leq m} \phi_k(x, y), \tag{19}$$

with $\phi_k(x, y)$ denoting the probability of reaching node y in a k-step random walk that starts from node x. Subsequently, we normalise Eq. (19) to ensure that it is a valid probability distribution. In our experiments, we set $m = 2$, meaning that at most 2-step random walks will be considered. As shown in the main paper, this formulation leads to an increase in expressivity, and we expect that further exploration of the probability measures will be a fruitful direction for the future.

We now explore to what extent the ability of the curvature to count substructures can also be improved in this way. To do this, we used powers of the transition matrix as the probability measure, as it has been shown that the nth power provides information on the number of n-paths and can therefore provide substructure information for cycles of size n [42]. We find that powers of the transition matrix larger than 1 can perform better for counting the substructures, particularly for substructures larger than 3-cycles. There is also a difference between Regular and Erdős–Rényi (ER) graphs as the best transition power tends to be higher for ER graphs. We hypthosise that this may have something to do with the mixing time of the graph, as large powers should converge to the stationary distribution, and regular graphs are more 'expander-like'. The best results are obtained by taking multiple landscapes using the transition matrix powers (up to $n = 5$) and then averaging them. We show that combined with a single layer MLP, this method can perform better than using Graph Neural Network based approaches and OR curvature with the uniform measure.

Method	Counting Substructures (MAE ↓)			
	Triangle	Tailed Tri.	Star	4-Cycle
GCN	0.4186	0.3248	**0.1798**	0.2822
κ_{OR} Filtration	0.2321	0.2395	0.3393	0.3089
κ_{OR} Filtration with transition matrix powers	**0.1956**	**0.2095**	0.3212	**0.2680**

Method	Optimal Transition Power	
	ER	Regular
Triangle	2	1
Tailed Triangle	4	3
Star	4	2
Chordal Cycle	2	2
4-Cycle	8	3

F Computational Complexity

Persistence diagrams of 1-dimensional simplicial complexes, i.e. graphs, can be computed in $\mathcal{O}(m \log m)$ time where m denotes the number of edges. Empirically, when calculating different curvature measures for different sizes of graphs, we find that Forman curvature scales well to large graphs, whereas OR and resistance curvatures can be used for smaller graphs and in cases that require a more expressive measure. Note that there are significantly faster ways to calculate resistance curvature as an approximation [64]. A majority of works on GGMs focus on small molecule generation, where any of these curvatures can be used with minimal pre-computation. Table 4 depicts the computational complexity of various curvature calculations on Erdős–Rényi graphs whilst Table 5 and Table 6 compares the complexity to methods based on MMD. We find that calculating persistence diagrams, turning these to persistence landscapes, averaging these and then calculating a distance takes a similar amount of time compared to MMD for different sizes of graphs and for different numbers of graphs in the reference set. Interestingly, our approach scales better than MMD as both the number of graphs in the reference set increases and when the size of the graphs increases. This will be important for comparing distributions of large data sets such as the commonly used Zinc dataset or QM9. Overall, we find that our method can be easily applied in practical use cases, especially given that models for graph generation tyically generate graphs with well under 1000 nodes.

Table 4: Computation time in seconds for discrete curvature on varying Erdős–Rényi graph sizes with $p = 0.3$.

No. nodes	κ_{FR}	κ_{OR}	κ_{R}
10	0.000	0.002	0.020
50	0.001	0.038	0.700
100	0.005	0.247	6.610
250	0.054	4.720	252.850
500	0.380	59.270	6414.970
1000	2.920	1040.700	74 366.070

Table 5: Computation time for different number of Erdős–Rényi graphs in reference set ($n = 10$ and $p = 0.3$.) with different distribution distance measures

Number of Graphs	Degree + MMD		Orbit + MMD		Curvature + MMD		Curvature + Landscapes	
10	2.2	ms	142.0	ms	2.4	ms	9.0	ms
20	3.6	ms	217.0	ms	4.2	ms	12.5	ms
50	10.9	ms	459.0	ms	12.4	ms	20.9	ms
100	34.1	ms	887.0	ms	37.9	ms	34.4	ms
200	120.0	ms	1960.0	ms	133.0	ms	80.4	ms
500	678.0	ms	6740.0	ms	727.0	ms	144.0	ms
1000	2620.0	ms	19 900.0	ms	2680.0	ms	359.0	ms

Table 6: Computation time for fixed number of Erdős–Rényi graphs in reference set with different sizes ($p = 0.3$.) with different distribution distance measures

Number of Graphs	Curvature + MMD		Curvature + Landscapes	
10	2.3	ms	9.0	ms
20	4.5	ms	12.5	ms
50	15.8	ms	20.9	ms
100	93.6	ms	34.4	ms
200	556.0	ms	80.4	ms
500	727.0	ms	144.0	ms
1000	2800.0	ms	364.0	ms

G Ethical Concerns

We have proposed a general framework for comparing graph distributions focusing primarily on method and theoretical development rather than on potential applications. We currently view drug discovery as being one of the main application areas, where further experiments may be required, but we have no evidence that our method enhances biases or causes harm in any way.