SIMILARITY AS A SIGNAL: COMPARING REPRESENTATIONS ACROSS MULTIPLE SCALES

Anonymous authors
Paper under double-blind review

ABSTRACT

Representational similarity measures provide a quantitative assessment of alignment between representations learned by different models, modalities, and training procedures. Existing measures typically characterize similarity at a single predefined scale, although different aspects of alignment may be reflected in structure ranging from local neighborhoods to broader global organization. Intuitively, local neighborhood structure may be important for retrieval or instance-level matching, whereas broader geometric organization may capture higher-level semantic relationships. We introduce a multiscale framework that treats representational similarity as a function of scale rather than as a single scalar. Motivated by the connection between RBF kernels and heat diffusion, we use the RBF bandwidth as the scale parameter and introduce the *multiscale similarity signal* to characterize representational alignment across scales. On controlled synthetic data, we show that perturbations at local, intermediate, and global scales produce distinct signatures in the signal. Moreover, on learned neural representations, the signal characterizes more complex transformations spanning a broad range of scales. Finally, we derive scalar scores by aggregating similarity across scales and evaluate them on the ReSi benchmark spanning graph, vision, and language domains, where our primary multiscale score achieves the best 90th-percentile rank in all three domains. Our results show that viewing representational similarity across multiple scales provides both an interpretable tool for analyzing representation geometry and an effective framework for quantifying representational similarity.

1 INTRODUCTION

Alignment across representations is a well-established mechanism in human learning, enabling learners to identify corresponding relational structure across examples and sensory modalities and generalize it to new contexts (Gentner & Medina, 1998; Christie & Gentner, 2010; Gentner & Hoyos, 2017; Aho et al., 2023). A similar principle has become important in machine learning, where representational alignment enables information to be transferred across models, modalities, and tasks (Sucholutsky et al., 2023). This principle underlies cross-modal systems such as CLIP (Radford et al., 2021), similarity-preserving knowledge distillation (Tung & Mori, 2019), and generative modeling methods such as REPA (Yu et al., 2025). As alignment becomes an objective and a mechanism in modern learning systems, a fundamental question arises: how can the degree of alignment between representations learned by different models, modalities, or training procedures be quantified? This question is particularly relevant for foundation models, whose representations are reused across tasks and application domains, including robotics (Driess et al., 2023), electronic health records (Steinberg et al., 2021), and medical imaging (Ma et al., 2024). Understanding whether and how such representations share structure therefore requires reliable tools for comparing their geometry.

Numerous measures of representational similarity have been proposed (Sucholutsky et al., 2023). Among them, centered kernel alignment (CKA) has become one of the most widely used (Kornblith et al., 2019). Given two representations of the same samples, CKA constructs pairwise kernel matrices and measures their normalized alignment. With the commonly used linear kernel, CKA summarizes shared global second-order structure. It has been used to compare convolutional and vision transformer architectures (Raghu et al., 2021), investigate the effects of network width and depth (Nguyen et al., 2021), and study training-related phenomena such as catastrophic forgetting (Ramasesh et al., 2021) and representational consistency across training objectives (Ciernik et al.,

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.

2025). Recent work, however, has identified limitations of this global characterization: CKA can be sensitive to transformations applied to subsets of samples (Davari et al., 2023) and can exhibit nonzero similarity between independent high-dimensional representations (Murphy et al., 2024; Gröger et al., 2026). A complementary perspective compares local neighborhood structure, for example through mutual k-nearest-neighbor overlap. These local and global approaches operate at different scales, where scale reflects the extent of representation structure captured by the comparison. The contrast between local and global measures became apparent in work on the Platonic Representation Hypothesis (Huh et al., 2024), which found weak convergence trends with CKA but substantially stronger local neighborhood agreement. Local and global approaches therefore capture complementary aspects of representational alignment, emphasizing neighborhood-level structure and broader organization, respectively. These findings suggest that representational alignment can vary substantially across scales, yet existing measures typically evaluate similarity at a single predefined scale.

In this work, we introduce a multiscale framework for representational similarity that treats scale as a variable. Motivated by the connection between radial basis function (RBF) kernels and heat diffusion, we use RBF bandwidth as the scale parameter and define the *multiscale similarity signal* to measure alignment across scales. We apply the framework to CKA and introduce uncentered kernel alignment (UKA) as a complementary signal construction, while showing how the formulation naturally extends to other kernel-based operators. The similarity signals characterize and visualize how alignment varies across scales, while their normalized area under the curve provides similarity scores for comparing representations. We demonstrate both uses in controlled, learned, and benchmark settings. On synthetic data, we show that perturbations of local, intermediate, and global structure affect distinct regions of the signal, revealing changes obscured by single-scale measures. We then analyze the gLocal transformation suggested by Muttenthaler et al. (2023), which aligns global representational structure with human similarity judgments while balancing preservation of the original local geometry. The multiscale signal characterizes this more complex transformation, whose effect spans a broad range of scales. Finally, we evaluate the similarity scores on the Representational Similarity (ReSi) benchmark (Klabunde et al., 2025b), spanning graph, vision, and language domains. While the original ReSi evaluation found no single measure, nor even a broader family of measures, that performed best across domains, our multiscale scores achieve the best mean rank in each domain, while our primary score achieves the best 90th-percentile rank in all three.

2 Preliminaries

Learned representations often exhibit structured, locally low-dimensional geometry, with samples associated with common objects, concepts, or task variables forming manifold-like regions (Cohen et al., 2020; Chung, 2021; Chou et al., 2025). The *union-of-manifolds hypothesis* formalizes this view by modeling a representation as a collection of such regions with potentially different dimensions and geometries (Brown et al., 2023; Wang & Wang, 2024). This naturally motivates a multiscale description: local neighborhoods characterize within-manifold geometry, while broader scales capture the relative organization of manifolds across the representation space. We next review the connection between RBF kernels and heat diffusion, followed by a formulation of the representational similarity problem and a review of notable similarity measures.

2.1 Heat Kernels and Manifold Learning

A d-dimensional Riemannian manifold $\mathcal{M} \subset \mathbb{R}^p$ is a space that is locally Euclidean, with its intrinsic geometry specified by a Riemannian metric. We denote by μ a probability measure on $\mathcal{M}$ describing the data distribution. A common assumption in machine learning is that data concentrates near a low-dimensional manifold, known as the manifold assumption (Belkin et al., 2006). Manifold learning exploits this assumption by using local ambient-space distances to characterize the underlying geometry. We consider the RBF kernel on $\mathbb{R}^p$, restricted to points $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{M}$,

$$k_\sigma^{\mathbb{R}^p}(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\|\boldsymbol{x} - \boldsymbol{x}'\|^2 / 2\sigma^2\right), \tag{1}$$

where the bandwidth $\sigma > 0$ controls the scale. For any real-valued function $f : \mathcal{M} \to \mathbb{R}$, the kernel induces the following integral operator

$$(\mathcal{K}_\sigma f)(\boldsymbol{x}) = \int_{\mathcal{M}} k_\sigma^{\mathbb{R}^p}(\boldsymbol{x}, \boldsymbol{x}') f(\boldsymbol{x}')\, d\mu(\boldsymbol{x}'), \tag{2}$$

and the operator family $\{\mathcal{K}_\sigma\}_{\sigma>0}$ provides a scale-dependent description of the data geometry. This scale interpretation is closely related to heat diffusion. The heat kernel provides a natural multiscale description of intrinsic geometry, with diffusion time controlling the transition from local to broader structure. In p-dimensional Euclidean space, the heat kernel satisfies $h_t^{\mathbb{R}^p}(\boldsymbol{x}, \boldsymbol{x}') = (4\pi t)^{-p/2} k^{\mathbb{R}^p}_{\sqrt{2t}}(\boldsymbol{x}, \boldsymbol{x}')$. For nearby points on a smooth manifold, ambient Euclidean distances locally approximate intrinsic distances, relating this RBF form to the short-time manifold heat kernel (Sun et al., 2009). Thus, the correspondence, $t = \sigma^2/2$, connects RBF bandwidth to diffusion time, motivating its interpretation as a scale parameter. The short-scale regime reflects local manifold geometry, while increasing σ progressively extends the comparison beyond this local regime; see Appendix A for further discussion.

2.2 Representational Similarity Measures

Representational similarity measures quantify alignment between representations using different notions of shared structure. Let $(\mathcal{Z}, \mu)$ denote the input probability space, and consider two representation maps $F_X : \mathcal{Z} \to \mathcal{X} \subset \mathbb{R}^{p_X}$, $F_Y : \mathcal{Z} \to \mathcal{Y} \subset \mathbb{R}^{p_Y}$. Given n realizations $\boldsymbol{z}_i \sim \mu$, we obtain paired representation outputs $\boldsymbol{x}_i = F_X(\boldsymbol{z}_i)$ and $\boldsymbol{y}_i = F_Y(\boldsymbol{z}_i)$, and denote the corresponding representation matrices by $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]^\top \in \mathbb{R}^{n \times p_X}$ and $\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n]^\top \in \mathbb{R}^{n \times p_Y}$.

Centered kernel alignment (CKA) compares representations through their kernel matrices (Kornblith et al., 2019). Given kernel functions $k^X(\cdot,\cdot)$ and $k^Y(\cdot,\cdot)$, define kernel matrices $\boldsymbol{K}^X, \boldsymbol{K}^Y \in \mathbb{R}^{n \times n}$ such that $[\boldsymbol{K}^X]_{ij} = k^X(\boldsymbol{x}_i, \boldsymbol{x}_j)$ and $[\boldsymbol{K}^Y]_{ij} = k^Y(\boldsymbol{y}_i, \boldsymbol{y}_j)$. For nonzero matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, define their Frobenius cosine similarity as $\cos_F(\boldsymbol{A}, \boldsymbol{B}) = \frac{\langle \boldsymbol{A}, \boldsymbol{B} \rangle_F}{\|\boldsymbol{A}\|_F \|\boldsymbol{B}\|_F}$, where $\langle \cdot, \cdot \rangle_F$ and $\| \cdot \|_F$ denote the Frobenius inner product and norm. CKA is then

$$\mathrm{CKA}(\boldsymbol{K}^X, \boldsymbol{K}^Y) = \cos_F(\boldsymbol{H}\boldsymbol{K}^X\boldsymbol{H}, \boldsymbol{H}\boldsymbol{K}^Y\boldsymbol{H}), \tag{3}$$

where $\boldsymbol{H} = \boldsymbol{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$ is the centering matrix. Equivalently, CKA is the normalized empirical Hilbert-Schmidt independence criterion (HSIC). CKA is commonly applied with the linear kernel, $k_{\mathrm{lin}}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i^\top \boldsymbol{x}_j$, yielding $\mathrm{CKA}_{\mathrm{lin}}(\boldsymbol{X}, \boldsymbol{Y}) = \mathrm{CKA}(\boldsymbol{X}\boldsymbol{X}^\top, \boldsymbol{Y}\boldsymbol{Y}^\top)$, which compares global second-order structure. CKA can also be applied with the RBF kernel of Equation 1. Throughout, σ denotes a relative bandwidth obtained by normalizing distances within each representation by its median pairwise distance, and we write $\mathrm{CKA}_{\mathrm{RBF},\sigma}(\boldsymbol{X}, \boldsymbol{Y}) = \mathrm{CKA}(\boldsymbol{K}^X_\sigma, \boldsymbol{K}^Y_\sigma)$. This normalization makes the measure invariant to isotropic rescaling of either representation and provides a common relative notion of scale. Unlike $\mathrm{CKA}_{\mathrm{lin}}$, $\mathrm{CKA}_{\mathrm{RBF},\sigma}$ depends explicitly on the scale σ.

A complementary approach compares local neighborhood structure. For the i-th sample, let $\mathcal{N}_k^X(i)$ and $\mathcal{N}_k^Y(i)$ denote its k nearest neighbors in $\boldsymbol{X}$ and $\boldsymbol{Y}$. Following Huh et al. (2024), mutual k-nearest-neighbor similarity is

$$\mathrm{MNN}_k(\boldsymbol{X}, \boldsymbol{Y}) = \frac{1}{n} \sum_{i=1}^{n} \frac{\left|\mathcal{N}_k^X(i) \cap \mathcal{N}_k^Y(i)\right|}{k}, \tag{4}$$

the average fraction of neighbors shared by the two representations, taking values in $[0, 1]$. Here, k controls the neighborhood scale, with smaller values emphasizing more local structure. Huh et al. (2024) also introduce CKNNA, which combines linear CKA with nearest-neighbor restriction and therefore incorporates both kernel and local neighborhood structure. This approach has recently gained popularity in several applications (Leng et al., 2025; Edamadaka et al., 2025; Bi et al., 2026).

Classical representational similarity measures have focused on global structure, while recent work has emphasized local alignment. These two contrasting perspectives motivate a multiscale view of representational similarity spanning the full range from local to global geometry.

3 Multiscale Representational Similarity

We propose to compare representations across a continuum of scales rather than at a single predefined scale. Our construction uses a family of kernel-based integral operators parametrized by bandwidth, which controls the scale of comparison. We first define the resulting similarity signal in the population setting and its finite-sample counterpart, and then introduce several signal constructions. Finally, we derive similarity scores from the signals and characterize their endpoint behavior to interpret these scores.

3.1 MULTISCALE SIMILARITY SIGNAL

We view a representation as inducing a geometry on the underlying data through pairwise distances, characterized at the population level by a kernel-based integral operator on functions over the common input space $\mathcal{Z}$. For any real-valued function $f : \mathcal{Z} \to \mathbb{R}$, define

$$(\mathcal{K}_\sigma^X f)(z) = \int_{\mathcal{Z}} k_\sigma^{\mathbb{R}^{p_X}} \big(F_X(z), F_X(z')\big) f(z')\, d\mu(z'), \tag{5}$$

and define $\mathcal{K}_\sigma^Y$ analogously. Intuitively, $(\mathcal{K}_\sigma^X f)(z)$ aggregates the values of f over inputs z', with weights determined by the distance between $F_X(z)$ and $F_X(z')$ in the representation space $\mathcal{X}$ at scale σ. If f is concentrated around z_0, then $\mathcal{K}_\sigma^X f$ reflects the scale-σ proximity pattern around $F_X(z_0)$. Although F_X and F_Y map into different spaces, their induced operators act on functions over the same input space $\mathcal{Z}$. This provides an operator-based view of representational similarity, in which similar representations are expected to induce similar operators on the input space $\mathcal{Z}$.

Definition 1. The population multiscale similarity signal between the representations induced by F_X and F_Y is the function

$$s_{\mathcal{J}}^{\mathcal{X},\mathcal{Y}} : (0,\infty) \to [0,1], \qquad s_{\mathcal{J}}^{\mathcal{X},\mathcal{Y}}(\sigma) = \mathcal{J}\left(\mathcal{K}_\sigma^X, \mathcal{K}_\sigma^Y\right), \tag{6}$$

where $\mathcal{J}$ denotes an operator similarity functional taking values in $[0,1]$ and invariant to positive rescaling of either argument.

Varying σ therefore describes how representational alignment changes across scales. This interpretation is particularly natural under the union-of-manifolds model (Brown et al., 2023) and is motivated by the connection between RBF kernels and heat diffusion mentioned in Section 2. At small scales, the kernel captures local manifold geometry; as the scale increases, it incorporates coarser manifold structure and the relative organization of different manifolds in the representation space.

In the finite-sample setting, we represent the population operators $\mathcal{K}_\sigma^X$ and $\mathcal{K}_\sigma^Y$ by the RBF kernel matrices $\boldsymbol{K}_\sigma^X$ and $\boldsymbol{K}_\sigma^Y$, whose entries are obtained by evaluating the corresponding kernels on the observed samples. As in Section 2.2, σ denotes a relative scale, with pairwise distances normalized by their median. Let J denote the finite-dimensional counterpart of $\mathcal{J}$ acting on the resulting kernel matrices. We define the *empirical multiscale similarity signal* by

$$s_J^{X,Y} : (0,\infty) \to [0,1], \qquad s_J^{X,Y}(\sigma) = J\left(\boldsymbol{K}_\sigma^X, \boldsymbol{K}_\sigma^Y\right), \tag{7}$$

which is the finite-sample analogue of Eq. 6. To simplify notation, we occasionally omit the superscript X, Y. We evaluate the similarity signal on a finite bandwidth grid $\sigma_1, \ldots, \sigma_m$, yielding

$$\boldsymbol{s}_J^{X,Y} = \left[s_J^{X,Y}(\sigma_1), \ldots, s_J^{X,Y}(\sigma_m)\right]^\top \in [0,1]^m. \tag{8}$$

We use geometrically spaced bandwidths, equivalent to uniform sampling in $\log(\sigma)$. This matches the relative scaling of the RBF kernel and provides finer resolution at smaller bandwidths. Since only the bandwidth varies across scales, pairwise distances are computed once and reused.

3.2 PROPOSED SIMILARITY SIGNALS

We construct several multiscale similarity signals by applying the Frobenius cosine similarity to different operators derived from the RBF kernel matrices. For symmetric positive semidefinite (SPSD) matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, $\cos_F(\boldsymbol{A}, \boldsymbol{B})$ is invariant to positive rescaling of either argument and takes values in $[0,1]$ as required for J.

CKA signal. CKA computes the similarity between centered RBF kernel matrices,

$$s_{\mathrm{CKA}}^{X,Y}(\sigma) = J_{\mathrm{CKA}}(\boldsymbol{K}_\sigma^X, \boldsymbol{K}_\sigma^Y) = \cos_F\left(\boldsymbol{H}\boldsymbol{K}_\sigma^X\boldsymbol{H}, \boldsymbol{H}\boldsymbol{K}_\sigma^Y\boldsymbol{H}\right) = \mathrm{CKA}_{\mathrm{RBF},\sigma}(\boldsymbol{X}, \boldsymbol{Y}). \tag{9}$$

Rather than selecting a single bandwidth as in standard CKA-RBF, the CKA signal follows the similarity trend across the full range of σ. To provide intuition for $s_{\mathrm{CKA}}(\sigma)$, observe the spectral form of $\cos_F$. Let $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$ and $\boldsymbol{B} = \boldsymbol{V}\boldsymbol{M}\boldsymbol{V}^\top$ be the spectral decompositions of two nonzero SPSD matrices, with eigenvalues λ_i, μ_j and eigenvectors $\boldsymbol{u}_i, \boldsymbol{v}_j$. Then

$$\cos_F(\boldsymbol{A}, \boldsymbol{B}) = \sum_{i,j} \tilde{\lambda}_i \tilde{\mu}_j (\boldsymbol{u}_i^\top \boldsymbol{v}_j)^2, \tag{10}$$

where $\tilde{\lambda}_i = \lambda_i / \left(\sum_k \lambda_k^2\right)^{1/2}$ and $\tilde{\mu}_j = \mu_j / \left(\sum_k \mu_k^2\right)^{1/2}$. For s_{CKA}, $\boldsymbol{A} = \boldsymbol{H}\boldsymbol{K}_\sigma^X\boldsymbol{H}$ and $\boldsymbol{B} = \boldsymbol{H}\boldsymbol{K}_\sigma^Y\boldsymbol{H}$, and their eigendecompositions correspond to kernel PCA of the two representations. Equation 10 therefore interprets CKA as a weighted alignment between their kernel principal modes: modes with large eigenvalues contribute more strongly, while $(\boldsymbol{u}_i^\top \boldsymbol{v}_j)^2$ measures the agreement between their principal directions. As σ varies, both the kernel principal modes and their relative importance change, so $s_{\mathrm{CKA}}(\sigma)$ tracks the agreement between the scale-dependent spectral geometries of the two representations. This interpretation also clarifies the effect of centering in CKA. Since $\boldsymbol{H}\boldsymbol{K}_\sigma^X\boldsymbol{H}\,\mathbf{1} = \mathbf{0}$, centering forces $\mathbf{1}$ to be an eigenvector with eigenvalue zero. Consequently, all nonzero spectral modes lie in the subspace orthogonal to $\mathbf{1}$, modifying the eigenspace on which the alignment is computed. This motivates a complementary construction that compares the uncentered kernel operators directly.

Uncentered Kernel Alignment (UKA) signal. Consider the uncentered alignment,

$$s_{\mathrm{UKA}}^{X,Y}(\sigma) = J_{\mathrm{UKA}}(\boldsymbol{K}_\sigma^X, \boldsymbol{K}_\sigma^Y) = \cos_F\left(\boldsymbol{K}_\sigma^X, \boldsymbol{K}_\sigma^Y\right), \tag{11}$$

which we denote uncentered kernel alignment (UKA). CKA and UKA use the same kernel family and alignment functional, differing only in whether centering is applied before the comparison.

Beyond CKA and UKA, the multiscale kernel formulation also makes it natural to introduce constructions from manifold learning into representational similarity. In particular, we normalize the kernels using the degree matrix associated with random walks on the data graph, reweighting affinities according to local connectivity and thereby reflecting local sample density (Nadler et al., 2006; Von Luxburg, 2007). Appendix D develops the resulting random walk kernel alignment (RWKA) and its degree-centered variant (dRWKA), which can be viewed as random-walk analogues of UKA and CKA, respectively. We report the empirical results obtained by s_{RWKA} and s_{dRWKA} alongside the results for s_{CKA} and s_{UKA} in Section 4.3.

3.3 From Similarity Signals to Similarity Scores

While the full signal is useful for diagnosing and visualizing similarity across scales, many applications require a single scalar score summarizing the similarity. We therefore aggregate the signal by averaging over the geometrically spaced bandwidths, which approximates the normalized area under the curve (AUC) on the log-bandwidth axis:

$$\mathrm{AUC}(\boldsymbol{s}_J^{X,Y}) = \frac{1}{m}\sum_{i=1}^{m} s_J^{X,Y}(\sigma_i) \approx \frac{1}{\log(\sigma_{\max}) - \log(\sigma_{\min})}\int_{\sigma_{\min}}^{\sigma_{\max}} s_J^{X,Y}(\sigma)\, d\log(\sigma). \tag{12}$$

The resulting score lies in $[0, 1]$ and gives equal weight to equal intervals in $\log(\sigma)$, or equivalently, to multiplicative changes in bandwidth. To understand what this aggregation captures, we characterize the endpoint behavior of the proposed signals, beginning with UKA.

Proposition 1. *For any finite paired samples $\boldsymbol{X}$ and $\boldsymbol{Y}$ with no duplicate representation vectors,*

$$\lim_{\sigma\to 0} s_{\mathrm{UKA}}^{X,Y}(\sigma) = 1, \qquad \lim_{\sigma\to\infty} s_{\mathrm{UKA}}^{X,Y}(\sigma) = 1. \tag{13}$$

At small bandwidths, $\boldsymbol{K}_\sigma \to \boldsymbol{I}$, so UKA compares two identity matrices and the similarity approaches one. At large bandwidths, $\boldsymbol{K}_\sigma \to \mathbf{1}\mathbf{1}^\top$, so the similarity again degenerates to one. Both endpoints are representation-independent and therefore cannot distinguish between representation pairs. As a result, differences in $\mathrm{AUC}(\boldsymbol{s}_{\mathrm{UKA}})$ are driven primarily by the nontrivial local-to-intermediate scales between these limits. CKA exhibits a different large-scale behavior.

Proposition 2. *For any finite paired samples $\boldsymbol{X}$ and $\boldsymbol{Y}$ with no duplicate representation vectors,*

$$\lim_{\sigma\to 0} s_{\mathrm{CKA}}^{X,Y}(\sigma) = 1, \qquad \lim_{\sigma\to\infty} s_{\mathrm{CKA}}^{X,Y}(\sigma) = \mathrm{CKA}_{\mathrm{lin}}(\boldsymbol{X}, \boldsymbol{Y}). \tag{14}$$

The small-bandwidth limit again follows from $\boldsymbol{K}_\sigma \to \boldsymbol{I}$, with the centered kernels converging to $\boldsymbol{H}$. At large bandwidths, centering removes the leading constant term of $\boldsymbol{K}_\sigma$, leaving the centered squared-distance structure, which is equivalent to the centered linear kernel. Accordingly, Alvarez (2022) showed that $s_{\mathrm{CKA}}(\sigma)$ converges to linear CKA as $\sigma \to \infty$. Unlike UKA, the large-scale endpoint of the CKA signal therefore remains representation-dependent. Its AUC summarizes multiscale representational agreement while retaining the contribution of this nontrivial large-scale limit. The same distinction carries over to the random-walk-based signals: in Appendix D.3, we prove that RWKA shares the endpoint behavior of UKA, whereas dRWKA shares that of CKA.

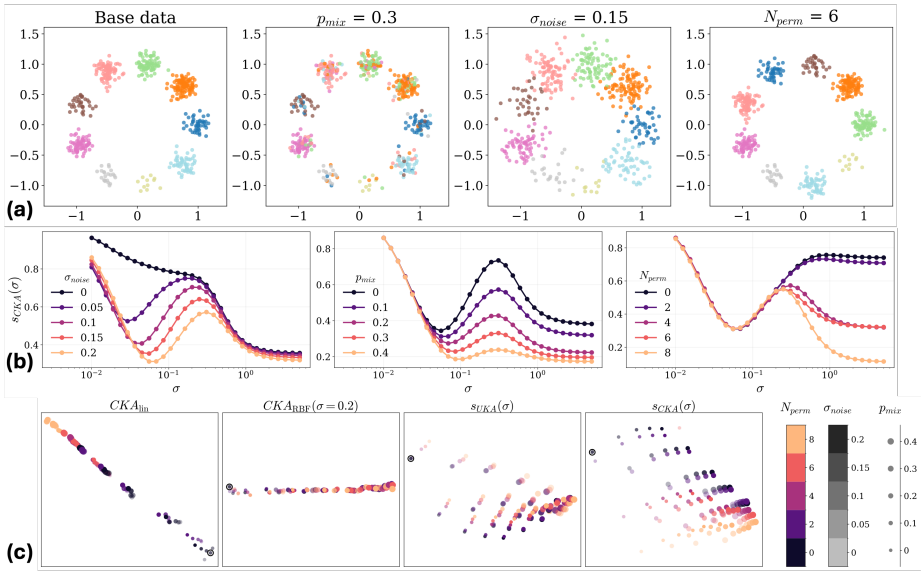


Figure 1: (a) Visualization of the base dataset and examples of the three transformations. (b) CKA-RBF similarity signals, $s_{\mathrm{CKA}}^{X_\theta, X_{ref}}(\sigma)$, for sweeps over the transformation parameters. (c) MDS embeddings of the similarity-to-base profiles over all parameter combinations, using $\mathrm{CKA}_{\mathrm{lin}}$ and $\mathrm{CKA}_{\mathrm{RBF},0.2}$, and the full signals $s_{\mathrm{UKA}}(\sigma)$ and $s_{\mathrm{CKA}}(\sigma)$. Point opacity, size, and color indicate σ_{noise}, p_{mix}, and N_{perm}, respectively and the base dataset is circled.

4 Experimental Results

We evaluate the multiscale framework in three complementary experiments. Controlled synthetic transformations first show that different regions of the signal capture local, intermediate, and global geometric changes, and demonstrate how the full signal can visualize representational variation across scales. We then apply the framework to gLocal (Muttenthaler et al., 2023) transformations of learned neural representations, illustrating its diagnostic use when transformations span multiple scales. Finally, we evaluate the AUC similarity score on the ReSi (Klabunde et al., 2025b) benchmark across Graph, Vision, and Language, where $\mathrm{AUC}(\boldsymbol{s}_{\mathrm{CKA}})$ is the only evaluated measure to achieve the best 90th-percentile rank in all three domains. Together, these experiments show that the multiscale approach provides both an interpretable diagnostic of representational change and a highly effective scalar representational similarity score. We provide code to reproduce our results on anonymous GitHub.

4.1 Synthetic Data

We first demonstrate that different regions of the multiscale similarity signal are sensitive to geometric changes occurring at different scales. To this end, we construct a synthetic dataset with an explicit hierarchical geometry and apply controlled transformations designed to primarily modify its local, intermediate, or global structure. The base dataset, shown in Figure 1a, consists of a mixture of nine isotropic Gaussian distributions in two dimensions, whose centers are evenly spaced on a circle of radius one. We draw $n = 500$ samples from this model where each center has a different probability and hence different number of samples. This construction contains several natural geometric scales: the arrangement of individual samples within each cluster, the cluster structure, and the global arrangement of the cluster centers. Starting from the same base dataset, we apply three transformations targeting progressively larger geometric scales, illustrated in Figure 1a:

1. **Local:** Add independent Gaussian noise, $\epsilon_i \sim \mathcal{N}(0, \sigma_{\mathrm{noise}}^2 I)$, perturbing fine-scale relationships while largely preserving the cluster structure.
2. **Intermediate:** Move each sample to a randomly selected cluster with probability p_{mix}, disrupting cluster membership while preserving the global arrangement of cluster centers.

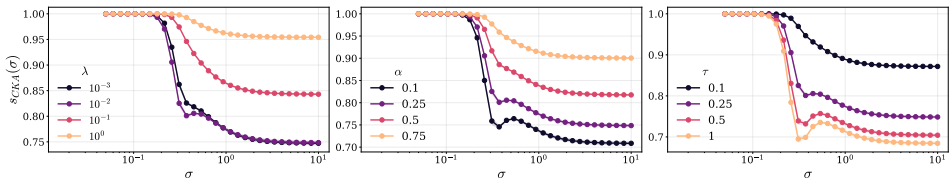


Figure 2: $s_{\mathrm{CKA}}^{X_{\boldsymbol{\phi}},X_0}$ similarity signals, for sweeps over the gLocal parameters λ, α, and τ. Curves show the mean over 10 CIFAR-100 batches, shading denotes one standard deviation across batches.

3. **Global:** Permute the positions of N_{perm} clusters, altering their global organization while preserving within-cluster geometry.

The transformations are therefore designed to have their dominant effects at progressively larger geometric scales. We create $5^3 = 125$ datasets for all parameter combinations $\sigma_{\mathrm{noise}} \in \{0, 0.05, 0.1, 0.15, 0.2\}$, $p_{\mathrm{mix}} \in \{0, 0.1, 0.2, 0.3, 0.4\}$, $N_{\mathrm{perm}} \in \{0, 2, 4, 6, 8\}$.

Let $\boldsymbol{\theta} = (\sigma_{\mathrm{noise}}, p_{\mathrm{mix}}, N_{\mathrm{perm}})$ denote the transformation parameters and $\boldsymbol{X}_{\boldsymbol{\theta}}$ the corresponding dataset. For the signal sweeps, we use the reference $\boldsymbol{\theta}_{\mathrm{ref}} = (0.2, 0, 4)$, with $\boldsymbol{X}_{\mathrm{ref}} = \boldsymbol{X}_{\boldsymbol{\theta}_{\mathrm{ref}}}$. Figure 1b shows $s_{\mathrm{CKA}}^{X_{\boldsymbol{\theta}},X_{\mathrm{ref}}}(\sigma)$ while varying one parameter at a time and fixing the others to their reference values. The signal is evaluated at 30 geometrically spaced bandwidths over $\sigma \in [10^{-2}, 5]$. Increasing σ_{noise} produces the largest separation at small bandwidths, p_{mix} primarily affects intermediate bandwidths, and N_{perm} produces the strongest differences at larger bandwidths. Thus, the region of maximal variation shifts toward larger bandwidths as the dominant geometric scale of the transformation increases. Appendix B.1 repeats the sweeps using the untransformed setting $\boldsymbol{\theta}_0 = (0, 0, 0)$ as reference.

We next compare the full signal with scalar similarity measures using all combinations of the transformation parameters. Taking $\boldsymbol{X}_0 = \boldsymbol{X}_{\boldsymbol{\theta}_0}$ as a common reference, each transformed dataset $\boldsymbol{X}_{\boldsymbol{\theta}}$ is represented by its similarity-to-base profile $\boldsymbol{s}_J^{X_{\boldsymbol{\theta}},X_0}$. For two settings $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$, we define $d_J(\boldsymbol{\theta}, \boldsymbol{\theta}') = \|\boldsymbol{s}_J^{X_{\boldsymbol{\theta}},X_0} - \boldsymbol{s}_J^{X_{\boldsymbol{\theta}'},X_0}\|_2$, with scalar measures reducing to the absolute difference between their similarity-to-base values. We visualize these pairwise dissimilarities using multidimensional scaling (MDS) (Kruskal, 1964), which embeds the transformed datasets in a low-dimensional space while approximately preserving their distances. Figure 1c shows the resulting two-dimensional embeddings for $\mathrm{CKA}_{\mathrm{lin}}$, $\mathrm{CKA}_{\mathrm{RBF},0.2}$, $\boldsymbol{s}_{\mathrm{UKA}}$, and $\boldsymbol{s}_{\mathrm{CKA}}$. Point color, size, and opacity encode the transformation parameters, and the base dataset is circled. For a scalar measure, the induced geometry is intrinsically one-dimensional, so the corresponding MDS embedding lies approximately along a line. The scalar measures nevertheless emphasize different transformations: $\mathrm{CKA}_{\mathrm{lin}}$ is particularly sensitive to N_{perm} and p_{mix}, whereas $\mathrm{CKA}_{\mathrm{RBF},0.2}$ is more sensitive to σ_{noise} and p_{mix}. In contrast, the multiscale signals retain variation across bandwidths and produce richer geometries in which different sources of change can be separated. In particular, $\boldsymbol{s}_{\mathrm{CKA}}$ separates local and global variation along distinct directions, whereas $\boldsymbol{s}_{\mathrm{UKA}}$ separates local and intermediate-scale variation.

4.2 Global-Local (gLocal) Transform

The synthetic transformations above are designed to act primarily at separate geometric scales. We apply our multiscale framework to analyze the gLocal transform of Muttenthaler et al. (2023), which learns to align neural network representations with human similarity judgments while preserving local neighborhood structure. This provides a practical setting in which geometric changes are not explicitly separated by scale. Given $\boldsymbol{x} \in \mathbb{R}^p$, gLocal learns an affine transformation, $\boldsymbol{y} = \boldsymbol{W}\boldsymbol{x} + \boldsymbol{b}$, by solving

$$\arg\min_{\boldsymbol{W},\boldsymbol{b}} (1-\alpha)\mathcal{L}_{\mathrm{global}}(\boldsymbol{W}, \boldsymbol{b}) + \alpha\mathcal{L}_{\mathrm{local}}(\boldsymbol{W}, \boldsymbol{b}; \tau) + \lambda \|\boldsymbol{W} - (\mathrm{tr}(\boldsymbol{W})/p)\, \boldsymbol{I}\|_F^2 . \tag{15}$$

The global loss is the negative log-likelihood of human odd-one-out judgments collected on the THINGS dataset (Hebart et al., 2019), encouraging the transformed representation to align with human similarity judgments. The local loss, evaluated on ImageNet images (Deng et al., 2009), preserves local structure by matching softmax distributions of pairwise cosine similarities, with

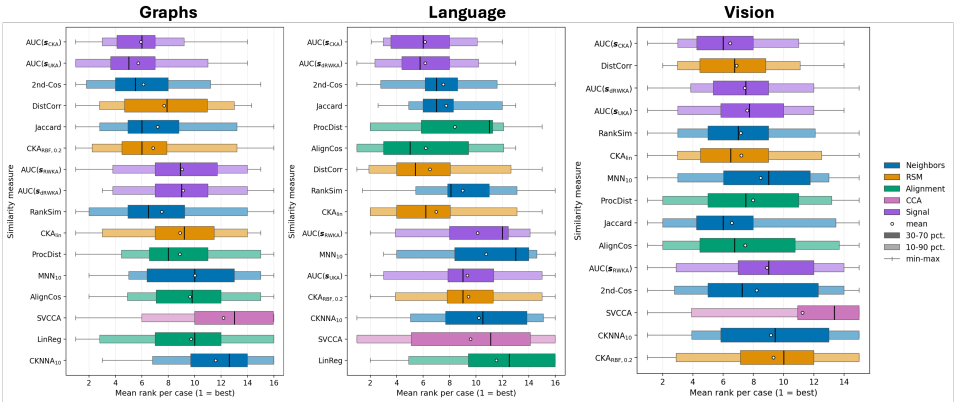


Figure 3: Per-domain ReSi performance. Distributions show within-case ranks, with lower values indicating better performance. Nested intervals denote the 10th-90th and 30th-70th percentiles; the black line marks the median, the white circle the mean, and whiskers the minimum and maximum. Measures are ordered by 90th-percentile rank.

temperature τ setting the locality scale. The parameter $\alpha \in [0, 1]$ balances the global and local objectives, while λ regularizes the transformation toward a scaled identity.

Following Muttenthaler et al. (2023), we train the gLocal transformation with SGD for 10 epochs with early stopping and a learning rate of 0.001. We train one transformation for each combination of $\lambda \in \{10^{-3}, 10^{-2}, 10^{-1}, 1\}$, $\alpha \in \{0.1, 0.25, 0.5, 0.75\}$, $\tau \in \{0.1, 0.25, 0.5, 1\}$.

Let $\boldsymbol{\phi} = (\lambda, \alpha, \tau)$ denote the gLocal parameters, $\boldsymbol{X}_0$ the original representation, and $\boldsymbol{X}_{\boldsymbol{\phi}}$ the corresponding transformed representation. Figure 2 shows $s_{\text{CKA}}^{X_{\boldsymbol{\phi}}, X_0}(\sigma)$, comparing each gLocal-transformed representation to the original representation $\boldsymbol{X}_0$. We vary one of λ, α, and τ at a time while fixing the remaining two to the reference setting $\boldsymbol{\phi}_{\text{ref}} = (10^{-2}, 0.25, 0.25)$. The signal is evaluated at 30 geometrically spaced bandwidths over $\sigma \in [5 \cdot 10^{-2}, 10]$, with larger bandwidth values than in Section 4.1 to accommodate the larger pairwise distances typical of high-dimensional representations. We evaluate OpenCLIP-ViT-L-14-2B representations on CIFAR-100 and average each signal over 10 batches of 1024 samples. Unlike the controlled transformations in Section 4.1, the effects of the gLocal parameters are not confined to distinct bandwidth ranges but extend across multiple scales. Increasing λ raises similarity across scales as stronger regularization restrains the transformation, while increasing α raises similarity most strongly at smaller bandwidths, reflecting stronger local preservation. In the τ sweep, smaller values yield higher similarity across scales, consistent with greater preservation of the original representation when the local and global objectives emphasize more distinct scales. As τ increases and their scales overlap, the representation changes more strongly rather than being further constrained by the local loss. Appendix B.2 presents results for additional architectures and an MDS analysis analogous to Figure 1c.

These results extend the controlled observations of Section 4.1 to learned neural representations. Even when geometric changes are not confined to a single scale, the multiscale signal reveals how a learned transformation affects representation geometry across scales, highlighting its potential as a diagnostic and visualization tool for analyzing representation transformations.

4.3 ReSi Benchmark

We evaluate the AUC-based multiscale similarity scores on a large-scale benchmark of representational similarity measures. The ReSi benchmark (Klabunde et al., 2025b) evaluates whether similarity measures capture known differences between models across Graph, Vision, and Language domains. Its tests include *Output Correlation*, measuring correlation with functional differences in model predictions using Jensen-Shannon divergence (JSD); *Accuracy Correlation*, measuring correlation with differences in model accuracy; *Label Randomization*, distinguishing models trained with different levels of label corruption; *Shortcut Affinity*, distinguishing models with shortcut features of varying strength; *Augmentation*, identifying differences in augmentation strength; and *Layer Monotonicity*, relating similarity to layer distance within a model.

The Graph domain includes Cora (Sen et al., 2008), Flickr (Rossi & Ahmed, 2015), and OGBN-Arxiv (Hu et al., 2020), with GCN (Kipf & Welling, 2017), GraphSAGE (Hamilton et al., 2017), and GAT (Veličković et al., 2018), and PGNN (You et al., 2019) for a subset of Cora experiments. Vision uses ImageNet100 and CIFAR100 (Krizhevsky et al., 2009) with ResNet18/34/101 (He et al., 2016), VGG11/19 (Simonyan & Zisserman, 2015), and ViT-B/32 and ViT-L/32 (Dosovitskiy et al., 2021). Language uses SST2 (Socher et al., 2013) and MNLI (Williams et al., 2018) with BERT-L (Devlin et al., 2019) and SmolLM2-1.7B (Allal et al., 2025). Thus, the benchmark spans both different architecture families and multiple model scales within them.

We evaluate four multiscale measures using $\mathrm{AUC}(\boldsymbol{s}_{\mathrm{CKA}})$, $\mathrm{AUC}(\boldsymbol{s}_{\mathrm{RWKA}})$, $\mathrm{AUC}(\boldsymbol{s}_{\mathrm{UKA}})$, and $\mathrm{AUC}(\boldsymbol{s}_{\mathrm{dRWKA}})$, with 100 geometrically spaced bandwidths over $\sigma \in [5 \cdot 10^{-2}, 10]$. We use a denser grid than in Section 4.2 because the benchmark spans domains and architectures with different informative bandwidth ranges. These multiscale measures form the *Signal* category in our evaluation. We compare against measures spanning four categories from the taxonomy of Klabunde et al. (2025a): *RSM*, *Neighbors*, *Alignment*, and *CCA*. Specifically, we include nine top-ranking measures from the original ReSi benchmark (Klabunde et al., 2025b): Aligned Cosine Similarity (Hamilton et al., 2016b), Linear CKA (Kornblith et al., 2019), Distance Correlation (Székely et al., 2007), Jaccard Similarity (Wang et al., 2020), Linear Regression (Kornblith et al., 2019), Procrustes Size-and-Shape Distance (Williams et al., 2021), Rank Similarity (Wang et al., 2020), SVCCA (Raghu et al., 2017), and Second-Order Cosine Similarity (Hamilton et al., 2016a). We additionally include three local measures, $\mathrm{CKA}_{\mathrm{RBF},0.2}$, MNN_{10}, and CKNNA_{10}. We use this common subset because results for the complete ReSi suite are not available across all architectures in the extended benchmark. Further implementation details are provided in Appendix C.

Because ReSi spans many datasets, architectures, and tasks, we aggregate performance separately within each domain. A *case* is a comparable evaluation context defined by a task, dataset, architecture, and quality measure. Within each case, we rank the measures separately for each seed, or layer pair where applicable, for which they return a finite value and average these ranks. Thus, each case contributes one rank per measure, regardless of the number of seeds or layer pairs it contains. To avoid favoring measures evaluated only on easier subsets, we exclude a measure from a domain if it returned NaN for more than 10% of its cases. This removes only Linear Regression from Vision, where it has 79.5% coverage. Figure 3 therefore compares sixteen measures in Graph and Language and fifteen in Vision. Measures are ordered by their 90th-percentile rank, favoring methods that perform consistently well across cases rather than those that are occasionally best.

Figure 3 shows that $\mathrm{AUC}(\boldsymbol{s}_{\mathrm{CKA}})$ achieves the best 90th-percentile rank in all three domains, a consistency not matched by any evaluated baseline. This advantage is not specific to the tail-based ordering: when measures are instead ranked by mean performance (Appendix C.2), a multiscale signal ranks first in every domain. $\mathrm{AUC}(\boldsymbol{s}_{\mathrm{CKA}})$ ranks first in Vision and Language and second in Graph, where $\mathrm{AUC}(\boldsymbol{s}_{\mathrm{UKA}})$ ranks first. Thus, across both robust and average-case summaries, multiscale signals consistently occupy the top of the ReSi rankings across domains. $\mathrm{AUC}(\boldsymbol{s}_{\mathrm{dRWKA}})$ also ranks highly, supporting the broader multiscale framework rather than a single signal construction. Appendix C.2 reports performance separately by task and domain, showing that the best-performing measure still varies across individual benchmark settings.

5 Conclusions and Future Work

We introduced a multiscale framework for representational similarity that treats scale as a variable rather than fixing it in advance. The resulting similarity signal captures alignment from local to global structure and can be aggregated into a scalar score. Across controlled and learned transformations, different geometric changes produced distinct multiscale signatures. On the ReSi benchmark, our multiscale scores performed consistently across graph, vision, and language domains, with the main score achieving the best 90th-percentile rank in all three, highlighting geometric scale as an informative dimension of representational similarity.

Future work includes studying the framework in settings where representational alignment is central, including CLIP-style vision–language models, REPA-based generative modeling, and knowledge distillation (Tung & Mori, 2019; Liang et al., 2022; Leng et al., 2025). Beyond these applications, extending the multiscale signal beyond Frobenius cosine similarity could capture complementary notions of alignment across scales.

REFERENCES

Kaarina Aho, Brett D Roads, and Bradley C Love. Signatures of cross-modal alignment in children's early concepts. *Proceedings of the National Academy of Sciences*, 120(42):e2309688120, 2023.

Loubna Ben Allal, Anton Lozhkov, Elie Bakouch, Gabriel Martín Blázquez, Guilherme Penedo, Lewis Tunstall, Andrés Marafioti, Hynek Kydlíček, Agustín Piqueres Lajarín, Vaibhav Srivastav, Joshua Lochner, Caleb Fahlgren, Xuan-Son Nguyen, Clémentine Fourrier, Ben Burtenshaw, Hugo Larcher, Haojun Zhao, Cyril Zakka, Mathieu Morlon, Colin Raffel, Leandro von Werra, and Thomas Wolf. SmolLM2: When smol goes big – data-centric training of a small language model, 2025. URL https://arxiv.org/abs/2502.02737.

Sergio A Alvarez. Gaussian rbf centered kernel alignment (cka) in the large-bandwidth limit. *IEEE transactions on pattern analysis and machine intelligence*, 45(5):6587–6593, 2022.

Mikhail Belkin, Partha Niyogi, and Vikas Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of machine learning research*, 7 (11), 2006.

Tianci Bi, Xiaoyi Zhang, Yan Lu, and Nanning Zheng. Vision foundation models can be good tokenizers for latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 43310–43319, 2026.

Bradley CA Brown, Anthony L. Caterini, Brendan Leigh Ross, Jesse C Cresswell, and Gabriel Loaiza-Ganem. Verifying the union of manifolds hypothesis for image data. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=Rvee9CAX4fi.

Chi-Ning Chou, Hang Le, Yichen Wang, and SueYeon Chung. Feature learning beyond the lazy-rich dichotomy: Insights from representational geometry. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=gKdjHLrHDS.

Stella Christie and Dedre Gentner. Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition and Development*, 11(3):356–373, 2010.

SueYeon Chung. Neural population geometry: An approach for understanding biological and artificial neural networks. *Current opinion in neurobiology*, 70:137–144, 2021.

Laure Ciernik, Lorenz Linhardt, Marco Morik, Jonas Dippel, Simon Kornblith, and Lukas Muttenthaler. Objective drives the consistency of representational similarity across datasets. In *International Conference on Machine Learning*, pp. 10920–10948. PMLR, 2025.

Uri Cohen, SueYeon Chung, Daniel D Lee, and Haim Sompolinsky. Separability and geometry of object manifolds in deep neural networks. *Nature communications*, 11(1):746, 2020.

Ronald R Coifman and Stéphane Lafon. Diffusion maps. *Applied and computational harmonic analysis*, 21(1):5–30, 2006.

MohammadReza Davari, Stefan Horoi, Amine Natik, Guillaume Lajoie, Guy Wolf, and Eugene Belilovsky. Reliability of cka as a similarity measure in deep learning. In *The Eleventh International Conference on Learning Representations*, 2023.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Li Kai, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255. IEEE, 2009.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, 2019. doi: 10.18653/v1/N19-1423.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *ICLR*, 2021.

Danny Driess, Fei Xia, Mehdi SM Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzaan Wahid, Jonathan Tompson, Quan Vuong, Tianhe Yu, et al. Palm-e: an embodied multimodal language model. In *Proceedings of the 40th International Conference on Machine Learning*, pp. 8469–8488, 2023.

Sathya Edamadaka, Soojung Yang, and Rafael Gomez-Bombarelli. Universally converging representations of matter across scientific foundation models. In *UniReps: 3rd Edition of the Workshop on Unifying Representations in Neural Models*, 2025.

Dedre Gentner and Christian Hoyos. Analogy and abstraction. *Topics in cognitive science*, 9(3): 672–693, 2017.

Dedre Gentner and Jose Medina. Similarity and the development of rules. *Cognition*, 65(2-3): 263–297, 1998.

Fabian Gröger, Shuo Wen, and Maria Brbić. Revisiting the platonic representation hypothesis: An aristotelian view. *arXiv preprint arXiv:2602.14486*, 2026.

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

William L Hamilton, Jure Leskovec, and Dan Jurafsky. Cultural shift or linguistic drift? comparing two computational measures of semantic change. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pp. 2116–2121, 2016a.

William L Hamilton, Jure Leskovec, and Dan Jurafsky. Diachronic word embeddings reveal statistical laws of semantic change. In *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 1: long papers)*, pp. 1489–1501, 2016b.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Martin N. Hebart, Adam H. Dickter, Alexis Kidder, Wan Y. Kwok, Anna Corriveau, Caitlin Van Wicklin, and Chris I. Baker. THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *PLOS ONE*, 14(10):e0223792, 2019. doi: 10.1371/journal.pone.0223792. URL https://doi.org/10.1371/journal.pone.0223792.

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

Minyoung Huh, Brian Cheung, Tongzhou Wang, and Phillip Isola. The platonic representation hypothesis. *arXiv preprint arXiv:2405.07987*, 2024.

Atharva Khandait and Jan E Gerken. From layers to networks: Comparing neural representations via diffusion geometry. *arXiv preprint arXiv:2605.15901*, 2026.

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

Max Klabunde, Tobias Schumacher, Markus Strohmaier, and Florian Lemmerich. Similarity of neural network models: A survey of functional and representational measures. *ACM Computing Surveys*, 57(9):1–52, 2025a.

Max Klabunde, Tassilo Wald, Tobias Schumacher, Klaus Maier-Hein, Markus Strohmaier, and Florian Lemmerich. Resi: A comprehensive benchmark for representational similarity measures. In *International Conference on Learning Representations*, volume 2025, pp. 18701–18740, 2025b.

Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey Hinton. Similarity of neural network representations revisited. In *International conference on machine learning*, pp. 3519–3529. PMLR, 2019.

Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. *Technical Report, University of Toronto*, 2009.

Joseph B Kruskal. Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, 29 (2):115–129, 1964.

Roy R Lederman and Ronen Talmon. Learning the geometry of common latent variables using alternating-diffusion. *Applied and Computational Harmonic Analysis*, 44(3):509–536, 2018.

Xingjian Leng, Jaskirat Singh, Yunzhong Hou, Zhenchang Xing, Saining Xie, and Liang Zheng. Repa-e: Unlocking vae for end-to-end tuning of latent diffusion transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 18262–18272, 2025.

Victor Weixin Liang, Yuhui Zhang, Yongchan Kwon, Serena Yeung, and James Y Zou. Mind the gap: Understanding the modality gap in multi-modal contrastive representation learning. *Advances in Neural Information Processing Systems*, 35:17612–17625, 2022.

Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. Segment anything in medical images. *Nature communications*, 15(1):654, 2024.

Harel Mendelman and Ronen Talmon. Supervised and semi-supervised diffusion maps with label-driven diffusion. In *International Conference on Learning Representations*, volume 2025, pp. 73278–73312, 2025.

Alex Graeme Murphy, Joel Zylberberg, and Alona Fyshe. Correcting biased centered kernel alignment measures in biological and artificial neural networks. In *ICLR Workshop on Representational Alignment*, 2024.

Lukas Muttenthaler, Lorenz Linhardt, Jonas Dippel, Robert A Vandermeulen, Katherine Hermann, Andrew Lampinen, and Simon Kornblith. Improving neural network representations using human similarity judgments. *Advances in neural information processing systems*, 36:50978–51007, 2023.

Boaz Nadler, Stéphane Lafon, Ronald R Coifman, and Ioannis G Kevrekidis. Diffusion maps, spectral clustering and reaction coordinates of dynamical systems. *Applied and Computational Harmonic Analysis*, 21(1):113–127, 2006.

Thao Nguyen, Maithra Raghu, and Simon Kornblith. Do wide and deep networks learn the same things? uncovering how neural network representations vary with width and depth. In *International Conference on Learning Representations*, 2021.

Maks Ovsjanikov, Mirela Ben-Chen, Justin Solomon, Adrian Butscher, and Leonidas Guibas. Functional maps: a flexible representation of maps between shapes. *ACM Transactions on Graphics (ToG)*, 31(4):1–11, 2012.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PMLR, 2021.

Maithra Raghu, Justin Gilmer, Jason Yosinski, and Jascha Sohl-Dickstein. Svcca: Singular vector canonical correlation analysis for deep learning dynamics and interpretability. *Advances in neural information processing systems*, 30, 2017.

Maithra Raghu, Thomas Unterthiner, Simon Kornblith, Chiyuan Zhang, and Alexey Dosovitskiy. Do vision transformers see like convolutional neural networks? *Advances in neural information processing systems*, 34:12116–12128, 2021.

Vinay Venkatesh Ramasesh, Ethan Dyer, and Maithra Raghu. Anatomy of catastrophic forgetting: Hidden representations and task semantics. In *International Conference on Learning Representations*, 2021.

Ryan A. Rossi and Nesreen K. Ahmed. The network data repository with interactive graph analytics and visualization. In *AAAI Conference on Artificial Intelligence*, pp. 4292–4293, 2015. URL http://networkrepository.com.

Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina Eliassi-Rad. Collective classification in evolutionary heterogeneous networks. *AI magazine*, 29(3):93–93, 2008.

Tal Shnitzer, Mirela Ben-Chen, Leonidas Guibas, Ronen Talmon, and Hau-Tieng Wu. Recovering hidden components in multimodal data with composite diffusion operators. *SIAM Journal on Mathematics of Data Science*, 1(3):588–616, 2019.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations (ICLR)*, 2015.

Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 1631–1642, 2013.

Ethan Steinberg, Ken Jung, Jason A Fries, Conor K Corbin, Stephen R Pfohl, and Nigam H Shah. Language models are an effective representation learning technique for electronic health record data. *Journal of biomedical informatics*, 113:103637, 2021.

Ilia Sucholutsky, Lukas Muttenthaler, Adrian Weller, Andi Peng, Andreea Bobu, Been Kim, Bradley C Love, Erin Grant, Iris Groen, Jascha Achterberg, et al. Getting aligned on representational alignment. *arXiv preprint arXiv:2310.13018*, 2023.

Jian Sun, Maks Ovsjanikov, and Leonidas Guibas. A concise and provably informative multi-scale signature based on heat diffusion. *Computer Graphics Forum*, 28(5):1383–1392, 2009.

Gábor J. Székely, Maria L. Rizzo, and Nail K. Bakirov. Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6):2769 – 2794, 2007.

Frederick Tung and Greg Mori. Similarity-preserving knowledge distillation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1365–1374, 2019.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph Attention Networks. *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=rJXMpikCZ. accepted as poster.

Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.

Chenxu Wang, Wei Rao, Wenna Guo, Pinghui Wang, Jun Liu, and Xiaohong Guan. Towards understanding the instability of network embedding. *IEEE Transactions on Knowledge and Data Engineering*, 34(2):927–941, 2020.

Yi Wang and Zhiren Wang. CW complex hypothesis for image data. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=XXioxiADDC.

Bar Weiss, Hau-Tieng Wu, and Ronen Talmon. Extracting common components from partially observed views using diffusion geometry. *Transactions on Machine Learning Research*, 2026.

Adina Williams, Nikita Nangia, and Samuel Bowman. A broad-coverage challenge corpus for sentence understanding through inference. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pp. 1112–1122, 2018.

Alex H Williams, Erin Kunz, Simon Kornblith, and Scott Linderman. Generalized shape metrics on neural representations. *Advances in neural information processing systems*, 34:4738–4750, 2021.

Jiaxuan You, Rex Ying, and Jure Leskovec. Position-aware graph neural networks. In *International conference on machine learning*, pp. 7134–7143. PMLR, 2019.

Sihyun Yu, Sangkyung Kwak, Huiwon Jang, Jongheon Jeong, Jonathan Huang, Jinwoo Shin, and Saining Xie. Representation alignment for generation: Training diffusion transformers is easier than you think. In *The Thirteenth International Conference on Learning Representations*, 2025.

AI USE STATEMENT

In this work, we used generative AI tools to formulate mathematical claims, assist in the writing of proofs, design or provide feedback on research methodology or experiments, implement methods, and retrieval and discovery of related work. We have not used generative AI tools to generate synthetic data sets, help develop theoretical models or conceptual frameworks, provide critical ingredients for proving mathematical claims, propose or refine hypotheses, assist with translation, clean and reformat dataset, support qualitative and thematic data analysis and interpret results. Additionally, we used generative AI tools to draft parts of a research paper, summarize or analyse existing literature, edit a research paper to improve readability, propose a title or keywords for a research paper. We have reviewed all AI-assisted work. Specifically, we checked and rewrote the assisted proofs, reviewed the methodologies, and tested the AI-written code before publication. Additionally, we carefully cross-checked all drafted text against the literature and verified all keywords. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

A HEAT DIFFUSION AND MULTISCALE GEOMETRY

Heat diffusion provides a canonical multiscale description of manifold geometry. It is governed by the Laplace–Beltrami operator $\Delta_{\mathcal{M}}$, the generalization of the Euclidean Laplacian to a Riemannian manifold, and its fundamental solution is the heat kernel $h_t^{\mathcal{M}}(\boldsymbol{x}, \boldsymbol{x}')$. As emphasized by Sun et al. (2009), the diffusion time t determines the scale: short times emphasize local structure, while longer times incorporate progressively broader neighborhoods.

For small t, the heat kernel has the local form

$$h_t^{\mathcal{M}}(\boldsymbol{x}, \boldsymbol{x}') \approx (4\pi t)^{-d/2} \exp\left(-\frac{d_{\mathcal{M}}(\boldsymbol{x}, \boldsymbol{x}')^2}{4t}\right), \tag{16}$$

where $d_{\mathcal{M}}$ is the geodesic distance on the manifold. For nearby points, the ambient Euclidean distance agrees with the geodesic distance to leading order. Thus, under the correspondence $t = \sigma^2/2$, the RBF kernel locally approximates the Gaussian term of the manifold heat kernel. This explains why small bandwidths probe local manifold geometry and motivates viewing increasing σ as extending the comparison toward progressively broader structure.

The Gaussian underlying the RBF kernel satisfies the semigroup property: a Gaussian at scale t can be decomposed into a narrower Gaussian at scale s followed by Gaussian smoothing at scale $t - s$. In particular, up to normalization,

$$k^{\mathbb{R}^p}_{\sqrt{2t}}(\boldsymbol{x}, \boldsymbol{x}') \propto \int_{\mathbb{R}^p} k^{\mathbb{R}^p}_{\sqrt{2(t-s)}}(\boldsymbol{x}, \boldsymbol{y}) k^{\mathbb{R}^p}_{\sqrt{2s}}(\boldsymbol{y}, \boldsymbol{x}')\, d\boldsymbol{y}, \qquad 0 < s < t. \tag{17}$$

For sufficiently small s, the narrow RBF restricted locally to $\mathcal{M}$ approximates the manifold heat kernel, $k^{\mathbb{R}^p}_{\sqrt{2s}}(\boldsymbol{y}, \boldsymbol{x}') \propto h_s^{\mathcal{M}}(\boldsymbol{y}, \boldsymbol{x}')$. For data supported on $\mathcal{M}$, this motivates the manifold-restricted interpretation

$$k^{\mathbb{R}^p}_{\sqrt{2t}}(\boldsymbol{x}, \boldsymbol{x}') \approx C_s \int_{\mathcal{M}} k^{\mathbb{R}^p}_{\sqrt{2(t-s)}}(\boldsymbol{x}, \boldsymbol{y}) h_s^{\mathcal{M}}(\boldsymbol{y}, \boldsymbol{x}')\, d\mathrm{vol}_{\mathcal{M}}(\boldsymbol{y}), \tag{18}$$

where C_s absorbs the normalization and $\mathrm{vol}_{\mathcal{M}}$ is the volume measure on $\mathcal{M}$. Thus, larger RBF bandwidths can be viewed as Euclidean smoothing of a local heat-kernel description of the manifold.

This Euclidean behavior at larger scales is desirable in our setting. While intrinsic heat diffusion remains constrained to within-manifold geometry, broader RBF kernels use ambient distances and can therefore capture the relative organization of different manifold components.

For intrinsic heat diffusion, the same scale interpretation is reflected spectrally through

$$h_t^{\mathcal{M}}(\boldsymbol{x}, \boldsymbol{x}') = \sum_{\ell=0}^{\infty} e^{-\lambda_\ell t} \phi_\ell(\boldsymbol{x}) \phi_\ell(\boldsymbol{x}'), \tag{19}$$

where $-\Delta_{\mathcal{M}} \phi_\ell = \lambda_\ell \phi_\ell$ with $\lambda_\ell \geq 0$. The eigenvalues provide a generalized notion of frequency on the manifold: low-frequency eigenfunctions vary smoothly and describe coarse, large-scale geometry, while high-frequency eigenfunctions capture finer, more local structure. Increasing t suppresses

higher-frequency modes more strongly, leaving progressively coarser geometric structure. This also adds to the interpretation of alignment in the spectral form of cosine similarity in equation 10.

B Supplementary Experimental Results

Here, we provide additional results on the experiments presented in Sections 4.1 and 4.2.

B.1 Synthetic Data

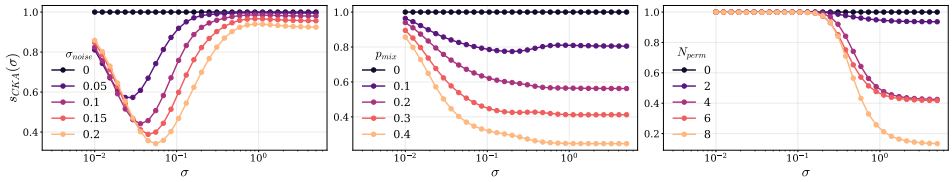
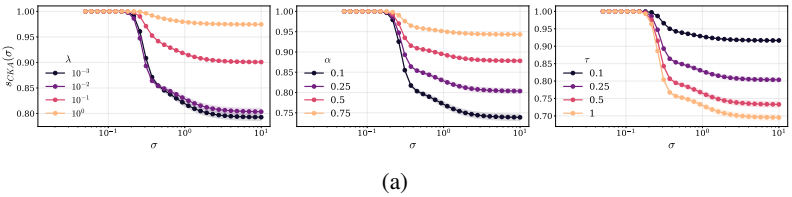


Figure 4: Synthetic experiment $s_{\mathrm{CKA}}^{X_\theta, X_{\theta_0}}$ signal for $\theta_0 = (0, 0, 0)$ swept over the synthetic transformation parameters $\sigma_{\mathrm{noise}}, p_{\mathrm{mix}}$ and N_{perm}.

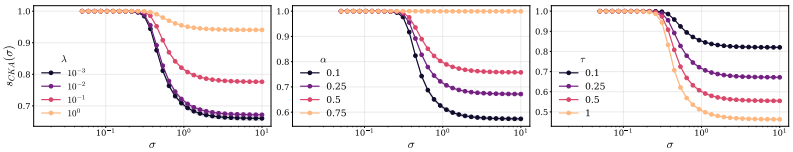
As an additional control, we repeat the synthetic experiment while varying one transformation at a time and setting the remaining transformation parameters to zero. The resulting signals are shown in Figure 4. This isolated setting yields the same qualitative scale dependence observed in the main experiment: Gaussian noise primarily affects small bandwidths, cluster mixing produces its largest changes at intermediate bandwidths, and cluster permutation is most visible at larger bandwidths. The main text uses sweeps around a nontrivial reference setting to demonstrate that these scale-dependent effects persist when multiple geometric perturbations are present simultaneously.

B.2 gLocal Transform

Here, we present additional results for the gLocal experiment from Section 4.2.



(a)



(b)

Figure 5: $s_{\mathrm{CKA}}^{X_\phi, X_0}$ multiscale similarity signals for sweeps over the gLocal parameters λ, α, and τ for (a) CLIP-ResNet50 and (b) DINOv2-ViT-Large. Each parameter is varied while the remaining two are fixed to the reference setting. Curves show the mean over 10 CIFAR-100 batches, shading denotes one standard deviation across batches.

We first examine whether the scale-dependent behavior observed for OpenCLIP-ViT-L-14-2B extends to other representation architectures. Figure 5 shows the multiscale similarity signals for sweeps over the gLocal parameters λ, α, and τ for CLIP-ResNet50 and DINOv2-ViT-Large. As in the main experiment, each parameter is varied while the remaining two are fixed to the reference setting. Across architectures, the signals exhibit similar qualitative behavior: increasing λ increases similarity across scales, while changes in α and τ affect broader and partially overlapping ranges of

the signal. These results indicate that the multiscale patterns observed in Section 4.2 are not specific to the OpenCLIP representation.

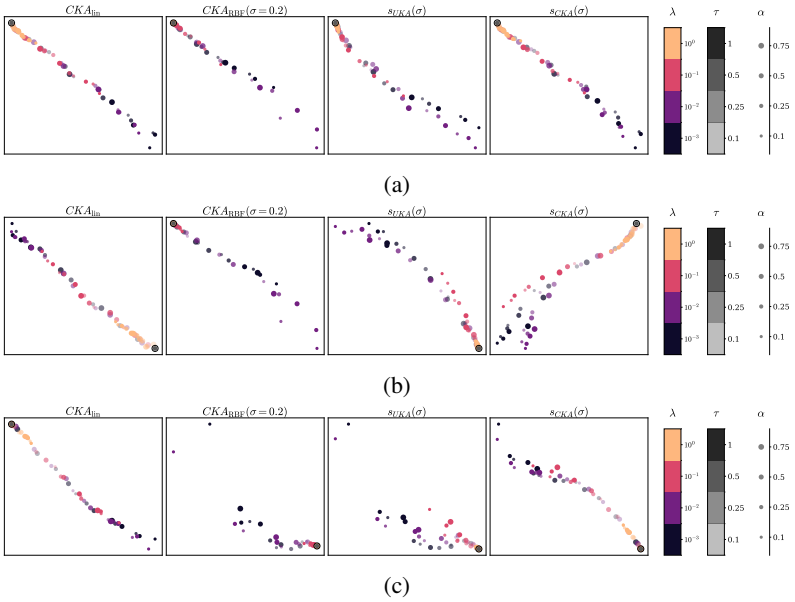


Figure 6: MDS embeddings of the similarity-to-original profiles over the full gLocal parameter grid for (a) CLIP-ResNet50, (b) OpenCLIP-ViT-L-14-2B, and (c) DINOv2-ViT-Large. We compare CKA_{lin}, $\text{CKA}_{\text{RBF},0.2}$, and the full signals $s_{\text{UKA}}(\sigma)$ and $s_{\text{CKA}}(\sigma)$.

To examine variation over the full gLocal parameter grid, we next apply the similarity-profile MDS analysis introduced in Section 4.1. For each transformation, we concatenate its similarity profiles across batches before computing pairwise distances and applying MDS. Figure 6 compares CKA_{lin}, $\text{CKA}_{\text{RBF},0.2}$, and the full signals $s_{\text{UKA}}(\sigma)$ and $s_{\text{CKA}}(\sigma)$, for CLIP-ResNet50, OpenCLIP-ViT-L-14-2B, and DINOv2-ViT-Large. The multiscale embeddings reveal parameter-dependent structure in the transformed representations, most clearly for OpenCLIP-ViT-L-14-2B with s_{CKA}, where different values of λ form distinct branches with additional within-branch variation associated with α and τ. For CLIP-ResNet50 and DINOv2-ViT-Large, the branches overlap more, but a weaker version of the same λ-associated organization can still be identified. The scalar measures also exhibit parameter-dependent structure, although their essentially one-dimensional embeddings make the different patterns harder to distinguish, and different measures emphasize different parameters. Overall, these embeddings provide a complementary qualitative visualization of variation across the gLocal parameter grid.

C ReSi Experiment Additional Details and Results

C.1 Experiment Details

Benchmark execution and baseline provenance. We integrate our measures into the original ReSi evaluation pipeline and use ReSi's model checkpoints, representation extraction, benchmark construction, and quality measures without modification. Baseline results have two sources. For architectures covered by the released ReSi results, we use the published result files directly. ReSi does not provide baseline results for CIFAR100, SmolLM2-1.7B, or PGNN; for these settings, we recompute the nine baseline measures reported in the main text using the original ReSi implementation. We restrict the comparison to this common baseline set because several additional ReSi measures are either unavailable for some architectures or substantially more expensive to recompute, which would otherwise result in different comparison sets across cases.

Quality measures and ranked tasks. We retain ReSi's original quality measures. Augmentation, label-randomization, and shortcut experiments are evaluated using AUPRC and violation rate; layer

Table 1: Structural coverage gaps in the ReSi experiments.

Domain	Setting	Missing evaluation
Graph	Augmentation	PGNN
Graph	Accuracy Correlation	All ReSi baseline measures
Vision	CIFAR100 Augmentation	ViT-B/32, ViT-L/32
Language	Augmentation	SmolLM2-1.7B
Language	Accuracy Correlation	SmolLM2-1.7B

monotonicity uses correlation and violation rate; and prediction-correlation experiments compare representational similarity with functional model differences. Following ReSi's main reported results, the ranking uses AUPRC for the group-separation tests, Spearman correlation for prediction correlation, and correlation for layer monotonicity. Violation rate is reported as its complement, *Conformity Rate*, but is not given an additional rank so that these tests do not receive twice the weight of correlation-based tasks. Similarly, Output Correlation is ranked using Jensen–Shannon divergence, while prediction disagreement is reported only as an additional evaluation in the appendix. Absolute accuracy difference is treated separately as the Accuracy Correlation task.

Ranking and missing values. Within each case, non-finite entries are omitted and the remaining measures are ranked using average ranks for ties. A measure missing from an individual comparison does not affect the ranks of measures that are present. Ranks are then averaged over all comparisons belonging to the case, so each task–dataset–architecture setting contributes one value per measure regardless of the number of seeds or layer pairs. To prevent a measure from receiving an artificially favorable rank by being evaluated only on an easier subset, we additionally exclude it from a case when more than 50% of its comparisons are non-finite and from a domain when it is available in fewer than 90% of that domain's cases. The latter criterion removes only Linear Regression in Vision. The threshold is intended as a coverage safeguard rather than a tuned parameter; the next-lowest relevant coverage is substantially higher.

Unavailable and invalid evaluations. A few gaps are inherited from the released benchmark. PGNN has no augmentation checkpoints and is therefore absent from that graph experiment. CIFAR100 augmentation does not include the ViT architectures, and language augmentation contains BERT-L but no SmolLM2 checkpoints. ReSi also provides no accuracy-difference results for its graph baselines, so Accuracy Correlation is omitted from the Graph domain. For SmolLM2, the corresponding models exist but their accuracy values were not released, so its Accuracy Correlation cases are also unavailable. We additionally exclude two corrupted Vision random-label checkpoints: ViT-L/32 seed 4 contains non-finite representations, while ViT-B/32 seed 0 exhibits a collapsed representation with unresolved pairwise distances. These exclusions are applied before ranking.

C.2 Additional Results

Per-Domain Mean Sorted In Figure 7, we present the same per-domain results as Figure 3 from Section 4.3 just sorted by the mean rank. The signal-based measures also perform strongly under mean-rank ordering. Specifically, $\mathrm{AUC}(\boldsymbol{s}_{\mathrm{CKA}})$ ranks first or second in every domain, with $\mathrm{AUC}(\boldsymbol{s}_{\mathrm{UKA}})$ ranking first in the Graph domain and $\mathrm{AUC}(\boldsymbol{s}_{\mathrm{dRWKA}})$ ranking second in the Language domain.

Per-Task Cross Domain Results Figure 8 further breaks down performance by ReSi task. Within each case, measures are first ranked over the available comparisons and these ranks are averaged to obtain one case-level rank per measure. For this per-task analysis, case-level ranks are then averaged within each subtask, defined by domain, dataset, task, and evaluation measure, collapsing differences in architectures and other case-specific identifiers. Each subtask therefore contributes one point to the distribution, preventing settings with more evaluated architectures from receiving disproportionate weight.

The results show substantial task dependence, with no measure dominating every evaluation. Nevertheless, $\mathrm{AUC}(\boldsymbol{s}_{\mathrm{CKA}})$ performs consistently well, ranking first on Output Correlation and Layer

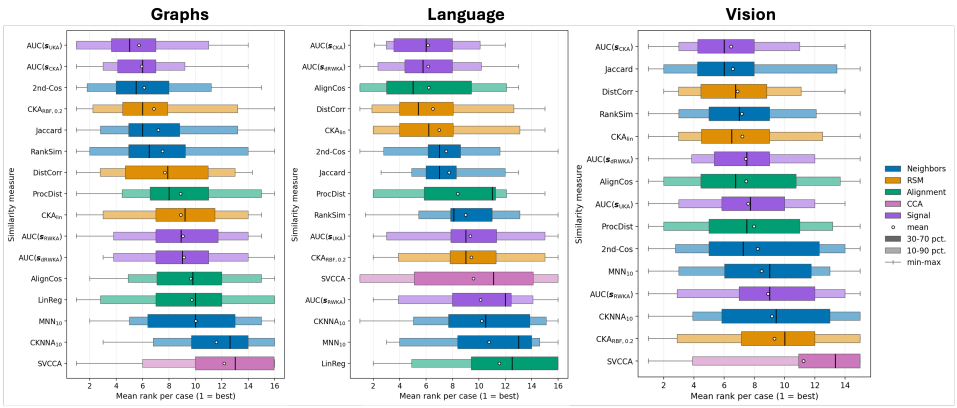


Figure 7: Per-domain ReSi performance. Distributions show within-case ranks, with lower values indicating better performance. Nested intervals denote the 10th-90th and 30th-70th percentiles; the black line marks the median, the white circle the mean, and whiskers the minimum and maximum. Measures are ordered by mean rank. These are the same results as in Figure 3 with just the ordering changed.

Monotonicity and among the leading measures on Accuracy Correlation, Shortcuts, and Augmentation. Its main exception is Label Randomization, where Distance Correlation and neighborhood-based measures perform better. Other multiscale signals also exhibit task-specific strengths, including $\mathrm{AUC}(\boldsymbol{s}_{\mathrm{UKA}})$ on Label Randomization and Shortcuts and $\mathrm{AUC}(\boldsymbol{s}_{\mathrm{dRWKA}})$ on Accuracy Correlation. Overall, the strong aggregate performance of the multiscale measures reflects consistency across heterogeneous tasks rather than dominance on a single benchmark.

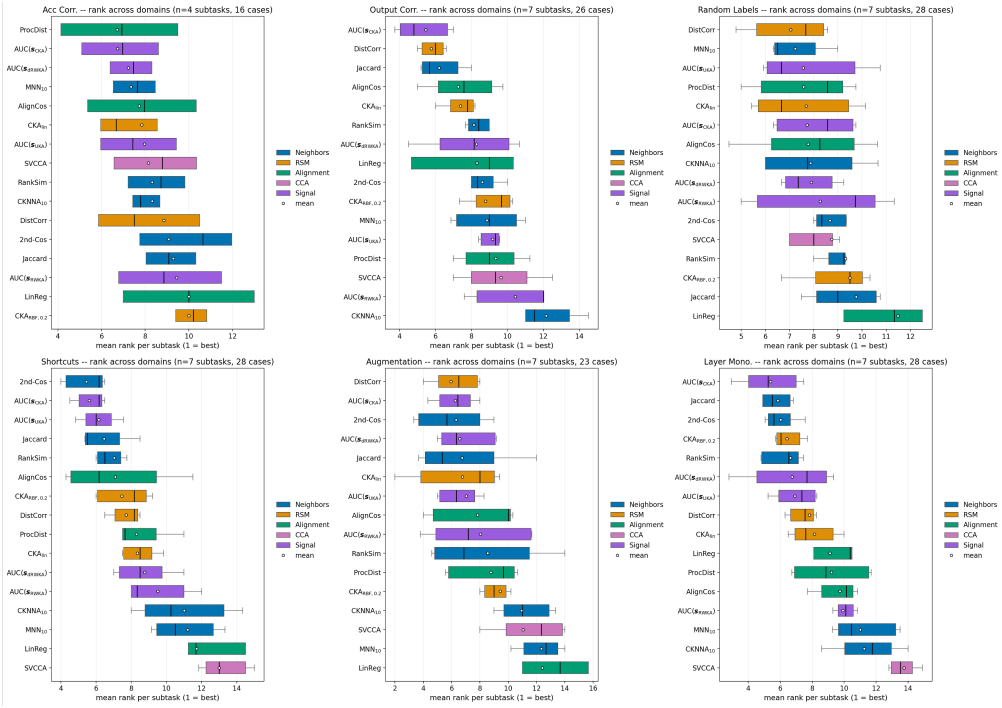


Figure 8: Per-task performance on the ReSi benchmark. Each panel shows the distribution of mean ranks across the available domain-dataset subtasks for one ReSi task, with lower ranks indicating better performance. Boxes summarize variation across subtasks, and markers denote the mean. Measures are grouped by similarity category. The number of subtasks and underlying cases is reported in each panel title.

Table 2: Averaged results over every model and dataset per domain (Graph: 4 models on 3 datasets; Text: 2 models on 2 datasets; Vision: 7 models on 2 datasets). Each cell averages seeds first, then that model's datasets, then the domain's models, so every model counts once regardless of how many datasets or seeds it ran. Higher values indicate that a similarity measure better reflects the ground truths from our tests. Best value in each column in bold, second best underlined, third best in italics.

	Type	Grounding by Prediction								Grounding by Design											
	Test	Output Corr.						Acc Corr.		Random Labels			Shortcuts			Augmentation			Layer Mono.		
	Eval.	Disagreement			JSD			Acc Diff		AUPRC			AUPRC			AUPRC			Spearman		
	Domain	Graph	Text	Vision	Graph	Text	Vision	Text	Vision	Graph	Text	Vision	Graph	Text	Vision	Graph	Text	Vision	Graph	Text	Vision
Neighbors	2nd-Cos	0.15	0.09	*0.11*	0.17	0.17	0.13	0.02	0.00	0.58	0.68	0.66	0.68	0.71	0.97	**0.88**	0.42	0.67	0.91	0.95	0.81
	CKNNA_{10}	-0.02	-0.06	-0.05	-0.02	0.03	0.01	0.08	0.01	0.50	0.75	**0.85**	0.47	0.63	0.79	0.68	0.39	0.50	0.61	0.74	0.76
	Jaccard	0.17	0.07	**0.19**	0.17	0.23	**0.27**	-0.02	0.05	0.45	0.76	0.74	0.67	0.71	**0.98**	0.86	0.37	0.77	0.98	*0.97*	0.77
	MNN_{10}	0.17	0.08	0.06	0.17	0.09	0.14	0.14	**0.14**	0.51	0.77	*0.78*	0.54	0.62	0.80	0.68	0.36	0.54	0.72	0.74	0.72
	RankSim	0.08	0.04	0.00	0.15	0.20	0.14	0.03	0.10	0.45	0.74	0.76	0.65	*0.70*	**0.98**	0.81	0.35	0.69	**0.99**	0.93	0.77
RSM	CKA_{lin}	0.19	0.26	0.07	0.17	0.32	*0.17*	0.06	*0.11*	0.54	*0.83*	0.77	0.58	0.66	0.97	0.79	**0.54**	0.66	0.84	0.95	0.80
	$\text{CKA}_{RBF,0.2}$	0.11	0.01	0.06	*0.23*	0.03	0.01	-0.02	0.06	*0.59*	0.63	0.71	**0.75**	0.69	0.86	0.74	0.42	0.45	0.93	0.94	**0.86**
	DistCorr	**0.22**	**0.31**	0.07	0.22	0.38	0.18	0.03	0.10	0.54	*0.83*	*0.78*	0.62	0.67	0.97	0.82	0.53	0.67	0.83	0.95	*0.82*
Alignment	AlignCos	0.12	0.21	0.15	0.15	**0.46**	0.16	*0.13*	0.05	0.52	**1.00**	0.69	0.64	0.66	**0.98**	0.68	0.40	**0.78**	0.71	0.92	0.59
	LinReg	*0.18*	0.17		0.15	0.22		-0.10		0.37	0.62		0.57	0.49		0.71	0.32		0.72	0.69	
	ProcDist	0.02	0.21	0.06	-0.05	*0.36*	0.18	**0.19**	0.12	0.54	0.96	0.71	0.59	0.65	0.97	0.79	0.38	0.70	0.91	0.88	0.61
CCA	SVCCA	-0.01	*0.24*	0.01	0.03	0.35	0.03	0.12	0.04	0.49	0.78	0.77	0.49	0.55	0.63	0.58	*0.52*	0.40	0.50	0.74	0.39
Signal (ours)	$\text{AUC}(\boldsymbol{s}_{\text{CKA}})$	0.17	0.20	0.08	**0.29**	0.30	0.18	0.12	0.10	0.54	0.74	*0.78*	0.72	0.71	0.97	0.78	0.42	*0.71*	0.94	0.98	0.84
	$\text{AUC}(\boldsymbol{s}_{\text{UKA}})$	0.10	0.05	0.03	0.24	0.11	0.13	0.02	0.08	**0.61**	0.70	0.79	*0.71*	**0.72**	*0.96*	*0.83*	0.41	0.64	*0.96*	0.94	*0.82*
	$\text{AUC}(\boldsymbol{s}_{\text{dRWKA}})$	0.10	0.08	0.06	0.10	0.26	0.15	0.08	0.09	0.56	0.75	0.77	0.62	0.68	*0.96*	0.76	0.49	0.64	0.85	**0.99**	0.79
	$\text{AUC}(\boldsymbol{s}_{\text{RWKA}})$	0.04	0.01	0.03	-0.01	0.16	0.12	-0.04	0.05	0.60	0.70	*0.78*	0.62	0.64	0.82	0.77	0.43	0.56	0.82	0.91	0.72

Average Absolute Scores Table 2 complements the rank-based analysis by reporting the underlying ReSi quality scores averaged within each domain. For each task and evaluation measure, we first average over seeds within each model–dataset setting. We then average the resulting values across datasets for each model, and finally across models within the domain. This hierarchical averaging gives each model equal weight, regardless of the number of datasets or seeds for which it was evaluated.

The results further illustrate the heterogeneity of the ReSi benchmark: no similarity measure dominates across all tasks and domains. Different baseline families perform best in different settings, including Distance Correlation on prediction disagreement, Jaccard and Aligned Cosine Similarity on several design-grounded tests, and Procrustes distance on accuracy correlation. The multiscale signals are similarly complementary, but remain competitive across a broad range of evaluations. In particular, $\text{AUC}(\boldsymbol{s}_{\text{CKA}})$ achieves the highest Graph JSD correlation, $\text{AUC}(\boldsymbol{s}_{\text{UKA}})$ performs best on Graph label randomization and Language shortcuts, and $\text{AUC}(\boldsymbol{s}_{\text{dRWKA}})$ achieves the highest Language layer-monotonicity score. Together with the rank results in Figure 3, these absolute scores indicate that the strong aggregate performance of the multiscale measures reflects broad consistency across heterogeneous tasks rather than dominance on a single evaluation.

D Random Walk Based Signals

The kernel-operator formulation of Section 3 naturally extends to operators from diffusion geometry (Coifman & Lafon, 2006). Rather than comparing the RBF kernel matrices directly, diffusion methods normalize pairwise affinities by their local connectivity, yielding operators whose spectral structure describes diffusion over the data geometry. Such normalized operators form the basis of diffusion maps (Coifman & Lafon, 2006; Nadler et al., 2006) and are closely related to other Laplacian-based geometric constructions such as functional maps (Ovsjanikov et al., 2012). Diffusion operators have also been used to extract and combine common geometric structure across multiple views through alternating diffusion and related methods (Lederman & Talmon, 2018; Shnitzer et al., 2019; Mendelman & Talmon, 2025; Weiss et al., 2026). More recently, Khandait & Gerken (2026) used alternating diffusion maps to fuse representations across layers before comparing neural networks.

D.1 Proposed Similarity Signals

Motivated by these connections, we introduce two multiscale similarity signals based on degree-normalized RBF kernels. These signals retain the Frobenius cosine similarity of Section 3, but replace the raw kernel matrices by operators associated with random walks on the data graph. This provides random-walk analogues of UKA and CKA that incorporate the density normalization commonly used in diffusion and manifold learning.

Random-Walk Kernel Alignment (RWKA) signal. Following Nadler et al. (2006), define the degree vector $\boldsymbol{d}_\sigma^X = \boldsymbol{K}_\sigma^X \mathbf{1}$, the degree matrix $\boldsymbol{D}_\sigma^X = \text{diag}(\boldsymbol{d}_\sigma^X)$, and define the symmetric random-

walk operator

$$\boldsymbol{S}_\sigma^X = (\boldsymbol{D}_\sigma^X)^{-1/2}\boldsymbol{K}_\sigma^X(\boldsymbol{D}_\sigma^X)^{-1/2}, \tag{20}$$

with analogous definitions for $\boldsymbol{Y}$. We define the random-walk kernel alignment (RWKA) signal as

$$s_{\mathrm{RWKA}}^{X,Y}(\sigma) = J_{\mathrm{RWKA}}\left(\boldsymbol{K}_\sigma^X, \boldsymbol{K}_\sigma^Y\right) = \cos_F\left(\boldsymbol{S}_\sigma^X, \boldsymbol{S}_\sigma^Y\right). \tag{21}$$

The eigenspaces of $\boldsymbol{S}_\sigma$ correspond to the diffusion modes of the representation. Equation 10 therefore interprets RWKA as a weighted alignment between the diffusion modes of the two representations at each scale. Degree normalization distinguishes RWKA from UKA by reweighting affinities according to local connectivity and thereby accounting for variation in sampling density.

The operator $\boldsymbol{S}_\sigma^X$ has leading eigenvalue one, with normalized stationary eigenvector $\boldsymbol{q}_\sigma^X = (\boldsymbol{d}_\sigma^X)^{1/2}/\|(\boldsymbol{d}_\sigma^X)^{1/2}\|_2$. This stationary mode is retained in RWKA and can dominate the alignment at broad scales, analogously to the constant mode retained by UKA. This motivates removing it before comparing the remaining diffusion geometry.

Degree-Centered Random-Walk Kernel Alignment (dRWKA) signal. Define the projection orthogonal to the stationary mode by

$$\boldsymbol{H}_{d,\sigma}^X = \boldsymbol{I} - \boldsymbol{q}_\sigma^X(\boldsymbol{q}_\sigma^X)^\top, \tag{22}$$

and analogously $\boldsymbol{H}_{d,\sigma}^Y$. We define the degree-centered random-walk kernel alignment (dRWKA) signal as

$$s_{\mathrm{dRWKA}}^{X,Y}(\sigma) = J_{\mathrm{dRWKA}}\left(\boldsymbol{K}_\sigma^X, \boldsymbol{K}_\sigma^Y\right) = \cos_F\left(\boldsymbol{H}_{d,\sigma}^X\boldsymbol{S}_\sigma^X\boldsymbol{H}_{d,\sigma}^X, \boldsymbol{H}_{d,\sigma}^Y\boldsymbol{S}_\sigma^Y\boldsymbol{H}_{d,\sigma}^Y\right). \tag{23}$$

Degree centering removes the stationary diffusion mode before comparing the remaining spectral geometry. CKA and dRWKA therefore apply analogous operations to different scale-dependent operators: CKA removes the uniform constant mode of the RBF kernel, whereas dRWKA removes the degree-dependent stationary mode of the diffusion operator. Consequently, RWKA and dRWKA can be viewed as random-walk analogues of UKA and CKA, respectively.

D.2 Endpoint Behavior

The distinction between retaining and removing the stationary mode leads to endpoint behavior analogous to that of UKA and CKA in Section 3.3.

Proposition 3. *For any finite paired samples $\boldsymbol{X}$ and $\boldsymbol{Y}$ with no duplicate representation vectors,*

$$\lim_{\sigma\to 0} s_{\mathrm{RWKA}}^{X,Y}(\sigma) = 1, \qquad \lim_{\sigma\to\infty} s_{\mathrm{RWKA}}^{X,Y}(\sigma) = 1. \tag{24}$$

At small bandwidths, $\boldsymbol{K}_\sigma \to \boldsymbol{I}$ and hence $\boldsymbol{S}_\sigma \to \boldsymbol{I}$, so the two operators coincide. At large bandwidths, $\boldsymbol{K}_\sigma \to \mathbf{1}\mathbf{1}^\top$ and $\boldsymbol{D}_\sigma \to n\boldsymbol{I}$, giving $\boldsymbol{S}_\sigma \to n^{-1}\mathbf{1}\mathbf{1}^\top$. The RWKA similarity therefore approaches one at both endpoints. As for UKA, these limits are representation-independent, so differences in $\mathrm{AUC}(\boldsymbol{s}_{\mathrm{RWKA}})$ arise primarily from the nontrivial scales between them.

Removing the stationary mode changes the large-scale behavior.

Proposition 4. *For any finite paired samples $\boldsymbol{X}$ and $\boldsymbol{Y}$ with no duplicate representation vectors,*

$$\lim_{\sigma\to 0} s_{\mathrm{dRWKA}}^{X,Y}(\sigma) = 1, \qquad \lim_{\sigma\to\infty} s_{\mathrm{dRWKA}}^{X,Y}(\sigma) = \mathrm{CKA}_{\mathrm{lin}}(\boldsymbol{X}, \boldsymbol{Y}). \tag{25}$$

For $\sigma \to 0$, we have $\boldsymbol{S}_\sigma \to \boldsymbol{I}$ and $\boldsymbol{q}_\sigma \to n^{-1/2}\mathbf{1}$, so $\boldsymbol{H}_{d,\sigma} \to \boldsymbol{H}$. The degree-centered operators therefore both converge to $\boldsymbol{H}$, and their cosine similarity approaches one. At large bandwidths, removing the stationary mode eliminates the leading representation-independent component of the diffusion operator. The remaining leading-order term is proportional to the centered squared-distance matrix, yielding linear CKA after normalization. The proof for this is quite similar to the CKA large scale limit shown by Alvarez (2022). We provide proof for this in Section D.3.

Thus, CKA-RBF and dRWKA provide two different multiscale continuations of linear CKA toward smaller geometric scales. CKA-RBF compares centered RBF kernel geometry directly, whereas

dRWKA first incorporates degree normalization and then removes the stationary diffusion mode. The two signals therefore share the same global endpoint while inducing different comparisons at local and intermediate scales. More broadly, RWKA and dRWKA illustrate how the multiscale framework can incorporate geometric operators from manifold learning without restricting representational similarity to a single operator construction.

D.3 dRWKA Large-Bandwidth Limit Proof

Here, we provide a proof of the large-bandwidth statement in Proposition 4. The argument follows a similar expansion to Alvarez (2022) for the large-bandwidth limit of $\mathrm{CKA}_{\mathrm{RBF},\sigma}$. Let $\epsilon = 2\sigma^2$ and define the normalized squared-distance matrix $\boldsymbol{\Delta}_X$ by $\boldsymbol{\Delta}_X(i,j) = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2/(d_{\mathrm{med}}^X)^2$, where d_{med}^X is the median pairwise distance used to normalize the RBF bandwidth. By a Taylor expansion of the exponential,

$$\boldsymbol{K}_\sigma^X = \mathbf{1}\mathbf{1}^\top - \epsilon^{-1}\boldsymbol{\Delta}_X + O(\epsilon^{-2}). \tag{26}$$

Let $\boldsymbol{\delta}_X = \boldsymbol{\Delta}_X\mathbf{1}$. The degree vector therefore satisfies

$$\boldsymbol{d}_\sigma^X = n\mathbf{1} - \epsilon^{-1}\boldsymbol{\delta}_X + O(\epsilon^{-2}). \tag{27}$$

Applying $(1+u)^{-1/2} = 1 - \frac{1}{2}u + O(u^2)$ elementwise gives

$$(\boldsymbol{D}_\sigma^X)^{-1/2} = \frac{1}{\sqrt{n}}\boldsymbol{I} + \frac{\epsilon^{-1}}{2n^{3/2}}\operatorname{diag}(\boldsymbol{\delta}_X) + O(\epsilon^{-2}). \tag{28}$$

Substituting these expansions into $\boldsymbol{S}_\sigma^X = (\boldsymbol{D}_\sigma^X)^{-1/2}\boldsymbol{K}_\sigma^X(\boldsymbol{D}_\sigma^X)^{-1/2}$ yields

$$\boldsymbol{S}_\sigma^X = \frac{1}{n}\mathbf{1}\mathbf{1}^\top - \frac{\epsilon^{-1}}{n}\boldsymbol{\Delta}_X + \frac{\epsilon^{-1}}{2n^2}\left(\boldsymbol{\delta}_X\mathbf{1}^\top + \mathbf{1}\boldsymbol{\delta}_X^\top\right) + O(\epsilon^{-2}). \tag{29}$$

We next expand the stationary mode $\boldsymbol{q}_\sigma^X = (\boldsymbol{d}_\sigma^X)^{1/2}/\|(\boldsymbol{d}_\sigma^X)^{1/2}\|_2$. Using $(1+u)^{1/2} = 1 + \frac{1}{2}u + O(u^2)$ gives

$$(\boldsymbol{d}_\sigma^X)^{1/2} = \sqrt{n}\,\mathbf{1} - \frac{\epsilon^{-1}}{2\sqrt{n}}\boldsymbol{\delta}_X + O(\epsilon^{-2}), \tag{30}$$

and

$$\|(\boldsymbol{d}_\sigma^X)^{1/2}\|_2 = n - \frac{\epsilon^{-1}}{2n}\mathbf{1}^\top\boldsymbol{\delta}_X + O(\epsilon^{-2}). \tag{31}$$

Thus,

$$\boldsymbol{q}_\sigma^X = \frac{1}{\sqrt{n}}\mathbf{1} - \frac{\epsilon^{-1}}{2n^{3/2}}\boldsymbol{\delta}_X + \frac{\epsilon^{-1}}{2n^{5/2}}(\mathbf{1}^\top\boldsymbol{\delta}_X)\mathbf{1} + O(\epsilon^{-2}), \tag{32}$$

and consequently

$$\boldsymbol{q}_\sigma^X(\boldsymbol{q}_\sigma^X)^\top = \frac{1}{n}\mathbf{1}\mathbf{1}^\top - \frac{\epsilon^{-1}}{2n^2}\left(\boldsymbol{\delta}_X\mathbf{1}^\top + \mathbf{1}\boldsymbol{\delta}_X^\top - \frac{2}{n}(\mathbf{1}^\top\boldsymbol{\delta}_X)\mathbf{1}\mathbf{1}^\top\right) + O(\epsilon^{-2}). \tag{33}$$

Since $\boldsymbol{S}_\sigma^X\boldsymbol{q}_\sigma^X = \boldsymbol{q}_\sigma^X$ and $\boldsymbol{S}_\sigma^X$ is symmetric, degree centering gives

$$\widetilde{\boldsymbol{S}}_\sigma^X = \boldsymbol{H}_{d,\sigma}^X\boldsymbol{S}_\sigma^X\boldsymbol{H}_{d,\sigma}^X = \boldsymbol{S}_\sigma^X - \boldsymbol{q}_\sigma^X(\boldsymbol{q}_\sigma^X)^\top. \tag{34}$$

Combining Equations 29 and 33, and using $\boldsymbol{H} = \boldsymbol{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$, gives

$$\widetilde{\boldsymbol{S}}_\sigma^X = -\frac{\epsilon^{-1}}{n}\boldsymbol{H}\boldsymbol{\Delta}_X\boldsymbol{H} + O(\epsilon^{-2}). \tag{35}$$

The same expansion holds for $\boldsymbol{Y}$. As shown by Alvarez (2022), for Euclidean squared distances,

$$\boldsymbol{H}\boldsymbol{\Delta}_X\boldsymbol{H} = -\frac{2}{(d_{\mathrm{med}}^X)^2}\boldsymbol{H}\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{H}. \tag{36}$$

Define $\boldsymbol{G}_X = \boldsymbol{H}\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{H}$ and $c_X = 2/[n(d^X_{\mathrm{med}})^2] > 0$, and analogously $\boldsymbol{G}_Y$ and c_Y. Equation 35 then gives

$$\widetilde{\boldsymbol{S}}^X_\sigma = \epsilon^{-1}c_X\left(\boldsymbol{G}_X + O(\epsilon^{-1})\right), \qquad \widetilde{\boldsymbol{S}}^Y_\sigma = \epsilon^{-1}c_Y\left(\boldsymbol{G}_Y + O(\epsilon^{-1})\right). \tag{37}$$

Since $\cos_F$ is invariant to positive rescaling, the factors $\epsilon^{-1}c_X$ and $\epsilon^{-1}c_Y$ cancel, giving

$$s^{X,Y}_{\mathrm{dRWKA}}(\sigma) = \cos_F\left(\boldsymbol{G}_X + O(\epsilon^{-1}), \boldsymbol{G}_Y + O(\epsilon^{-1})\right). \tag{38}$$

Here, the matrix-valued $O(\epsilon^{-1})$ terms are understood in Frobenius norm. Since $\boldsymbol{G}_X$ and $\boldsymbol{G}_Y$ are nonzero,

$$\left\langle \boldsymbol{G}_X + O(\epsilon^{-1}), \boldsymbol{G}_Y + O(\epsilon^{-1})\right\rangle_F = \langle \boldsymbol{G}_X, \boldsymbol{G}_Y\rangle_F + O(\epsilon^{-1}), \tag{39}$$

while

$$\|\boldsymbol{G}_X + O(\epsilon^{-1})\|_F = \|\boldsymbol{G}_X\|_F + O(\epsilon^{-1}), \qquad \|\boldsymbol{G}_Y + O(\epsilon^{-1})\|_F = \|\boldsymbol{G}_Y\|_F + O(\epsilon^{-1}). \tag{40}$$

Therefore,

$$s^{X,Y}_{\mathrm{dRWKA}}(\sigma) = \frac{\langle \boldsymbol{G}_X, \boldsymbol{G}_Y\rangle_F}{\|\boldsymbol{G}_X\|_F\|\boldsymbol{G}_Y\|_F} + O(\epsilon^{-1}) = \mathrm{CKA}_{\mathrm{lin}}(\boldsymbol{X}, \boldsymbol{Y}) + O(\epsilon^{-1}). \tag{41}$$

Finally, since $\epsilon = 2\sigma^2$,

$$s^{X,Y}_{\mathrm{dRWKA}}(\sigma) = \mathrm{CKA}_{\mathrm{lin}}(\boldsymbol{X}, \boldsymbol{Y}) + O(\sigma^{-2}), \qquad \sigma \to \infty, \tag{42}$$

which proves the large-bandwidth statement in Proposition 4.