Extended Abstract Track

Exploring Geometric Representational Alignment through Ollivier-Ricci Curvature and Ricci Flow

Editors: List of editors' names

Abstract

Representational analysis explores how input data of a neural system are encoded in high-dimensional spaces of its distributed neural activations, and how we can compare different systems, for instance, artificial neural networks and brains, on those grounds. While existing methods offer important insights, they typically do not account for local intrinsic geometrical properties within the high-dimensional representation spaces. To go beyond these limitations, we explore Ollivier-Ricci curvature and Ricci flow as tools to study the alignment of representations between humans and artificial neural systems on a geometric level. As a proof-of-principle study, we compared the representations of face stimuli between VGG-Face, a human-aligned version of VGG-Face, and corresponding human similarity judgments from a large online study. Using this discrete geometric framework, we were able to identify local structural similarities and differences by examining the distributions of node and edge curvature and higher-level properties by detecting and comparing community structure in the representational graphs.

Keywords: Representational alignment, Ollivier-Ricci curvature, Ricci flow, Graph-based representation

1. Introduction

Artificial neural networks (ANNs) can match human performance in image recognition and classification tasks among others (Mohsenzadeh et al., 2020). This led to the investigation of how ANNs encode, transform, and generalize information, and if and how these processes can be related to brains (representational alignment) (Richards et al., 2019). One key direction is to study how the geometry of internal representations reflects item similarity, categorical divisions, or latent variation within the input data(Chung and Abbott, 2021). The similarity between vectors in representational space is commonly measured using a kernel or a Representational Dissimilarity Matrix (RDM); RDMs are then commonly compared by applying Representational Similarity Analysis (RSA) by computing the correlations between RDMs (Kriegeskorte et al., 2008). Despite its popularity, RSA has notable limitations, e.g. Dujmović et al. (2022) showed that the correspondence between activation patterns in different systems can depend on the dataset, and that seemingly similar representational geometries actually encode different features. To improve this, it is crucial to capture intrinsic geometric properties of high-dimensional data, assuming that this data lies on a lower-dimensional manifold within the ambient space, where relationships between points reflect meaningful variations and similarities in the data (Lin and Zha, 2008). However, RSA imposes an ambient space geometry (Euclidean) that can distort these intrinsic geometrical relationships.

To respect the manifold's geometry and to overcome the limitations of classic RSA, we utilize a graph representation of the data (as a discrete counterpart of a manifold) and

© 2024 .

employ the Ollivier Ricci curvature (ORC) method, a discrete, graph-based analog of Ricci curvature in Riemannian geometry (Ollivier, 2007). ORC reflects the graph geometry by considering local neighborhoods and computing the optimal transport between probability measures at each node in the graph. At the same time, ORC avoids imposing a metric on the data manifold from the ambient space. ORC measures the deviation of a neighborhood structure from being flat (in the discrete case: having a grid-like topology). This notion of non-flatness relates to patterns of local connectivity in the graph. Positive curvature of an edge means that there are more connections and negative curvature indicates fewer connections compared to a grid. The Ricci flow process then allows us to reveal the communities of a graph which resemble regions in Riemannian manifolds of large positive curvature. For more details, see Appendix A.

We employed ORC to data from a human online experiment (REMOVED), in which 1,397 participants were asked to judge similarities between 100 generated faces in a triplet-odd-one-out task (161,700 triplet combinations). Subsequently, human behavior was modeled using a custom adaptation of VGG-Face (Parkhi et al., 2015) that was trained to predict human choices given sets of three face images (Aligned VGG-Face), see Appendix B and REMOVED.

The tools from discrete geometry enabled us to compare similarity spaces from human responses, and activation patterns from both the VGG-Face and Aligned VGG-Face model. For the former, we used activations in the FC7 layer. For the latter, we took activations in the VGG-bridge, a dense layer connecting a frozen, pre-trained part of the original VGG-Face with a new decision block specific to the similarity judgment task (see Figure A1 and Appendix B.2). Then, we constructed graph representations based on the human responses (Human judgment graph), Aligned VGG-Face, and VGG-Face. We applied ORC to analyze local neighborhood structures in the graphs and to gauge how the additional information introduced by the alignment process remodels the structural properties of representations in the original VGG-Face. By utilizing edge properties derived from edge curvature values and applying Ricci flow, we detected communities within graph representations. To gain a broader understanding of representational geometries, we compared these results with various community metrics, such as modularity.

2. Results

A comparative analysis of node and edge curvatures across models. In our graphs, nodes represented faces, and an edge between two nodes indicated that the corresponding images were considered similar by human participants or in terms of the vector representation within the two ANNs (details on graph constructions are given in Appendix A). We first computed ORC on the graphs' edge level and then derived the node curvature by averaging the curvature values across all edges connected to each node.

We observed a notable difference in ORC between models, with the Aligned VGG-Face matching the curvature structure in the Human judgment graph more closely than the original VGG-Face as shown in Figure A2 in the Appendix. The distribution of edge curvature values for both the Human judgment graph and Aligned VGG-Face showed similar value ranges with a mean shifted towards positive curvatures and a left-skewed shape, indicating a higher proportion of positive curvature with a sub-population of edges with strongly neg-

Extended Abstract Track

ative curvature. In contrast, the original VGG-Face displays a more symmetrical pattern distributed around zero. This implies that the data in the aligned network and the human similarity judgments both form local clusters connected by fewer (negatively curved) edges, indicating distinct structural characteristics compared to the original VGG-Face. To quantify the similarity between the distributions, we computed the Kullback-Leibler (KL) divergence between each VGG-Face version and the Human judgment graph. Indeed, we found a low KL divergence between the Human judgment graph and the Aligned VGG-Face, but a higher divergence with the Original VGG-Face (Table 1).

This distinction between the VGG-Face and Aligned VGG-Face structure was also reflected in the node curvature level. VGG-Face qualitatively shows no strong correlation with the Human judgment (Figure A2b), consistent with the differences in edge curvature distribution (Figure A2c, e). Interestingly, while the edge curvatures between the Human judgment graph and the Aligned VGG-Face were similar on the distribution level, their agreement fluctuated at individual nodes. That is, the node curvature shows a similar tendency (negative vs. positive) for most nodes, however, the node curvature values diverge for some nodes, indicating that these nodes belong to different local neighborhoods and thus, their interpretation in terms of cluster membership seemed to differ between humans and ANNs (sample images are shown in Figure A2a and b).

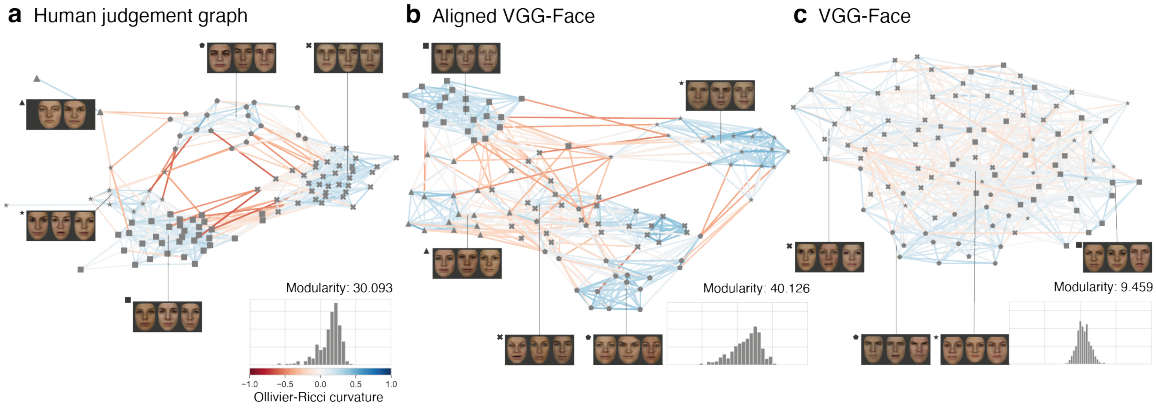

Figure 1: Detected communities (different symbols) by Ricci flow in the three experiments and the corresponding edge curvature value distribution. For each community we show a random sample of faces drawn from it. The edge colors in graphs indicate curvature values, with red shades representing negative curvature and blue shades indicating positive values. Modularity measurement as a community metric is attributed to each graph representation.

Ricci Flow analysis provides insights into structural properties of representational geometries. To detect and compare community structures based on our curvature measurements, we computed the discrete Ricci flow (Ni et al., 2019) in each representational graph. Ricci flow carves out the community structure of a graph by deforming space sim-

ilarly to heat diffusion: regions with large positive curvature contract, while those with strong negative curvature expand (see Appendix A for details). We then analyzed the structural properties of the processed graphs, as summarised in Table 1. In particular, we computed *conductance* (community separation by between-and-within edge ratio), *internal edge density* (edges within community vs. maximum possible), *modularity* (community strength by comparing edge densities to a random graph), and *average embeddedness* (number of shared neighbors for pairs of nodes within a community (see also Appendix A). These results (Table 1) clearly show that the Human judgment graph and the graph representation of the Aligned VGG-Face have very similar properties in terms of community structure. In contrast, the Original VGG-Face, despite being trained on face images, does not show a similarly high degree of structure.

Next, we wanted to visually assess the community structure and its relation to edge curvatures. As shown in Figure 1, the Human judgment graph and the Aligned VGG-Face exhibited similar community patterns overall, and both show similar tendencies to cluster, e.g., male and female face images (as indicated by the three randomly chosen samples per community). However, the cluster structure in the Aligned VGG-Face is less clearly defined compared to the Human judgment graph. In line with our other findings, the VGG-Face graph does not show a clear and coherent structure that we could relate to the Human judgment graph.

Table 1: Comparison of graph structure based on KL-divergence (KLD) between edge curvature distributions and community metrics derived from Ricc flow: *Conductance*, *Internal edge density* (IED), *Modularity*, and *Average embeddedness* (AE).

Graph	**KLD**	**Communities**	**Conductance**	**IED**	**Modularity**	**AE**
Human Judgment	0.000	5	0.297	0.563	30.093	0.767
Aligned VGG-Face	0.183	5	0.173	0.550	40.126	0.827
Original VGG-Face	0.697	4	0.417	0.386	9.459	0.604

3. Discussion

By employing Ollivier Ricci curvature as a tool for studying the similarity between the intrinsic geometries of neural representations independent of the ambient space, we aimed to go beyond the limitations of distance-based approaches, such as in classical RSA. Analyzing node and edge curvature yields insights into the local geometric properties of the underlying manifold. Additionally, we can refine this information using Ricci flow to obtain a broader perspective of the graph structure and detect communities within the representational spaces. With these methods, we could show how the alignment of an ANN to human behavior shifts its representational structure toward its target. We hope that this approach can motivate new avenues for understanding and analyzing the geometry of representations across neural systems.

Acknowledgments

Removed for review

References

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

Marin Dujmović, Jeffrey S Bowers, Federico Adolfi, and Gaurav Malhotra. Some pitfalls of measuring representational similarity using representational similarity analysis. *bioRxiv*, pages 2022–04, 2022.

Yao Feng, Haiwen Feng, Michael J Black, and Timo Bolkart. Learning an animatable detailed 3d face model from in-the-wild images. *ACM Transactions on Graphics (ToG)*, 40(4):1–13, 2021.

Nikolaus Kriegeskorte, Marieke Mur, and Peter A Bandettini. Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, 2:249, 2008.

Tong Lin and Hongbin Zha. Riemannian manifold learning. *IEEE transactions on pattern analysis and machine intelligence*, 30(5):796–809, 2008.

Debbie S. Ma, Joshua Correll, and Bernd Wittenbrink. The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, 47(4):1122–1135, 2015. doi: 10.3758/s13428-014-0532-5.

Yalda Mohsenzadeh, Caitlin Mullin, Benjamin Lahner, and Aude Oliva. Emergence of visual center-periphery spatial organization in deep convolutional neural networks. *Scientific reports*, 10(1):4638, 2020.

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

Yann Ollivier. Ricci curvature of metric spaces. *Comptes Rendus Mathematique*, 345(11): 643–646, 2007.

Omkar Parkhi, Andrea Vedaldi, and Andrew Zisserman. Deep face recognition. 2015.

FOR REVIEW REMOVED. Removed for review.

Blake A Richards, Timothy P Lillicrap, Philippe Beaudoin, Yoshua Bengio, Rafal Bogacz, Amelia Christensen, Claudia Clopath, Rui Ponte Costa, Archy de Berker, Surya Ganguli, et al. A deep learning framework for neuroscience. *Nature neuroscience*, 22(11):1761–1770, 2019.

Areejit Samal, RP Sreejith, Jiao Gu, Shiping Liu, Emil Saucan, and Jürgen Jost. Comparative analysis of two discretizations of ricci curvature for complex networks. *Scientific reports*, 8(1):8650, 2018.

Appendix A. Methods

A.1. Graph Construction

We generated a reference graph from the human judgment matrix by applying a threshold to its values, which range from 0 to 1, with higher values indicating greater similarity. Since the matrix directly reflects similarity measurements between images, thresholding is an effective method for graph construction. We set the threshold at 0.55 to ensure a connected graph, meaning that images with a similarity score above 0.55 are considered similar, while those below are not. This approach resulted in the creation of an unweighted graph. The neural network projects the data into a high-dimensional space, forming distinct data manifolds that fundamentally differ from the original input. Given the heterogeneity of the human judgment graph, including variations in edge distribution and node degrees, we utilized the nearest neighbor method to develop graphs, selecting the ten nearest neighbors for each node.

A.2. Ricci Curvature

In Riemannian geometry, curvature measures how a manifold deviates from being locally Euclidean, with Ricci curvature specifically quantifying this deviation in tangent directions Samal et al. (2018). It influences the average spread of geodesics in those directions and the rate at which the volume of distance balls and spheres expands. Geometrically, Ricci curvature determines the rate at which a ball's volume increases with its radius and the volume of overlap between two balls based on their radii and the distance between their centers. These properties are interconnected, as demonstrated by the following formula:

$$Vol_{\alpha}(\epsilon) = d\alpha\epsilon^{n-1}\big(1 - \frac{Ric(v)}{3}\epsilon^2 + o(\epsilon^2)\big) \tag{1}$$

where, n represents the dimension of the Riemannian manifold, while $\mathrm{Vol}_{\alpha}(\epsilon)$ denotes the $(n-1)$-dimensional volume formed within an n-dimensional solid angle $d\alpha$ by geodesics of length ϵ in the direction of the vector v.

A.3. Ollivier Ricci Curvature

In Riemannian geometry, curvature describes how a manifold deviates from being locally similar to Euclidean space, with Ricci curvature specifically measuring this deviation in various tangent directions Samal et al. (2018). Geometrically, Ricci curvature influences the rate at which the volume of a ball expands as its radius increases, as well as the volume of the overlap between two balls, depending on their radii and the distance between their centers. Moreover, the overlap volume between two balls is directly connected to the transportation cost needed to move one ball to the other—a greater overlap volume implies a lower transportation cost. This relationship highlights a connection between Ricci curvature and optimal transportation. Using this concept, Ollivier introduced a generalized form of Ricci curvature on metric measure spaces based on optimal transportation Ollivier (2007). For a metric space (X, d) equipped with a probability measure m_x for each $x \in X$, Ollivier's Ricci curvature (OR) along a path xy is defined as follows

Extended Abstract Track

$$\kappa_{xy} = 1 - \frac{W_1(m_x, m_y)}{d(x, y)} \tag{2}$$

where $W_1(m_x, m_y)$ is the Wasserstein distance.

A.4. Ricci Flow

The Ricci flow method, based on the geometric concept of curvature introduced by F. Gauss and B. Riemann, describes how space bends at each point. Areas with high positive curvature are denser, while regions with negative curvature are less so. Hamilton developed the Ricci flow, a curvature-driven diffusion process, which deforms space similarly to heat diffusion, regions with large positive curvature contract, while those with strong negative curvature expand. Ni et al. (2019) adapted Ricci flow from Riemannian geometry to discrete networks, using it to detect community structures within graphs. The discrete Ricci flow algorithm on a network is an evolving process. In each iteration, all edge weights update simultaneously by the following flow process:

$$w_{xy}^{i+1} = d^i(x, y) - \kappa^i(x, y).d^i(x, y) \tag{3}$$

where w_{xy}^i is the weight of the edge xy at the i-th iteration, and κ_{xy}^i is the Ricci curvature at the edge xy at the i-th iteration, and $d_{(x,y)}^i$ is the shortest path distance on the graph induced by the weights w_{xy}^i. Initially $w_{xy}^0 = w_{xy}$ and $d_{xy}^0 = d_{xy}$

A.5. Community Metrics

Conductance: It measures the ratio of the number of edges that cross the communities (external edges) to the total number of edges within the communities and external edges of the communities. Low conductance indicates that the community is well separated with few connections to other communities and high conductance suggests that the community is not well identified.

Modularity: It measures the strength of the division by comparing the density of edges inside communities to the density of edges in a random graph with the same degree distribution.

Average-Embeddedness: measures the number of shared neighbors for pairs of nodes within a community. it captures the cohesiveness of a community.

Internal Edge Density: It measures the density of edges within the community compared to the maximum possible number of edges in that community.

Appendix B. Data and Neural Networks

B.1. Face Stimuli and Human Similarity Judgments

Stimulus images were computed using the 3D reconstruction model DECA (Feng et al., 2021) applied to 2D portraits of the Chicago Face Dataset (Ma et al., 2015).

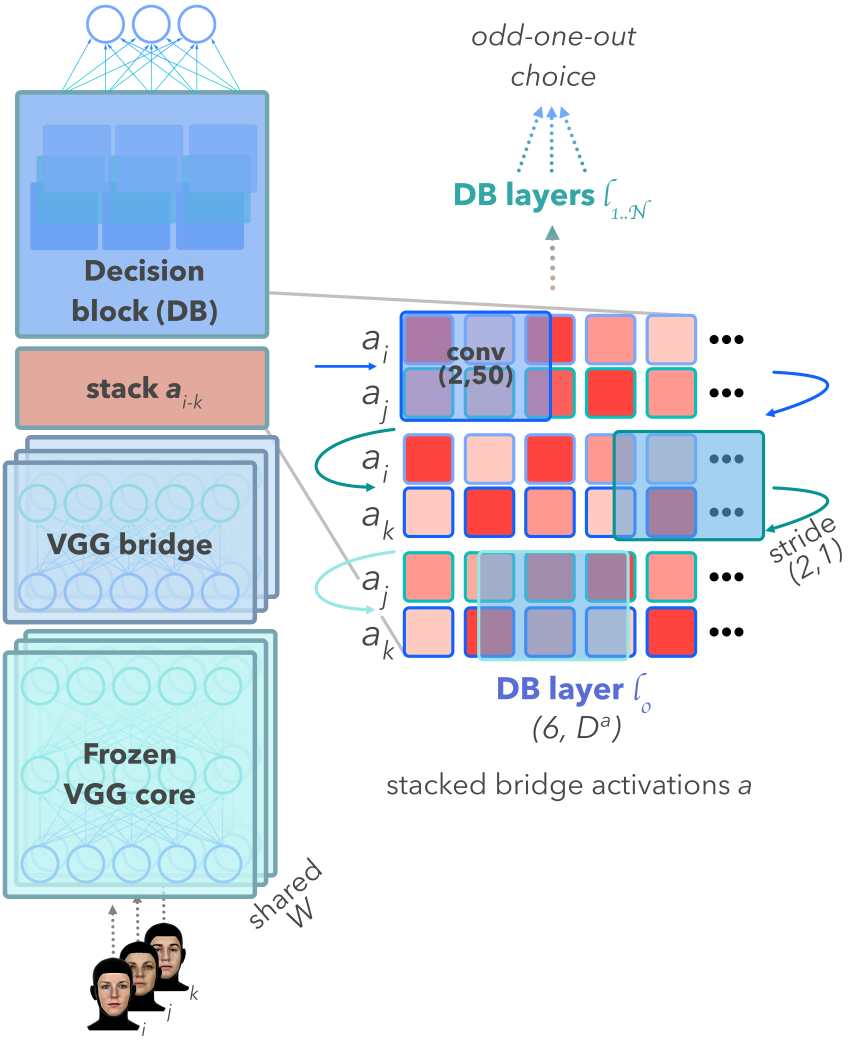

Figure A1: **Human-aligned VGG-Face**. The network is trained to predict human judgments in a face similarity task. The figure is adopted from REMOVED.

Human face similarity judgments (n = 194,261) were acquired in the form of a triplet-odd-one-out task from 1,397 participants (age range 18 - 65, mean age = 31.9 ± 11.2 years) in an online experiment. For more details on the stimulus set and experimental design, we refer to REMOVED.

B.2. Human-aligned VGG-Face

We adopted the pre-trained VGG-Face architecture (Parkhi et al., 2015) to predict human face similarity judgments in the experiment (Figure A1).

Extended Abstract Track

First, we froze all layers up to the fully connected layer FC7, making their weights non-trainable (*VGG core*). Second, we replaced subsequent layers with one FC layer (*VGG bridge*), which converts a 4,096-dimensional input to a 300-dimensional vector. We added a *decision block* consisting of convolutional layers. This block receives stacked activation maps from the bridge for each input image in a triplet (x_i, x_j, x_k), resulting in a 6x300 matrix $[a_i, a_j, a_i, a_k, a_j, a_k]$. The first convolutional layer in the decision block has 2 filters of size (2, 50) and stride (2, 1), producing an output of size (batch size, 2, 3, 251). After applying a ReLU activation, another convolutional layer with one filter of size (3, 100) and stride (1, 1) is applied, followed by another ReLU. This results in a (batch size, 2, 1, 152) output. Then, we down-sampled the signal to (batch size, 1, 1, 3) using two more convolutional layers (one filter each, kernel sizes: (1, 100) and (1, 51)) with an intermediate ReLU. The resulting 3-length output vector indicates the model's choice, where the highest value identifies the odd-one-out.

We trained the architecture using cross-entropy loss with the Adam optimizer, a learning rate of $5e^{-4}$, and a batch size of 16. The data (X: triplet images, Y: human choices) was split into training (70%; $n_{train} = 135,982$), validation (15%; $n_{val} = 29,139$), and test sets (15%; $n_{test} = 29,140$). For more details see REMOVED.

Appendix C. Supplementary Figure

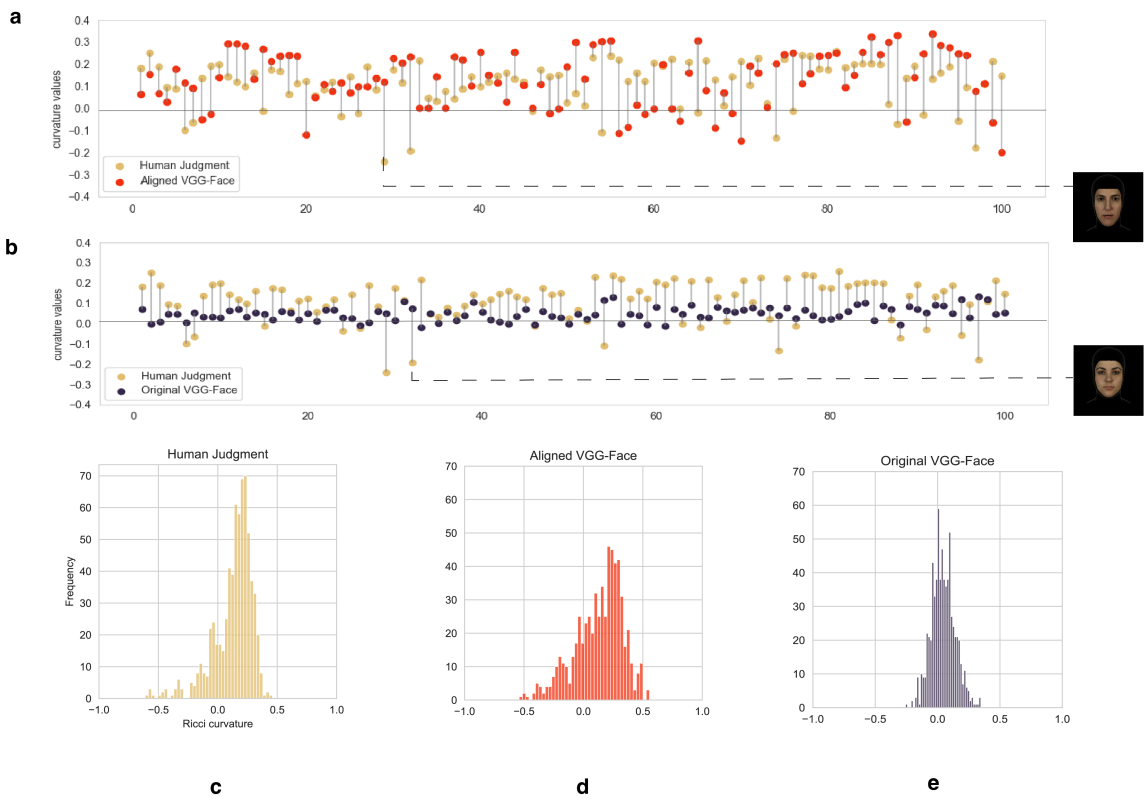

Figure A2: The node curvature distribution (a and b) and edge curvature distribution (c, d, and e) across different graphs. We compare the individual node curvatures of the Human Judgment graph (yellow dots) with that of the Aligned VGG-Face (red dots) in (a) and the Original VGG-Face (blue dots) in (b). The x-axis represents the nodes (images), while the y-axis shows the corresponding node curvature values. Two examples where node curvature diverges between models are highlighted in (a) and (b). The edge curvature distributions are displayed for the Human Judgment graph (c), Aligned VGG-Face (d), and Original VGG-Face (e).