**Exploratory image similarity calculation**

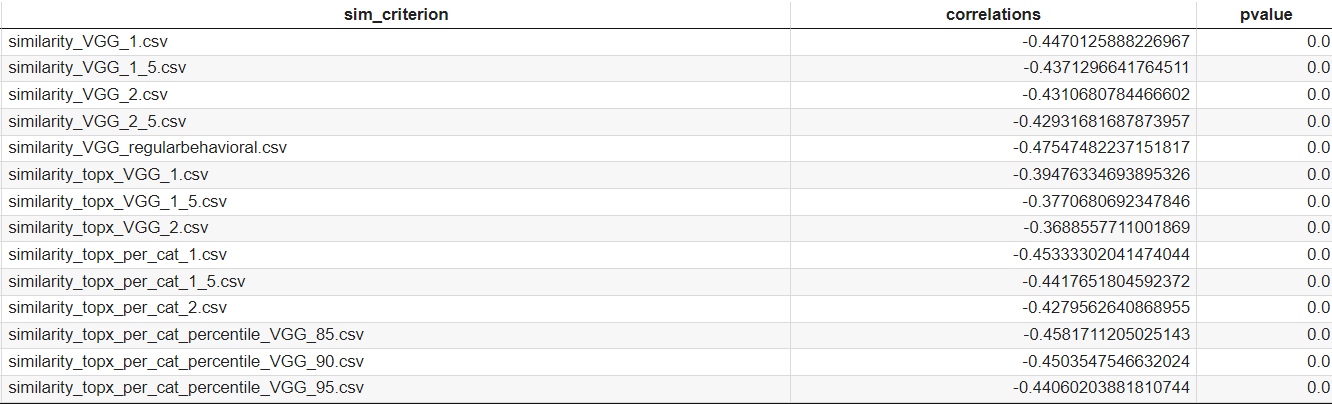
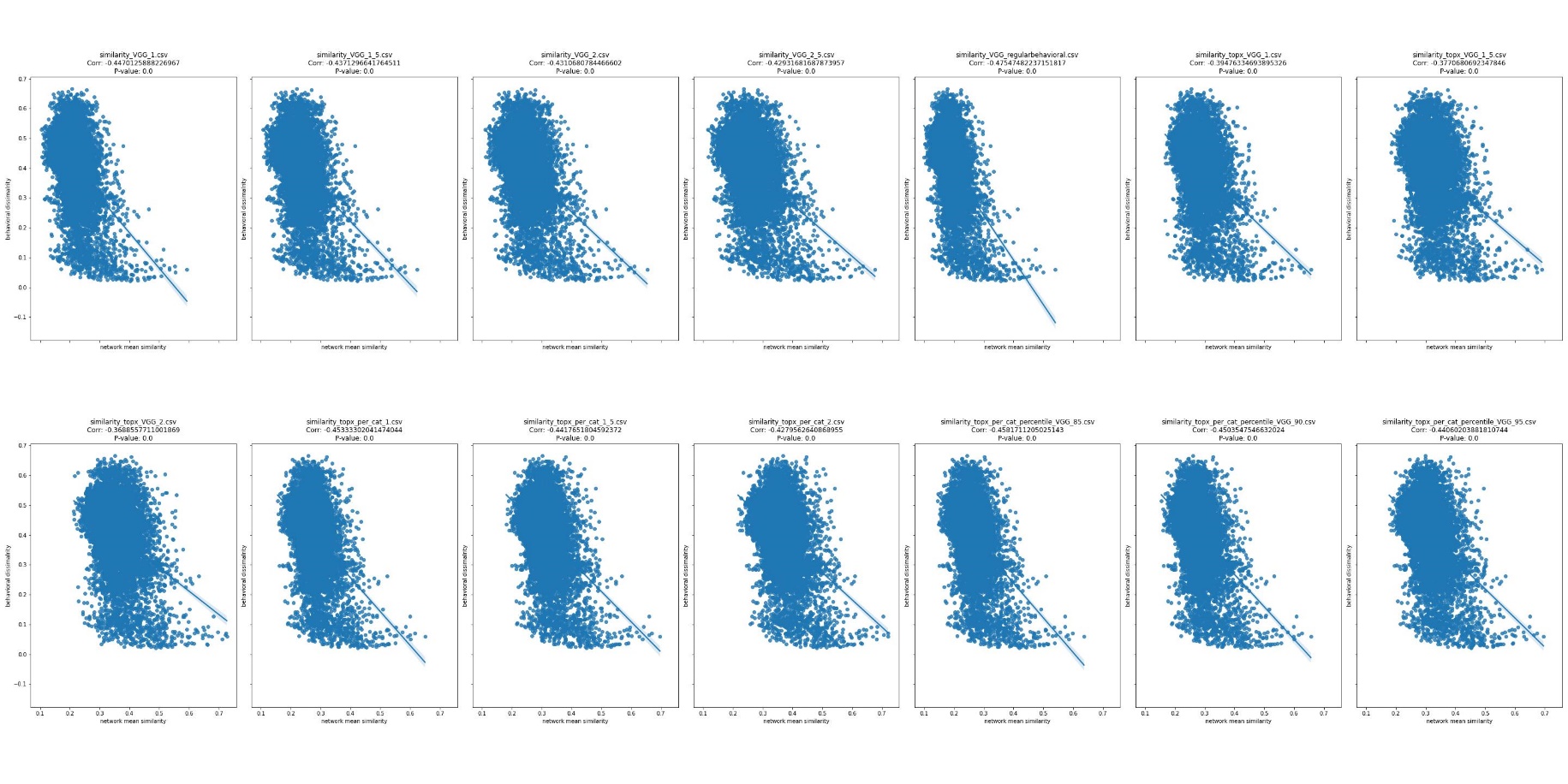
Due to multiple attempts creating different mean similarity trends for different subgroups of distractor images (target + layer 0, target + layer 3, target + layer 6) but unfortunately having no luck creating a non-significant different between the mean similarity of the groups, we started thinking about the way we defined similarity between two images.

Five different approaches were attempted:

1. **Threshold per pair:** for each pair of images, we calculate the mean similarity of each unit, normalize, and check which units pass this threshold to use for the similarity calculation. This approach results in different units being used for each similarity calculation.
2. **Most activated units over all the dataset:** we calculate the mean activation for all units in each layer over the whole dataset, normalize and use only the units that pass a threshold. This approach results in the same neurons being used for all similarity calculations.
3. **Most activated units per category with std:** we calculate the mean activation for all units in each layer over the dataset of each category separately. Per dataset we keep the index of the units that pass the threshold and use these units for the similarity calculation. This approach results in the same units being used per category.
4. **Most activated units per category with percentile:** we calculate the mean activation for all units in each layer over the dataset of each category separately. Per dataset we keep the index of the units that are the top percentile of activation and use these units for the similarity calculation. This approach results in the same units being used per category.
5. **Classic approach:** using all the activation vectors to calculate the similarity between all images.

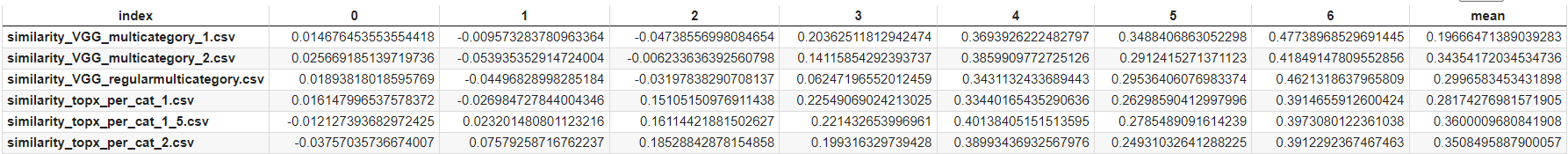
To evaluate the success of each approach we used two different methods:

1. **Correlation to human similarity judgments:** using behavioral similarity judgments from the [article](https://www.sciencedirect.com/science/article/pii/S1053811919303702?via%3Dihub) we calculate correlation between the behavioral similarity to each of the different approaches



1. **Check clustering accuracy using multi-dimensional scaling (MDS):**MDS is a statistical technique used to visualize and analyze the similarity or dissimilarity between objects based on a set of similarity or dissimilarity measures. In other words, MDS takes a set of objects and their pairwise distances or similarities and converts them into a set of coordinates in a lower-dimensional space (usually 2D or 3D) that preserves the relative distances between objects as much as possible. The resulting visualization is often referred to as an MDS plot or configuration.

MDS can be used as a pre-processing step to visualize the similarity or dissimilarity between objects, which can then be used to guide the clustering process. To use MDS for clustering, we use our different similarity measures from all different approaches and then use MDS to convert these distances or similarities into a set of coordinates in a lower-dimensional space. You could then use these coordinates as input to a clustering algorithm (e.g., k-means, hierarchical clustering, etc.) to group your images intoclusters**.**

To check what approach clusters better we use the Silhouette score. This score is a measure of how well each data point in a cluster is separated from other clusters. It ranges from -1 to 1, with a score closer to 1 indicating that the data point is well-matched to its own cluster and poorly matched to neighboring clusters.

Diagram

Description automatically generated