

Uncovering relationships in multi-channel EEG data using principal Hessian directions and Ricci flow

Benjamin J. Choi, Melanie Weber
Harvard University, Cambridge, MA USA

Keywords: network analysis, functional connectivity, EEG classification, principal Hessian directions, Ricci flow

Abstract

The high-dimensional nature of EEG data—especially multi-channel EEG—poses major challenges for downstream classification. Uncovering connectivity relationships and network structure between EEG channels can improve high-dimensional EEG signal recovery by identifying redundant inputs and guiding more targeted feature selection. In this paper, we present two complementary methods for inferring network structure and connectivity in EEG data: (1) a new supervised algorithm for inferring classification-relevant community structure based on principal Hessian directions (pHd), and (2) a discrete Ricci flow-based unsupervised community detection algorithm. We demonstrate these systematic methods on high-dimensional real-world EEG datasets involving classifying imagined digits versus non-digits and detecting emotional valence. We show that our pHd and Ricci flow methods—when combined—can detect interchannel relationships that meaningfully hold on unseen test classification data. Moreover, we demonstrate that this interchannel structure extracted via pHd and Ricci flow can enable subsequent improvements in downstream EEG signal recovery.

Note: supporting code stored in pHd-Ricci-flow GitHub repository.

1. Introduction

EEG data often exhibits high-dimensionality due to fast electrode sampling rates, substantial channel counts, and the time-intensive process of EEG data collection [1-3]. As a result, the number of features per sample in an EEG signal recovery context can far outstrip the number of samples themselves (i.e., $d >> n$) [4][5]. This high-dimensionality introduces several significant challenges in signal recovery and classification tasks. Overfitting, a well-documented problem when training classification models on neural data [6][7], can be particularly severe in high-dimensional settings; models can capture noise and spurious training data patterns, leading to poor generalization on unseen data [8]. This overfitting risk is even further exacerbated in the context of EEG, where the signal-to-noise ratio is already low due to various physiological and external artifacts [9-13]. Moreover, the redundant and highly-correlated features common in multichannel EEG data can obscure meaningful patterns, making it difficult to isolate the signals that are truly indicative of different cognitive states or conditions [14][15]. Model interpretability also suffers in high-dimensional settings [16-18], making it challenging to derive meaningful neuroscientific insights from the classification models. Ideal EEG classification systems not only perform well but also provide clear explanations of feature importance and interactions [19-21], which is difficult to achieve when dealing with thousands of (potentially collinear) features [17][18]. Uncovering structure in high-dimensional EEG data is thus crucial for many practical applications.

Beyond classical algorithms for general dimensionality reduction and feature selection like Principal Component Analysis (PCA) and minimum-redundancy-maximum-relevance (MRMR) [22][23], several studies have presented methods for creating effective embeddings of high-dimensional EEG by exploiting the underlying relationships between different EEG channels [24][25]. By detecting redundancy and collinearity among EEG channels while elucidating interchannel features carrying signal, developing methods to analyze network structure in EEG can lead to meaningful low-dimensional embeddings—and more robust downstream classification. In this vein, Behrouzi et al., 2022 and Choi, 2024 demonstrated that feeding graph embeddings of EEG data into downstream graph convolutional network (GCN)-based architectures led to superior test classification performance [26][27]. Similarly, Wang et al., 2022 showed that integrating graph topology information into EEG features led to performance increases on EEG classification tasks [28].

These recent strides in systematic, automated frameworks for graph-based EEG embeddings [26-28] build upon a significant body of literature on functional connectivity (FC) [29-35]. Traditional methods for inferring relationships between EEG channels exploit conventional hemispheric relationships [36-38]; for example, several studies present methods grounded in hemispheric asymmetry for informing classification model features [38-42]. However, traditional connectivity methods reliant on established EEG electrode placements may fail in cases of electrode displacement or dislocation, to which traditional EEG setups are highly susceptible [43][44]. Similarly, preprogrammed models that rely on fixed assumptions about EEG

relationships may overlook transient or task-specific patterns, which can be critical for understanding neural dynamics in real-world and clinical settings [45-47]. These shortcomings highlight the need for data-driven approaches that can adaptively capture connectivity structures.

Conventional data-driven methods for inferring functional connectivity in EEG include pairwise correlation [48], coherence [49], and Granger causality (in directed cases) [50]. While effective, these methods can fail to account for broader network topology in EEG and evolving dynamic patterns [51-54]—recent studies have shifted toward more elaborate algorithms for functional connectivity inference. For example, Pellegrini et al., 2023 proposed a flexible approach using beamforming for identifying EEG connectivity [53], while Xu et al., 2024 demonstrated EEG network-based preprocessing methods using a wavelet transform on musical EEG data [54]. These studies provide promising evidence for the utility of EEG connectivity and network analysis. Notably, we mark a distinction between versatile embedding or preprocessing methods designed to be applied to various settings and different downstream model architectures [55-57], and efforts aimed at developing state-of-the-art (SOTA) models themselves; the latter category includes many notable studies that focus on the development of SOTA neural network classification architectures rather than solely on upstream network analysis methods [58-60]. As opposed to building neural networks for EEG classification, the focus of this research is on demonstrating useful network analysis and embedding techniques.

In this vein, we propose a combined framework of two new methods for inferring network structure in EEG data. Notably, contrary to prior work relying on fixed interchannel assumptions [36-42], our methods are channel agnostic, enabling our framework to derive insights even in cases of electrode displacement or dislocation. Furthermore, rather than deriving general characterizations of overall channel behavior (as in conventional connectivity literature [48-50]), our novel combined framework is designed to generate insights that are useful to a specific downstream signal recovery problem. The two methods comprising our combined framework are delineated below.

First, we present a new application of the principal Hessian directions (pHd) method, first proposed by Li in 1992 [61], geared toward signal-relevant community detection among EEG channels. The pHd method, which identifies directions in the input space that exhibit significant curvature of a response surface, has been demonstrated on many applications involving dimensionality reduction, signal recovery, and visualization [61-64]. In this study, we extend the existing body of pHd literature by applying pHd to a novel community detection context. Our proposed method computes the pHd for each individual EEG channel with respect to the signal (here, in a classification setting) and evaluates the alignment of the principal Hessian direction across channels to identify signal-relevant communities. Though originally formulated for dimensionality reduction, we hypothesize that the pHd-based network analysis method presented herein can offer supplemental information that improves downstream signal recovery performance and can be applied systematically across different EEG preprocessing and connectivity contexts.

Second, emerging literature on graph processing [65-67] has highlighted the discovery of Ricci flow as a new SOTA technique for community detection that outperforms classical spectral clustering techniques [67-69]. Recent studies leveraging discrete analogs of the Ricci flow tailored toward graphs based on Ollivier’s notion of Ricci curvature have demonstrated promising results on applied clustering problems [67-71], and in this study, we present the first large-scale application of the Ricci flow for community detection on EEG data. Beyond inferring mere connectivity relationships, we anticipate that unsupervised partitioning of EEG channels into communities will enable the discovery of complementary structure in conjunction with the aforementioned pHd method. More specifically, by exploring where the pHd communities overlap with and depart from the Ricci flow communities, we can identify both redundant (in the overlapping case) and “subtly-informative” (in the departure case) interchannel relationships for signal recovery (as elaborated upon in *Section 3*). Our core hypothesis underlying the pHd-Ricci flow method is that it reveals interchannel relationships that are both (1) meaningful and (2) useful for downstream classification tasks; we conduct a rigorous validation of this hypothesis in *Section 3*.

2. Methods

We demonstrate our network analysis methods on two real-world EEG datasets: a small-scale EEG dataset on imagined digits (*MindBigData*, henceforth referred to as *MBD*) from Vivancos et al., 2022 [72] and a larger-scale EEG dataset on emotional valence (*SEED*) from Duan et al., 2013 [42, 73]. Both datasets are high-dimensional, and for the purposes of demonstrating our pHd method, center around classification tasks: we explore classifying imagined digits versus non-digits in a preliminary *MBD* setting, and later extend our methods to detect positive versus negative emotional valence in the *SEED* dataset.

2.1 Initial Demonstration on *MBD*

As a proof-of-concept, we first apply our proposed network analysis methods to imagined digit versus non-digit classification on small-scale 14-channel (Emotiv EPOC) data from Vivancos et al., 2022 [72]. This initial dataset was selected due to its high-dimensional nature and relative simplicity; we employed balanced digit and non-digit classes of two-second segments ($n=300$) with 14 channels at 128 Hz.

Here, we present a first-of-its-kind pipeline for implementing principal Hessian directions on multichannel EEG data. As our aforementioned notion of pHd alignment relies on the assumption of meaningful features with respect to the classification problem, we cannot leverage our initial data existing in the temporal domain as arbitrary timestamps on temporally-invariant signals convey no classification-relevant information. Thus, we apply a fast Fourier transform (FFT) to obtain a power spectrum (selecting only positive frequencies to avoid redundancy) for each channel in every sample. Since each individual EEG channel now conveys comparable class-relevant frequency data, we can viably apply our alignment notion. Note that all network

analysis methods are performed on a partitioned training dataset (in this case, with a 4:1 train-test split) in order to prevent data leakage.

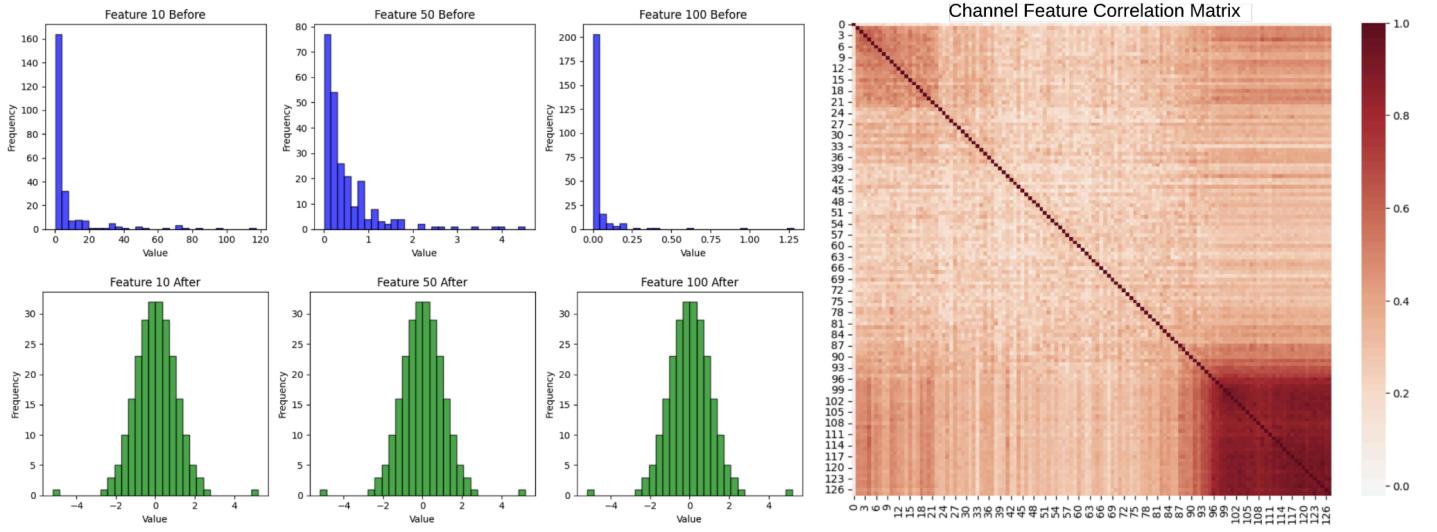


Figure 1: Power spectrum features before (upper left) and after Gaussianization (lower left); feature correlation matrix (right).

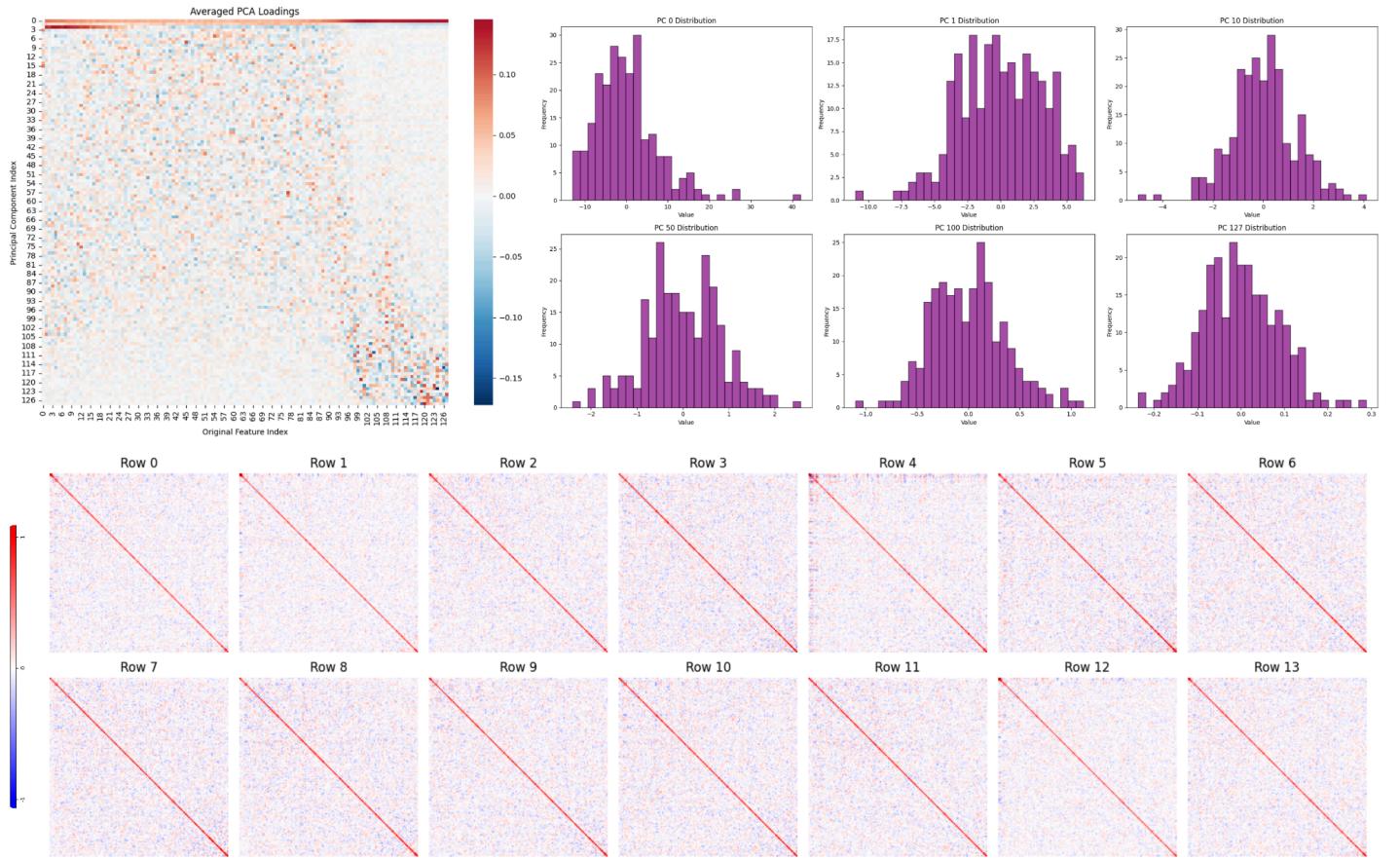


Figure 2: Averaged principal component loading matrix (upper left), feature distributions after principal component transformation (upper right), final feature correlation matrices for each EEG channel (bottom).

Notably, in order to leverage the classical technique with theoretical guarantees for estimating principal Hessian directions using Stein's lemma [61], we also require features to be independent and identically

distributed normal random variables with mean 0 and identity covariance. As shown in *Figure 1*, this is certainly not the case in our initial setting—power spectrum bins are not normally distributed across individual channels (*Figure 1*, left panel), nor are they empirically uncorrelated (*Figure 1*, right panel). We thus propose and demonstrate the following first-of-its-kind algorithm for pHd preprocessing in EEG data: (1) Gaussianize individual power spectrum features, (2) perform PCA with whitening on each individual channel, (3) aggregate and apply the same averaged set of PC loadings to transform every channel, and (4) renormalize the transformed features using identical cross-channel parameters with a new optimization based on the Kolmogorov-Smirnov normality test. This systematic process is designed to enable downstream application of conventional pHd estimation leveraging the theoretical guarantees of Stein’s lemma [61].

Initial Gaussianization is performed using a 100-bin quantile transformer on each power spectrum feature. While PCA could theoretically eliminate each channel’s cross-feature correlation shown in *Figure 1* (right panel), we cannot apply 14 different PC loadings to 14 different channels as this would disturb our goal to detect meaningful pHd alignment. Crucially, we must apply the same set of loadings to each channel so that the resulting PC-based features convey the exact same frequency information across different channels—otherwise, the pHd vectors would not have one-to-one comparable features. We rectify this concern by performing individual PCAs on each channel and then averaging the PC loadings (*Figure 2*, upper left panel) to form a single transformation matrix. This same averaged loading matrix is then applied to transform each EEG channel. While this does sacrifice the goal of perfectly uncorrelated features, we achieve our aim of ensuring meaningful comparability—and can observe empirically that this aggregated transformation yields approximately uncorrelated final features across each individual channel, as desired (*Figure 2*, lower panel).

However, we also empirically observe that the PC transformation has perturbed the gaussianity of our original features (*Figure 2*, upper left panel). Again, we cannot merely transform each channel’s features individually as our features now represent combinations of frequencies as determined by the PC loadings—each feature must undergo the same transformation across all 14 channels in order to preserve our alignment notion. To rectify this concern, we developed a new Kolmogorov-Sminrov (KS)-based optimization method.

Our KS-based method attempts to achieve approximate standard normality across PC-perturbed feature distributions by applying a Z-score transformation for each feature across all 14 channels. As mentioned, the algorithm must apply the same two transformation parameters (mean μ and standard deviation σ) across all 14 channels—meaning that tradeoffs between channels when it comes to normality are inevitable. In order to ensure a systematic process maximizing overall normality, we attempt to maximize the average KS normality test p-value across all 14 channels for each PC-transformed feature using the L-BFGS-B optimization algorithm as proposed by Zhu et al., 1997 [74]. As shown in *Figure 3* (upper panel), our optimization enables us to achieve a resulting feature set with relatively minimal statistically significant deviations from normality.

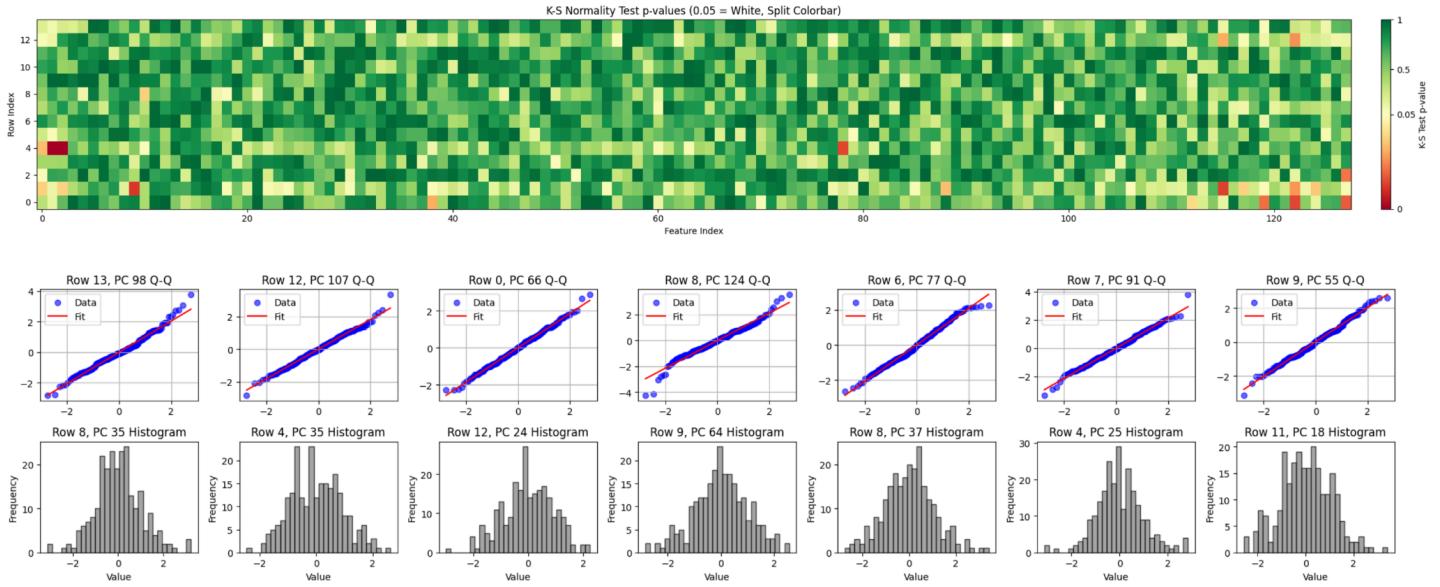


Figure 3: Results of the Kolmogorov-Smirnov normality optimization; p-values are plotted in the upper panel, with red denoting statistically significant deviations from normality while green denotes p-values above 0.05. The lower panel depicts quantile-quantile (Q-Q) normality plots and histograms for randomly selected features.

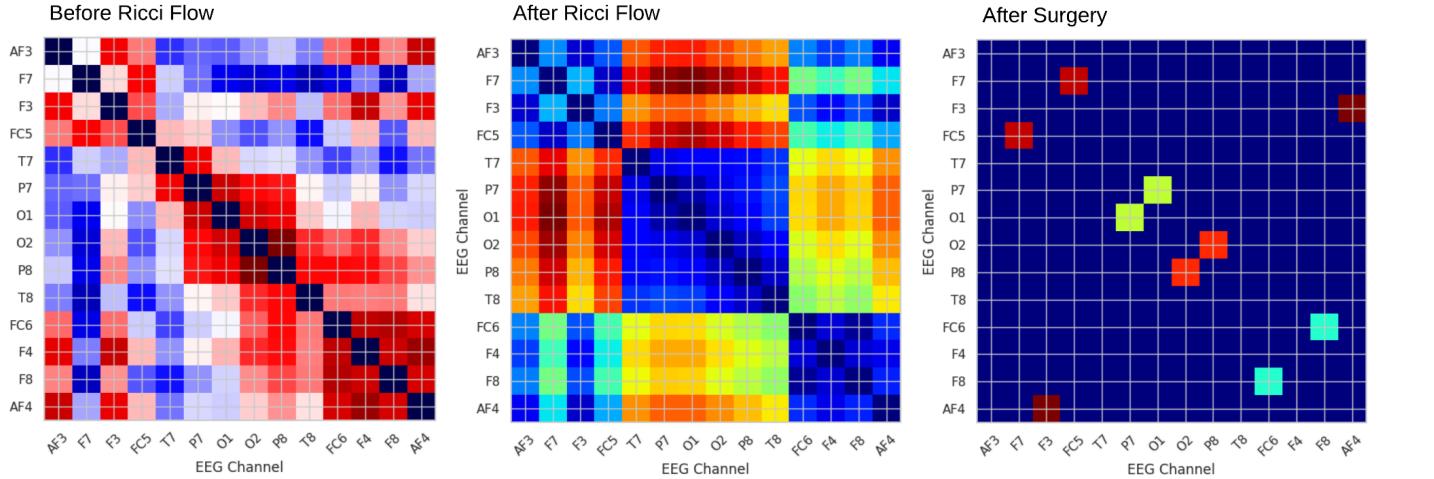
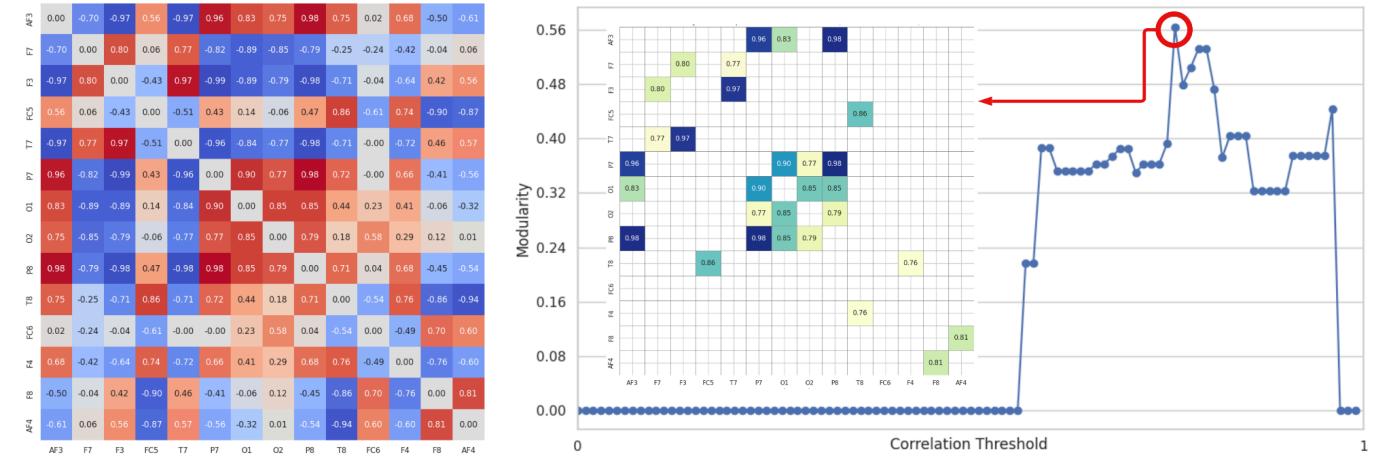


Figure 4: Results of both the principal Hessian directions-informed community detection algorithm (upper panels) and the Ricci flow-informed community detection algorithm (lower panels) on the initial 14-channel imagined digit versus non-digit dataset. The upper left panel depicts the pHd alignment matrix, while the upper right panel depicts the process of iteratively increasing the alignment (correlation) cutoff threshold and taking the adjacency matrix with maximum modularity. From left to right, the lower panels depict: (1) adjacency before Ricci flow, (2) adjacency after Ricci flow, and (3) partitioned communities after surgery.

After completing our feature transformation, we then apply the pHd method by computing the principal Hessian direction with respect to the classification problem: labels are set to +1 for imagined digits and -1 for imagined non-digits. To obtain alignment information, we performed the pHd method on each individual channel. Per convention, we multiplied labels with the outer product of each feature vector and took the leading eigenvector of the resulting matrix. We then computed the correlation between each EEG channel’s leading eigenvector to arrive at the final pHd alignment matrix for the dataset (*Figure 4*, upper left panel). To compute the pHd-informed community partitions, we started with this initial correlation-based adjacency matrix and iteratively increased the correlation cutoff value needed to maintain an edge between two channels. We then computed the modularity of all possible partitions and systematically took the set of communities with maximum modularity (as shown in *Figure 4*, upper right panel).

As discussed, we then shifted toward our complementary unsupervised community detection algorithm applying the Ricci flow. Building upon methods demonstrated in Sia et al., 2019 [67], Tian et al., 2023 [68], and Choi, 2024 [27], we used inter-channel correlation to initialize a weighted adjacency matrix and then applied the Ricci flow to the weighted graph using Ollivier’s notion of Ricci curvature [75]. We then evolved the weights of the graph using the Ricci flow and then performed “surgery” by cutting the graph’s heavily weighted edges [27, 67-68], using all unique cut ratios within two standard deviations of the total number of edges in order to obtain an exhaustive list of possible partitions while avoiding trivial scenarios (e.g., graphs with almost zero edges). We then obtained our final graph by selecting the community partition that maximized network modularity, as depicted in the bottom panel of *Figure 4*. Careful analyses and evaluations of the complementary pHd and Ricci flow community partitions are discussed in *Section 3*.

2.2 Higher-Dimensional Extension on *SEED*

After successfully demonstrating our network analysis methods on the 14-channel *MBD* case [72], we then applied our pHd and Ricci flow analyses to an even higher-dimensional setting: the 62-channel *SEED* dataset [42, 73], which we partitioned using a 9:1 train-test split. While our initial imagined digit versus non-digit demonstration involved two-second samples at 128 Hz, the *SEED* dataset enables network analysis across 130 seconds at 200 Hz. Our pHd-informed network analysis on *SEED* involves signal recovery related to positive versus negative emotional valence, with 450 total samples each containing 62-channel time series data obtained from 15 different subjects.

As with the 14-channel demonstration, we applied the same FFT, Gaussianization, and averaged PC loading pipeline. Notably, as the number of features per channel outstripped the total number of samples (13,000 versus 405), we reduced the dimensionality in the PC loading step to 405 frequency features before applying KS optimization. To further account for the high dimensionality of the dataset and the resulting high number of relatively uninformative features, we also harmonically weighted the pHd features when computing

correlation—features based on prominent PCs were weighted higher when determining channel alignment to account for the $d \gg n$ nature of the dataset. The resulting pHd-informed community detection process is depicted in the upper panel of *Figure 5*. As before, we also systematically applied the discrete Ricci flow algorithm for complementary unsupervised clustering, as depicted in *Figure 5* (lower panel). An evaluation of the combined network analysis and community detection on *SEED* is explored in *Section 3*.

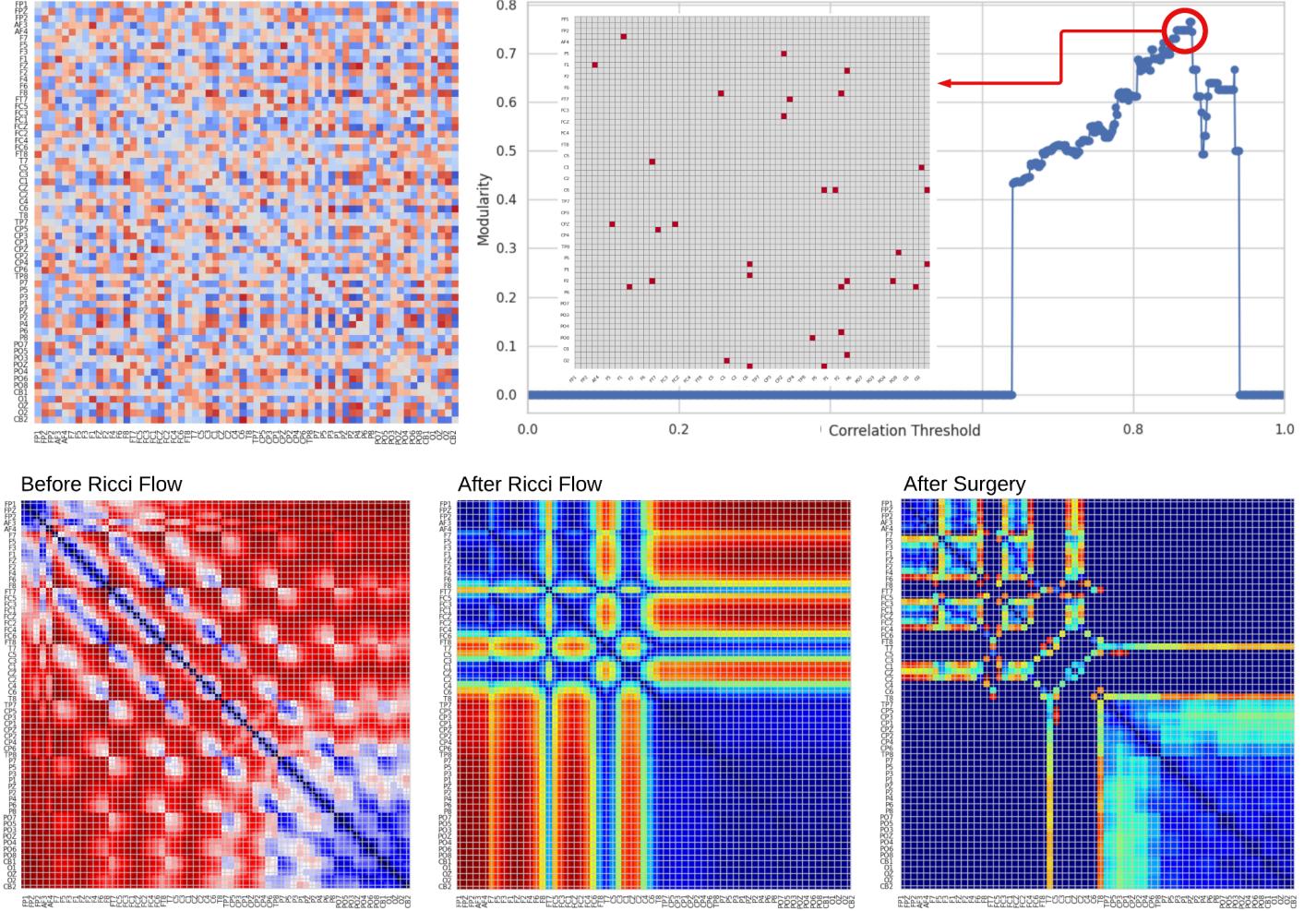


Figure 5: Results of both the principal Hessian directions-informed community detection algorithm (upper panels) and the Ricci flow-informed community detection algorithm (lower panels) on the 62-channel *SEED* dataset for emotional valence detection. The upper left panel depicts the pHd alignment matrix, while the upper right panel depicts the process of iteratively increasing the alignment (correlation) cutoff threshold and taking the adjacency matrix with maximum modularity. From left to right, the lower panels depict: (1) adjacency before Ricci flow, (2) adjacency after Ricci flow, and (3) partitioned communities after surgery.

3. Results

The pHd and Ricci flow-informed communities across all datasets and clustering methods were designed to maximize modularity in correspondence with conventional methods for optimal community detection [76-79]. Beyond merely providing interpretable insights into potential connectivity, however, we conduct a thorough demonstration of both the significance and downstream utility of our network analyses.

3.1 Significance Tests

Crucially, as previously outlined, our pHd and Ricci flow methods are built to be complementary. The unsupervised community detection enabled via Ricci flow allows for the identification of channel groups with similar outward characteristics, while the signal-relevant pHd method identifies groups of channels whose features interact similarly with the response signal. Thus, by identifying nodes that are in the same Ricci flow-informed group (indicating similarity in content) and in the same pHd-informed group (indicating similarity in how their content relates to the response), we can more conclusively identify pairs of nodes that are redundant—that is, nodes whose similar frequency content is undergirded by the redundant manner said frequencies interact with our signal of interest. Moreover, by applying our combined framework to identify nodes that belong to the same Ricci flow-informed group but not the same pHd-informed group, we can detect “subtly-informative” EEG channel relationships: channels that exhibit content-wise similarities but fundamentally relate differently to the downstream signal.

In accordance with the first component of our aforementioned hypothesis (*Section 1*), we conducted statistical tests to assess whether these pHd-Ricci identified redundant and subtly-informative relationships are indeed significant. In particular, we developed t -statistics to measure the significance of the (1) redundancy and (2) subtle-informativity of a given pair of EEG channels. On the redundancy front, we expect that if two channels are indeed redundant, averaging their feature content should less adversely affect downstream signal recovery compared to the comparable effect of averaging two randomly selected (presumably non-redundant) EEG channels. More precisely, we perform logistic regressions to assess the delta in test classification accuracy when regressing on both channels separately versus regressing on their averaged features. We quantify redundancy via the statistic defined as the averaged regression test accuracy minus the separate regression accuracy, normalized by the separate regression accuracy. These regressions are performed only on the selected channels (as opposed to the entire channel set) to isolate the nature of that specific relationship. Greater scores on this statistic correspond to greater redundancy (i.e., stronger performance of the averaged regression relative to the separate regression), while more negative scores indicate lower redundancy. On the subtle-informativity front, since the shared Ricci flow group and differing pHd group indicate a shared content structure but distinct interactions between each channel and the downstream signal, we hypothesized that the information encoded in the difference between the two channels would prove relatively more informative than averaging the two channels (since taking the difference effectively shifts the focus from the shared structure to the channels’ distinct signal-related components). We thus construct a new statistic corresponding to subtle-informativity defined as the test accuracy of regressing on the channel difference minus the test accuracy of regressing on the averaged channels, normalized by the separate regression accuracy; higher scores correspond to greater subtle informativity. For both our redundancy and subtle-informativity test statistics, we conducted randomization tests to determine whether the test statistics computed for our pHd-Ricci-identified channel relationships on both

MBD and *SEED* significantly exceed those obtained under a null hypothesis. Results are displayed below in *Figure 6*.

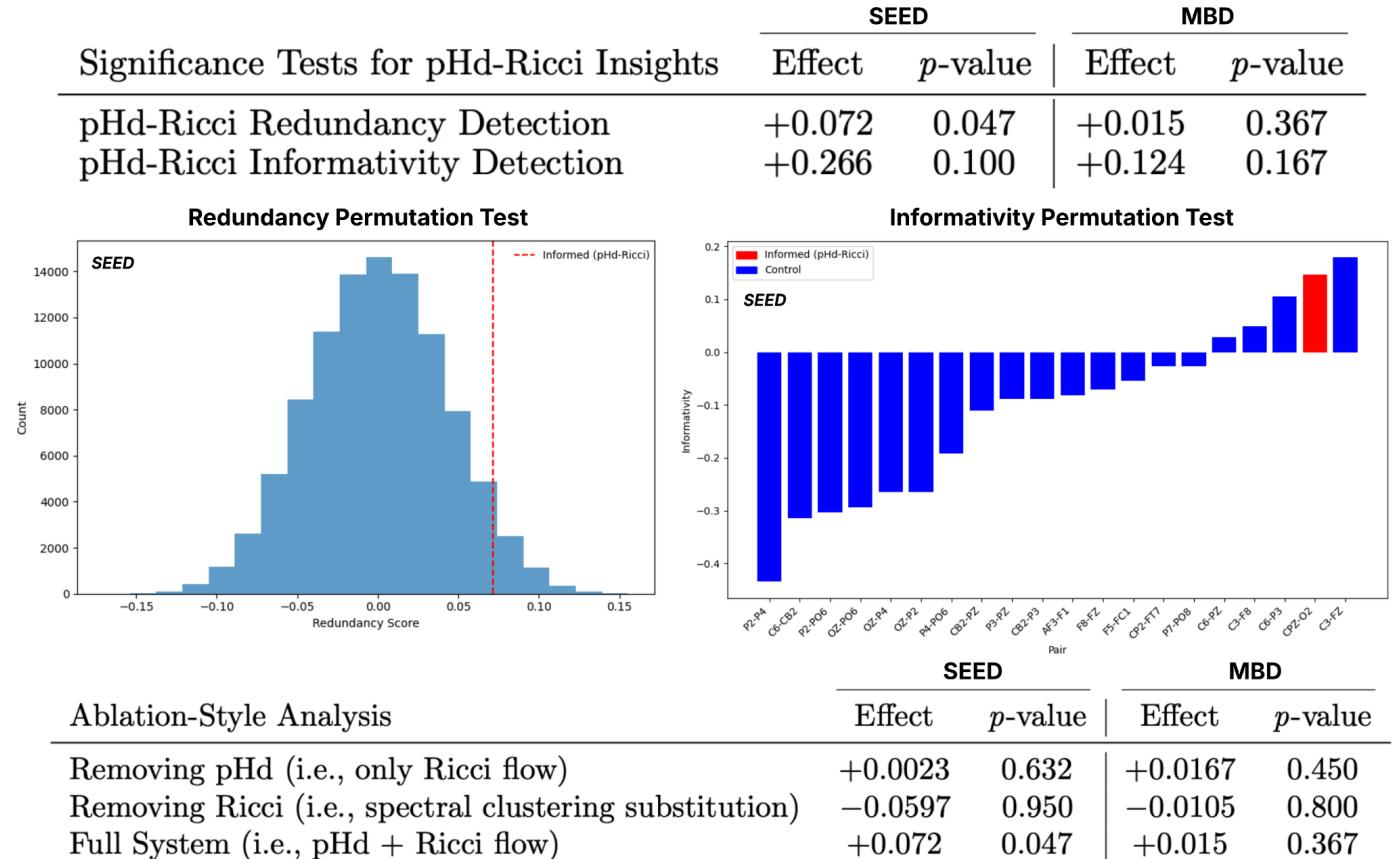


Figure 6: Results of significance tests and analyses for the pHd-Ricci-identified channel relationships. Top panel: observed effects and corresponding *p*-values for redundancy and subtle-informativity detection across *SEED* and *MBD* datasets. Middle left: null distribution of redundancy scores from control group permutations on *SEED*, with the observed pHd-Ricci-informed score given by the dashed red line. Middle right: subtle-informativity scores on *SEED* for both the control relationships (in blue) and the pHd-Ricci-informed relationship (in red). Lower panel: observed effects and corresponding *p*-values for ablation-style analyses on redundancy detection (excluding pHd, substituting out Ricci flow, and comparing with the full system).

By directly applying the results of our systematic pHd and Ricci flow analyses (from *Section 2*) in accordance with our predefined redundant and subtly-informative identification procedures, we identified two pairs of redundant nodes and two pairs of subtly informative nodes on *MBD* (see *Figure 7* for an illustration); our analysis on *SEED* identified nineteen redundant channel relationships and two channels with a subtly-informative relationship. To ensure a consistent identification procedure, we identified redundant nodes first before examining subtly-informative pairings; reference groups were given by the multi-node clusters obtained via the final (*Section 2*) partitions. For our significance tests, we quantify the overall “effect” score as the difference between the observed mean t-statistics under the alternative hypothesis (that the pHd-Ricci analysis identifies genuinely meaningful relationships) and the mean t-statistics expected under the null hypothesis (generated via randomization). We also provide an associated *p*-value—the proportion of randomizations for which the randomized effect score meets or exceeds the observed effect. We avoid

instituting an arbitrary critical value threshold [81-83] with respect to our use of p -values; we instead offer context in *Section 4* on the particular limitations and bounds on these p -values to allow them to speak for themselves within the broader context of effect interpretation and experimental uncertainty. (It is a positive “effect” score—not an arbitrarily-low p -value—that measures whether the detected relationships exhibit characteristics in alignment with the predetermined hypotheses.) For the redundancy significance tests, the initial control group is constructed to match the treatment (i.e., pHd-Ricci-identified) group size by randomly selecting channel pairs, with the ensuing p -value generated via random label permutation; control groups were exhaustively randomized on *MBD* and randomized 100,000 times on *SEED*. Note that since subtle-informativity is only defined relative to redundancy (i.e., subtly-informative pairs share structural similarity—same Ricci group—but are not hypothesized to be redundant with respect to signal recovery), our subtly-informativity significance test uses the identified redundant pairs as the control group; p -values again correspond to label randomization. Overall, we obtain positive effect scores (on both metrics and on both datasets)—i.e., the pHd-Ricci identified pairs did have higher mean scores on the redundancy and subtle-informativity t -statistics relative to the control groupings; these results are displayed in *Figure 6* and further discussed in *Section 4*.

Notably, we also conducted an ablation-style analysis to justify the pHd and Ricci flow components of our combined framework. Note that as opposed to a classical ablation, we cannot simply re-run an analysis sans-Ricci, since the pHd method alone does not provide a measure of redundancy or informativity—it only groups nodes with similar feature-signal relationships. However, we can assess whether nodes in the same Ricci flow group exhibit meaningful redundancy even without pHd corroboration; if not, this would justify the added value of incorporating pHd into our framework. Likewise, by comparing Ricci flow groupings to those produced by conventional spectral clustering [84-85], we can evaluate whether Ricci flow offers comparative utility in terms of redundancy-related insights—thus providing support for the use of Ricci flow’s curvature-based clustering approach in our overall framework. Our ablation-style analyses are focused on redundant relationship identification; we note that the detection of subtle informativity, by definition, requires both pHd and Ricci flow (and thus inherently does not exist if we exclude either component).

In our assessment of Ricci flow alone for redundancy detection, we evaluated pairs within the same Ricci group (without pHd corroboration) against a random-pair control group; results from this test are compared with the Ricci flow *and* pHd system to shed light on whether pHd inclusion is justified. We also evaluated traditional spectral clustering [84-85] as a direct substitute for our Ricci flow-based clustering in order to justify our use of Ricci flow itself. To reduce ambiguity in cluster selection, we used both the eigengap heuristic and modularity maximization; both approaches suggested the same optimal number of clusters for both *SEED* and *MBD*. Results from these analyses are displayed in *Figure 6* and discussed in *Section 4*.

3.2 Downstream Utility Tests

In addition to the aforementioned set of significance tests and analyses, we also conducted applied tests to show that our combined pHd-Ricci flow system offers utility in downstream signal recovery cases. Specifically, we aimed to verify whether the channel relationship insights (and subsequent feature generation) generated by the combined pHd-Ricci framework could improve test classification performance on *SEED* and *MBD*. In accordance with the identified relationships discussed in *Section 3.1*, we applied a systematic predetermined framework for incorporating pHd-Ricci insights into downstream embeddings; we transformed the identified subtly-informative channel pairs into new features describing the difference between the two channels in frequency space, and transformed redundant nodes by averaging their frequency content. We then evaluated the hypothetical downstream impact of these more robust initial embeddings on overall test classification performance.

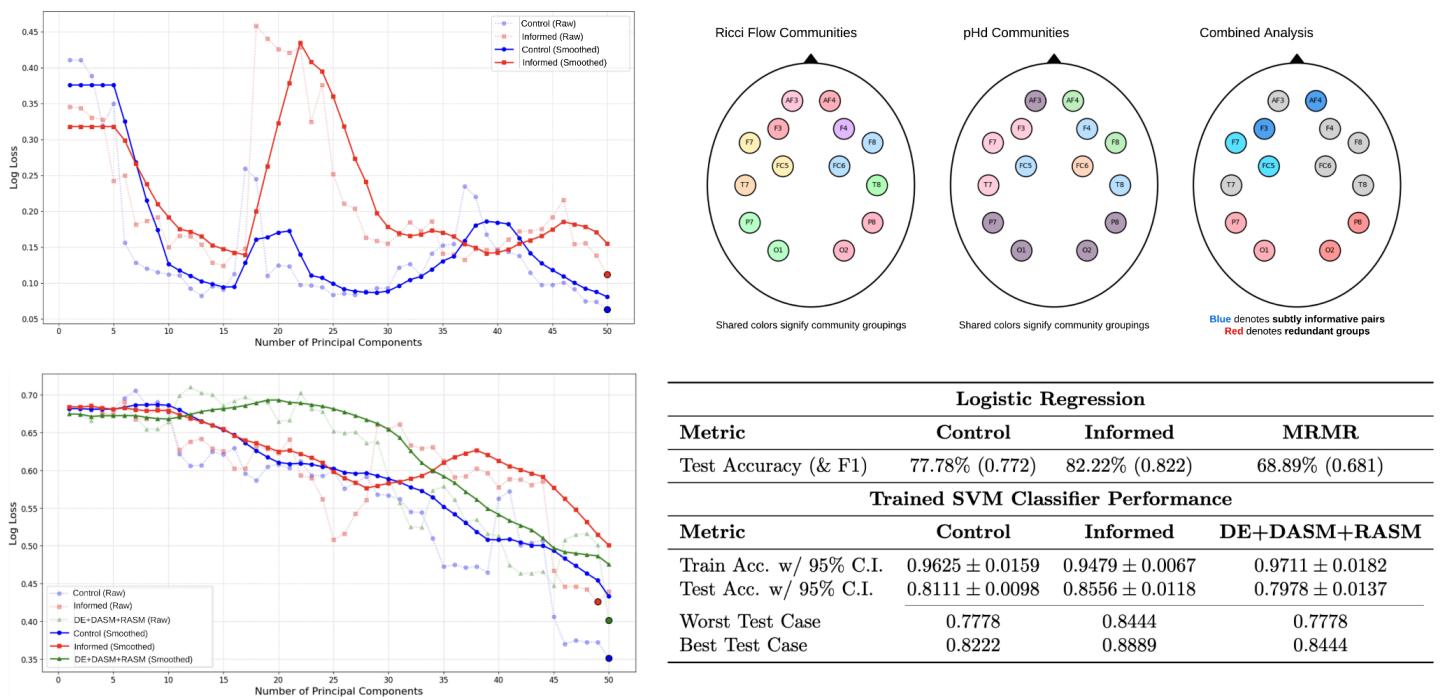


Figure 7: Results evaluating the impact of our pHd-Ricci-driven insights on downstream signal recovery performance across both datasets (*MBD* in the upper panels, *SEED* in the lower panels). The upper left panel depicts a performance (test loss) comparison between a pHd-Ricci-informed PC logistic regression (blue) and a control PC logistic regression (red). The upper right panel offers a visualization of our Ricci flow- and pHd-informed communities (in the 14-channel case) and the resulting subtly informative and redundant relationships. A PC logistic regression on *SEED* is depicted in the lower left panel, with three depicted performance curves: control (red), informed (blue), and the conventional hemispheric DE+DASM+RASM method (green). SVM classification test results using the control, informed, and DE+DASM+RASM features are depicted in the lower right panel.

As displayed in *Figure 7*, we conducted multiple careful analyses across both datasets to ensure that the derived relationships—beyond mere potential significance—could have a positive impact on signal recovery. On the imagined digit dataset, we conducted two logistic regressions—one with the pHd and Ricci-informed features (reducing redundant and subtly-informative rows), and one without as a control. We found the control regression achieved a mean test log loss of 0.228 with standard error (SE) 0.144 (corresponding to a test accuracy of 95.0%), while the informed regression achieved a mean test log loss of 0.0474 with SE 0.0181 (and

perfectly classified all test cases). Given the high dimensionality of the dataset, however, we also conducted principal component logistic regressions based on the informed vs. control feature sets (both with frequency domain features), plotting mean log loss on the test set versus the number of PC features used in the regression. As visualized in the upper left panel of *Figure 7*, we found that the pHd and Ricci flow-informed logistic regression demonstrated more stable test performance as the number of PC features increased compared to the control. The best global performance (marked by a colored dot in *Figure 7*, upper left panel) was also superior for the informed PC regression compared to the control.

To corroborate these preliminary results achieved on the simpler *MBD* case, we conducted multiple analyses of the utility of the pHd- and Ricci flow-informed network analysis on the *SEED* dataset. Our initial logistic regression test revealed that our pHd-Ricci-informed preprocessing method led to superior accuracy compared to both an uninformed control regression and a minimum-redundancy-maximum-relevance (MRMR)-informed regression (see *Figure 7*, lower right panel). The expanded channel count in *SEED* also enabled comparison with the highly-cited differential entropy (DE), differential asymmetry (DASM), and rational asymmetry (RASM) method for preprocessing [42]. In particular, to further assess the utility of our pHd- and Ricci-flow informed analysis and preprocessing, we trained ten kernel SVMs (via grid-searching identical parameter grids) on *SEED* on each of the three feature sets: control, informed, and DE+DASM+RASM. We found that our informed preprocessing method outperformed the other two methods by a statistically significant margin (*Figure 7*, lower right panel). In addition, when applied to the aforementioned PC logistic regression test, we found that the pHd- and Ricci flow-informed method again achieved the best global performance and empirically demonstrated relatively strong performance across PCs compared to the control and DE+DASM+RASM alternatives (*Figure 7*, lower left panel). Further implications of the relative efficacy of our pHd-Ricci framework are discussed in *Section 4*.

4. Discussion

We present a novel application of the principal Hessian directions method for community detection on EEG signals. It is important to note, however, that our use of pHd involves approximations and tradeoffs related to covariance and gaussianity; while we validate our method via the aforementioned studies on the utility of the pHd-informed network analysis, further testing will be required to determine the position of pHd within broader bodies of neural engineering and community detection literature. While our pHd application pipeline has the advantage of being systematic, we note that the principal component vector computation may become intensive in settings with exceedingly large numbers of samples and features; they are, however, relatively suitable for high-dimensional applications with small n and thus fewer computable PCs. Further generalization to broader settings may require setting PC cutoffs via eigenvalue scree plots or iterative eigenvector estimation methods. The K-S optimization method used for preprocessing also introduces a potential bottleneck in large settings via

its use of L-BFGS-B [74], as the number of features needed to be optimized is equivalent to the length of a channel and must be optimized over the number of channels. Follow-up studies may explore shortcut heuristics for computing constant normalization parameters across channels (e.g., taking simple or weighted averages of μ and σ parameters across channels).

Given that Ricci flow-based clustering algorithms have been shown to outperform SOTA clustering algorithms in recent literature [67-69], our use of Ricci flow for EEG network analysis is designed to be both a first-of-its-kind extended application and an intentional taxonomic choice for providing unsupervised clustering information in conjunction with pHd. As our Ollivier-Ricci curvature-based clustering algorithm is computed once across a single correlation-initialized graph (with node count equal to EEG channel count), the use of Ricci flow described herein is unlikely to prove prohibitively intensive for EEG applications, which rarely exceed 256 channels. However, we do note that the complexity of the Ricci flow community detection algorithm is typically quadratic due to the Sinkhorn-based computation [80] of Wasserstein distance for Ollivier Ricci curvature [68]; if necessary, potential linear-time approximations can be implemented as discussed in Tian et al., 2023 [68]. Notably, (as shown in *Section 3*) we also conduct an ablation-style analysis in which we validate our choice of Ricci flow relative to spectral clustering. We find that the Ricci flow method posted positive effect scores on both *SEED* and *MBD* (*Figure 6*)—indicating that the Ricci flow-identified relationships exhibited more redundancy than randomly selected relationships—while traditional spectral clustering posted negative effect scores on both *SEED* and *MBD*. Thus, beyond recent studies showing favorable performance of curvature-based clustering methods (i.e., Ricci flow) [67-68], we also observe empirical support for favoring Ricci flow over spectral clustering for structural node grouping. However, while these results may help support the relative use of Ricci flow over spectral clustering for helping identify meaningful structure, we note that we do not find compelling evidence that Ricci flow *in isolation* provides meaningful value in terms of redundancy identification given its corresponding *p*-values on *SEED* and *MBD* of 0.632 and 0.450; the overall effect score on *SEED*, at +0.0023, is also relatively weak. However, we do find that Ricci flow *combined with pHd* does yield effect scores with greater statistical power (*Figure 6*), thereby helping justify the contribution of pHd to our overall system. More specifically, on the full system significance test for redundancy detection, the combined pHd-Ricci framework posted positive effect scores of +0.072 on *SEED* (with *p*-value 0.047) and +0.015 on *MBD* (with *p*-value 0.367). We also find it particularly crucial to contextualize the *p*-values and statistical powering enabled by the *MBD* significance test. While the *SEED* redundancy significance tests—owing to the large number of potential channel relationships—are conducted across 100,000 permutations and therefore enable more granular *p*-values, *MBD*'s limited channel count means that we are extremely limited in the number of randomizations we can perform. For example, as opposed to the 38-channel treatment and control groups in *SEED*, our *MBD* randomization test groups contain only two pairs each; given the noise inherent in our logistic regression-style test, weaker power with our *MBD* tests is thus expected owing

to these small sample sizes (à la the Law of Large Numbers). In particular, the more restrictive *MBD* setting (relative to *SEED*) also places theoretical limitations on how low the resulting *p*-values can be. Notably, for the subtle-informativity significance test, we found that our combined pHd-Ricci framework produced positive informativity effect scores on both *SEED* and *MBD* (+0.266 and +0.124, respectively) with *p*-values of 0.1 and 0.167; as the *p*-values are derived via ranking pairs, *MBD p*-values lower than 0.167 are not possible given that our test is based off four permutable relationships (two of which are control). That is, the two pHd-Ricci identified pairs on *MBD* were more informative than the two control pairs (which occurs under random conditions with $\frac{1}{6}$ probability, yielding the lowest possible *p*-value of 0.167); on *SEED*, the pHd-Ricci-identified pair ranked higher on our informativity metric than eighteen of nineteen control pairs (yielding a *p*-value of 0.1). Thus, while our *p*-values are computed to offer supplemental insight, proper context is vital given that the noise and absolute lower-bound *p*-values are high to begin with. Overall, through our *Section 3.1* significance tests, we find our pHd-Ricci method yields uniformly positive effect scores across both datasets and relationship types, providing support for the notion that our framework detects meaningful relationships in alignment with our predetermined hypotheses.

We shed further light on these positive results on our significance test via our *Section 3.2* analyses of the downstream utility of our pHd-Ricci-derived information. As mentioned in the *Introduction*, we emphasize that the objective of this study is not to build a SOTA classification model or to overperform EEG benchmarks for specific classification tasks. Rather, we aim to demonstrate useful preprocessing and analysis methods that can be used across diverse applications or combined with a variety of SOTA classifiers downstream. Our preprocessing methods are intentionally light and avoid thoroughly transforming features (instead averaging doubly-redundant channels, transforming subtly-informative relationships into differential features, etc.) in order to provide breathing room for subsequent combination with use-case-specific classification techniques. By using our identified pHd-Ricci relationships to systematically inform downstream feature creation and selection, we showed (see *Section 3.2* and *Figure 7*) consistently positive improvements in downstream signal recovery performance compared to both uninformed techniques and traditional feature selection techniques, including the redundancy-based MRMR and the relationship-informativity-based DE+DASM+RASM. These *Section 3* significance and utility results across datasets, relationships, and methods help undergird the efficacy of our pHd-Ricci framework.

Overall, through this study, we hope to expand the toolkit of EEG preprocessing methods by providing a systematic analytical framework for classification-relevant network analysis. Though outside of the scope of this demonstration, we also note that as our derived relationships are grounded in terms of well-defined EEG channels, our methods confer potential advantages in the realm of interpretability. As opposed to methods like PCA, which may obscure direct interpretation of channel contributions by aggregating information into principal components, we instead establish plain, interpretable relationships that elucidate both channel-specific

similarities and the roles of individual nodes in conveying critical information. This explainability could potentially foster positive interactions with downstream regression coefficients; as an aside, we noted several shifts in regression coefficient weighting between control and informed models during testing. For example, FC6 features were heavily weighted in the 14-channel control model, while weights were more evenly distributed across transformed channels in the pHd- and Ricci flow-informed model (with the subtly-informative F3-AF4 relationship weighted strongest). It is crucial, however, to caveat this observation—there are strong limitations to coefficient interpretability given collinearity [86], and we make no formal assertions. Follow-up studies combining our methods with downstream explainable machine learning techniques could shed light on any hypothetical interpretability advantages. Future work could also explore more elaborate techniques to combine and exploit pHd alignment information beyond the systematic cluster-focused approach explored in this work (e.g., via developing a more elaborate composite score-based approach to weighting relationships), but we hope our methods presented herein offer a systematic starting point.

We emphasize that while our study demonstrates potentially useful network analysis and preprocessing techniques for EEG data, results are preliminary and there are important limitations. While we evaluate our data on both 14-channel and 62-channel data across 45 subjects, further studies involving out-of-laboratory data collection will be necessary to assess whether the pHd- and Ricci flow-based methods remain useful in noisy real-world EEG conditions. Furthermore, collecting EEG data from human subjects across longer timeframes may also reveal whether the derived network relationships remain persistent or need to be continuously reevaluated. Moreover, we have no evidence on how these methods would perform in non-EEG neural signal recovery settings (e.g., fMRI or fNIRS data), nor do we investigate their effectiveness in cross-modal tasks involving the integration of multiple neural data types.

References

1. Teplan, Michal. "Fundamentals of EEG measurement." *Measurement Science Review* 2.2 (2002): 1-11.
2. Gevins, Alan, et al. "High resolution EEG: 124-channel recording, spatial deblurring and MRI integration methods." *Electroencephalography and Clinical Neurophysiology* 90.5 (1994): 337-358.
3. Light, Gregory A., et al. "Electroencephalography (EEG) and event-related potentials (ERPs) with human participants." *Current Protocols in Neuroscience* 52.1 (2010): 6-25.
Hemmelmann, Claudia, et al. "Multivariate tests for the evaluation of high-dimensional EEG data." *Journal of Neuroscience Methods* 139.1 (2004): 111-120.
4. Hemmelmann, Claudia, et al. "Multivariate tests for the evaluation of high-dimensional EEG data." *Journal of Neuroscience Methods* 139.1 (2004): 111-120.

5. Liang, Zhen, et al. "EEGFuseNet: Hybrid unsupervised deep feature characterization and fusion for high-dimensional EEG with an application to emotion recognition." *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 29 (2021): 1913-1925.
6. Saqib, Mohammed, et al. "Regularization of deep neural networks for EEG seizure detection to mitigate overfitting." *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*. IEEE, 2020.
7. Zhou, Songchi, et al. "Less is More: Reducing Overfitting in Deep Learning for EEG Classification." *2023 Computing in Cardiology (CinC)*. Vol. 50. IEEE, 2023.
8. Hosseini, Mahan, et al. "I tried a bunch of things: The dangers of unexpected overfitting in classification of brain data." *Neuroscience & Biobehavioral Reviews* 119 (2020): 456-467.
9. Zhang, Haoming, et al. "A novel convolutional neural network model to remove muscle artifacts from EEG." *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021.
10. Choi, Benjamin, et al. "Targeted adversarial denoising autoencoders (TADA) for neural time series filtration." 2025. Accepted at AI4TS at AAAI 2025.
11. Übeyli, Elif Derya. "Statistics over features: EEG signals analysis." *Computers in Biology and Medicine* 39.8 (2009): 733-741.
12. Choi, Benjamin. "Removing neural signal artifacts with autoencoder-targeted adversarial transformers (AT-AT)." Accepted at CNS 2025.
13. Leite, Niago Moreira Nobre, et al. "Deep convolutional autoencoder for EEG noise filtering." *2018 IEEE international conference on bioinformatics and biomedicine (BIBM)*. IEEE, 2018.
14. McFarland, Dennis J. "Characterizing multivariate decoding models based on correlated EEG spectral features." *Clinical Neurophysiology* 124.7 (2013): 1297-1302.
15. Erguzel, Turker Tekin, et al. "Binomial logistic regression and artificial neural network methods to classify opioid-dependent subjects and control group using quantitative EEG power measures." *Clinical EEG and neuroscience* 50.5 (2019): 303-310.
16. de Mijolla, Damien, et al. "Human-interpretable model explainability on high-dimensional data." *arXiv preprint arXiv:2010.07384* (2020).
17. Deng, Wei Q., Radu V. Craiu, and Lei Sun. "Measuring the severity of multi-collinearity in high dimensions." *arXiv preprint arXiv:2203.10360* (2022).
18. Drobnič, Franc, Andrej Kos, and Matevž Pustišek. "On the interpretability of machine learning models and experimental feature selection in case of multicollinear data." *Electronics* 9.5 (2020): 761.

19. Mortaga, Maged, Alexander Brenner, and Ekaterina Kutafina. "Towards interpretable machine learning in EEG analysis." *German Medical Data Sciences 2021: Digital Medicine: Recognize–Understand–Heal*. IOS Press, 2021. 32-38.
20. Sturm, Irene, et al. "Interpretable deep neural networks for single-trial EEG classification." *Journal of neuroscience methods* 274 (2016): 141-145.
21. Jemal, Imene, et al. "An interpretable deep learning classifier for epileptic seizure prediction using EEG data." *IEEE Access* 10 (2022): 60141-60150.
22. Abdi, Hervé, and Lynne J. Williams. "Principal component analysis." *Wiley interdisciplinary reviews: computational statistics* 2.4 (2010): 433-459.
23. Bugata, Peter, and Peter Drotar. "On some aspects of minimum redundancy maximum relevance feature selection." *Science China Information Sciences* 63.1 (2020): 112103.
24. Tafreshi, Taban Fami, Mohammad Reza Daliri, and Mahrad Ghodousi. "Functional and effective connectivity based features of EEG signals for object recognition." *Cognitive neurodynamics* 13.6 (2019): 555-566.
25. Almohammadi, Abdullah, and Yu-Kai Wang. "Revealing brain connectivity: graph embeddings for EEG representation learning and comparative analysis of structural and functional connectivity." *Frontiers in Neuroscience* 17 (2024): 1288433.
26. Behrouzi, Tina, and Dimitrios Hatzinakos. "Graph variational auto-encoder for deriving EEG-based graph embedding." *Pattern Recognition* 121 (2022): 108202.
27. Choi, Benjamin. "Geometric machine learning on EEG signals." Accepted in *Proceedings of Machine Learning Research* (2024).
28. Wang, Huiyang, Hua Yu, and Haixian Wang. "EEG_GENet: A feature-level graph embedding method for motor imagery classification based on EEG signals." *Biocybernetics and Biomedical Engineering* 42.3 (2022): 1023-1040.
29. Chang, Catie, et al. "EEG correlates of time-varying BOLD functional connectivity." *Neuroimage* 72 (2013): 227-236.
30. Miljevic, Aleksandra, et al. "Alterations in EEG functional connectivity in individuals with depression: A systematic review." *Journal of Affective Disorders* 328 (2023): 287-302.
31. Lee, You-Yun, and Shulan Hsieh. "Classifying different emotional states by means of EEG-based functional connectivity patterns." *PloS one* 9.4 (2014): e95415.
32. Xie, Wanze, Russell T. Toll, and Charles A. Nelson. "EEG functional connectivity analysis in the source space." *Developmental Cognitive Neuroscience* 56 (2022): 101119.
33. Chu, Catherine Jean, et al. "EEG functional connectivity is partially predicted by underlying white matter connectivity." *Neuroimage* 108 (2015): 23-33.

34. Rojas, Gonzalo M., et al. "Study of resting-state functional connectivity networks using EEG electrodes position as seed." *Frontiers in neuroscience* 12 (2018): 235.
35. Barzegaran, Elham, and Maria G. Knyazeva. "Functional connectivity analysis in EEG source space: the choice of method." *PLoS One* 12.7 (2017): e0181105.
36. French, Christopher C., and J. Graham Beaumont. "A critical review of EEG coherence studies of hemisphere function." *International Journal of Psychophysiology* 1.3 (1984): 241-254.
37. Tucker, D. M., D. L. Roth, and T. B. Bair. "Functional connections among cortical regions: topography of EEG coherence." *Electroencephalography and clinical neurophysiology* 63.3 (1986): 242-250.
38. Li, Yang, et al. "A novel neural network model based on cerebral hemispheric asymmetry for EEG emotion recognition." *IJCAI*. 2018.
39. Pei, Guanxiong, et al. "Mixed Emotion Recognition Based on EEG Signals." *2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE, 2023.
40. Zheng, Fa, et al. "EEG emotion recognition based on hierarchy graph convolution network." *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2021.
41. Li, Rui, Le-Dian Liu, and Bao-Liang Lu. "Discrimination of decision confidence levels from EEG signals." *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2021.
42. Duan, Ruo-Nan, Jia-Yi Zhu, and Bao-Liang Lu. "Differential entropy feature for EEG-based emotion classification." *2013 6th international IEEE/EMBS conference on neural engineering (NER)*. IEEE, 2013.
43. Sommer, Bjoern, et al. "Investigation of subdural electrode displacement in invasive epilepsy surgery workup using neuronavigation and intraoperative MRI." *Neurological Research* 40.10 (2018): 811-821.
44. Wang, Yunhua, and Jean Gotman. "The influence of electrode location errors on EEG dipole source localization with a realistic head model." *Clinical Neurophysiology* 112.9 (2001): 1777-1780.
45. Hunyadi, Borbála, et al. "A dynamic system of brain networks revealed by fast transient EEG fluctuations and their fMRI correlates." *Neuroimage* 185 (2019): 72-82.
46. Mishra, Sudhakar, Narayanan Srinivasan, and Uma Shanker Tiwary. "Dynamic functional connectivity of emotion processing in beta band with naturalistic emotion stimuli." *Brain sciences* 12.8 (2022): 1106.
47. Huang, Yinghao, et al. "The Impact of Transient and Stable Patterns of Functional Connectivity in Emotion Recognition." *2024 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*. IEEE, 2024.
48. Wirsich, Jonathan, et al. "Complementary contributions of concurrent EEG and fMRI connectivity for predicting structural connectivity." *Neuroimage* 161 (2017): 251-260.
49. Srinivasan, Ramesh, et al. "EEG and MEG coherence: measures of functional connectivity at distinct spatial scales of neocortical dynamics." *Journal of neuroscience methods* 166.1 (2007): 41-52.

50. Friston, Karl, Rosalyn Moran, and Anil K. Seth. "Analysing connectivity with Granger causality and dynamic causal modelling." *Current opinion in neurobiology* 23.2 (2013): 172-178.
51. Sakkalis, Vangelis. "Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG." *Computers in biology and medicine* 41.12 (2011): 1110-1117.
52. Babu, Nirmal Varghese, and E. Grace Mary Kanaga. "Advanced functional connectivity analysis with integrated EEG signal enhancement and GTW-EEG-GAN for emotion detection." *International Journal of Computers and Applications* 46.12 (2024): 1152-1165.
53. Pellegrini, Franziska, et al. "Identifying good practices for detecting inter-regional linear functional connectivity from EEG." *NeuroImage* 277 (2023): 120218.
54. Xu, Jiayang, et al. "EEG decoding for musical emotion with functional connectivity features." *Biomedical Signal Processing and Control* 89 (2024): 105744.
55. Li, Guangqiang, Ning Chen, and Jing Jin. "Semi-supervised EEG emotion recognition model based on enhanced graph fusion and GCN." *Journal of Neural Engineering* 19.2 (2022): 026039.
56. Li, Peiyang, et al. "EEG based emotion recognition by combining functional connectivity network and local activations." *IEEE Transactions on Biomedical Engineering* 66.10 (2019): 2869-2881.
57. Chen, Chuangquan, et al. "Fusing frequency-domain features and brain connectivity features for cross-subject emotion recognition." *IEEE Transactions on Instrumentation and Measurement* 71 (2022): 1-15.
58. Wu, Xun, et al. "Investigating EEG-based functional connectivity patterns for multimodal emotion recognition." *Journal of neural engineering* 19.1 (2022): 016012.
59. Qiu, Xiangkai, et al. "A multi-head residual connection GCN for EEG emotion recognition." *Computers in Biology and Medicine* 163 (2023): 107126.
60. Jin, Ming, et al. "PGCN: Pyramidal graph convolutional network for EEG emotion recognition." *IEEE Transactions on Multimedia* (2024).
61. Li, Ker-Chau. "On principal Hessian directions for data visualization and dimension reduction: Another application of Stein's lemma." *Journal of the American Statistical Association* 87.420 (1992): 1025-1039.
62. Lue, Heng-Hui. "Principal Hessian directions for regression with measurement error." *Biometrika* 91.2 (2004): 409-423.
63. Nanga, Salifu, et al. "Review of dimension reduction methods." *Journal of Data Analysis and Information Processing* 9.3 (2021): 189-231.
64. Li, Ker-Chau, Heng-Hui Lue, and Chun-Houh Chen. "Interactive tree-structured regression via principal Hessian directions." *Journal of the American Statistical Association* 95.450 (2000): 547-560.

65. Ni, Chien-Chun, et al. "Community detection on networks with Ricci flow." *Scientific reports* 9.1 (2019): 9984.
66. Lai, Xin, Shuliang Bai, and Yong Lin. "Normalized discrete Ricci flow used in community detection." *Physica A: Statistical Mechanics and its Applications* 597 (2022): 127251.
67. Sia, Jayson, Edmond Jonckheere, and Paul Bogdan. "Ollivier-ricci curvature-based method to community detection in complex networks." *Scientific reports* 9.1 (2019): 9800.
68. Tian, Yu, Zachary Lubberts, and Melanie Weber. "Curvature-based clustering on graphs." *arXiv preprint arXiv:2307.10155* (2023).
69. Leite, Daniela, et al. "Community detection in networks by dynamical optimal transport formulation." *Scientific Reports* 12.1 (2022): 16811.
70. Brooks, Skylar J., et al. "Community detection in the human connectome: Method types, differences and their impact on inference." *Human Brain Mapping* 45.5 (2024): e26669.
71. Gao, Jie. "Graph Ricci Flow and Applications in Network Analysis and Learning." *2024 Joint Mathematics Meetings (JMM 2024)*. AMS, 2024.
72. Vivancos, David, and Felix Cuesta. "MindBigData 2022 A Large Dataset of Brain Signals." *arXiv preprint arXiv:2212.14746* (2022).
73. Zheng, Wei-Long, and Bao-Liang Lu. "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks." *IEEE Transactions on autonomous mental development* 7.3 (2015): 162-175.
74. Zhu, Ciyou, et al. "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization." *ACM Transactions on mathematical software (TOMS)* 23.4 (1997): 550-560.
75. Ollivier, Yann. "Ricci curvature of metric spaces." *Comptes Rendus Mathematique* 345.11 (2007): 643-646.
76. Chen, Mingming, Konstantin Kuzmin, and Boleslaw K. Szymanski. "Community detection via maximization of modularity and its variants." *IEEE Transactions on Computational Social Systems* 1.1 (2014): 46-65.
77. Yang, Liang, et al. "Modularity based community detection with deep learning." *IJCAI*. 2016.
78. Newman, Mark EJ. "Equivalence between modularity optimization and maximum likelihood methods for community detection." *Physical Review E* 94.5 (2016): 052315.
79. Zhang, Xiang-Sun, et al. "Modularity optimization in community detection of complex networks." *Europhysics Letters* 87.3 (2009): 38002.
80. Cuturi, Marco. "Sinkhorn distances: Lightspeed computation of optimal transport." *Advances in neural information processing systems* 26 (2013).

81. Blitzstein, Joseph K., and Neil Shephard. *Introduction to Statistics: Inference, Description, Prediction, and Causality*. 2025. Textbook for Harvard University's STAT 111 course, Spring 2025.
82. McShane, Blakeley B., et al. "Abandon statistical significance." *The American Statistician* 73.sup1 (2019): 235-245.
83. Betensky, Rebecca A. "The p-value requires context, not a threshold." *The American Statistician* 73.sup1 (2019): 115-117.
84. Von Luxburg, Ulrike. "A tutorial on spectral clustering." *Statistics and Computing* 17 (2007): 395–416.
85. Ng, Andrew, Michael Jordan, and Yair Weiss. "On spectral clustering: Analysis and an algorithm." *Advances in neural information processing systems* 14 (2001).
86. Tu, Yu-Kang, Valerie Clerehugh, and Mark S. Gilthorpe. "Collinearity in linear regression is a serious problem in oral health research." *European journal of oral sciences* 112.5 (2004): 389-397.