



Phase Angle Spatial Embedding (PhASE)

A Kernel Method for Studying the Topology of the Human Functional Connectome

Zachery Morrissey¹, Liang Zhan², Hyekyoung Lee⁵, Johnson Keiriz¹,
Angus Forbes³, Olusola Ajilore¹, Alex Leow¹(✉), and Moo Chung⁴

¹ University of Illinois at Chicago, Chicago, IL 60612, USA
aleow@psych.uic.edu

² University of Pittsburgh, Pittsburgh, PA 15260, USA

³ University of California Santa Cruz, Santa Cruz, CA 95064, USA

⁴ University of Wisconsin-Madison, Madison, WI 53706, USA

⁵ Seoul National University, Seoul, Republic of Korea

Abstract. Modern resting-state functional magnetic resonance imaging (rs-fMRI) provides a wealth of information about the inherent functional connectivity of the human brain. However, understanding the role of negative correlations and the nonlinear topology of rs-fMRI remains a challenge. To address these challenges, we propose a novel graph embedding technique, **phase angle spatial embedding (PhASE)**, to study the “intrinsic geometry” of the functional connectome. PhASE both incorporates negative correlations as well as reformulates the connectome modularity problem as a kernel two-sample test, using a kernel method that induces a *maximum mean discrepancy* (MMD) in a reproducing kernel Hilbert space (RKHS). By solving a graph partition that maximizes this MMD, PhASE identifies the most functionally distinct brain modules. As a test case, we analyzed a public rs-fMRI dataset to compare male and female connectomes using PhASE and minimum spanning tree inferential statistics. These results show statistically significant differences between male and female resting-state brain networks, demonstrating PhASE to be a robust tool for connectome analysis.

1 Introduction

Human connectomics has progressed rapidly in the last decade in part due to advances in the application of graph theory and complex network analysis to brain imaging data [2, 11]. For example, graph theoretical analysis of resting-state (i.e. task-negative) fMRI has been increasingly used to study the inherent functional connectivity of the brain in both healthy subjects and those with psychiatric or neurodegenerative conditions [8]. Thus, better understanding resting-state connectomes can elucidate how brain networks are organized differently.

In functional connectome analysis, graph theory is used to organize the pairwise correlations for all brain regions of interest (ROIs) into an adjacency matrix with signed edges [2]. However, there are some limitations with this approach.

One issue complicating traditional rs-fMRI graph analysis is how to properly model negative correlations between ROIs [9,14]. While applying thresholds or similar approaches to focus on positive correlations alone can simplify downstream analyses, this potentially removes biologically-relevant connectivity information [11]. Another issue is that while graph metrics can reveal important information regarding global/nodal characteristics of connectivity data, topological features are largely inaccessible to this analysis [13]. Thus, these limitations highlight the need for a way to represent functional connectomes in their native space.

The main contribution of this paper is the development of a kernel method approach for studying the topology of the functional connectome. To this end, we proposed **phase angle spatial embedding (PhASE)**, by modeling the functional coupling between two ROIs as a phase angle (from 0 to $\pi/2$, corresponding to full co-activation vs. anti-activation, respectively). With PhASE, topological features of resting-state connectivity can be intuitively understood by encoding each ROI using all phase angles that relate this ROI to the rest of the brain. This offers a mathematically robust method for working with positive *and* negative edges, as the full range of correlative information is preserved. Further, we equipped phase with a kernel, allowing us to conceptualize connectome modularity as a kernel two-sample problem, via the maximization of the corresponding maximum mean discrepancy (MMD) metric [4,5] in a reproducing kernel Hilbert space (RKHS). Last, we outlined a statistical inference procedure to compare the connectome topology between two groups (e.g., male vs. female) via the corresponding minimum spanning trees (MSTs) induced by the manifold geodesic distances in PhASE [7]. Thus, the functional connectome can be transformed into a new space representative of its “intrinsic geometry”, and can be used for state-of-the-art topological data analysis techniques [3].

2 Methods

Rationale of PhASE. The core procedure of PhASE is to encode the probability of two brain regions (i and j) exhibiting a coupling relationship of co- (p_{ij}^+) vs. anti- (p_{ij}^-) activation using a phase angle between 0 and $\pi/2$ (Fig. 1). One (simple) way of defining probability can be done using individual correlation matrices. In a dataset with Z subjects, if we treat the correlation between brain regions i and j as an edge (e_{ij}), then we can define the probability of there being a negative edge between i and j (p_{ij}^-) as the average occurrence of observing a negative correlation between i and j across all Z subjects, i.e. $p_{ij}^- = \sum_{z=1}^Z \mathbb{I}[e_{ij}^z < 0] / Z$, where e_{ij}^z is the edge for the z -th subject. Computing this probability for all ROIs then forms a *negative probability matrix* (NPM). By taking advantage of the simple relationship that $p_{ij}^- + p_{ij}^+ = 1$, the phase angle between i and j can be defined as

$$\theta_{ij} = \arctan \left(\frac{p_{ij}^-}{p_{ij}^+} \right)^{1/2}. \quad (1)$$

Furthermore, by fixing i and considering all possible j , the PhASE procedure encodes the i -th ROI in an N -dimensional Euclidean space (where N is the number of ROIs), such that the phase angle vector $\theta_i = [\theta_{i1}, \theta_{i2}, \dots, \theta_{iN}]^\top \in [0, \pi/2]^N$. (A phase angle of 0 represents fully co-activating ROIs and $\pi/2$ fully anti-activating ROIs). Given a PhASE matrix, state-of-the-art nonlinear dimensionality reduction techniques or topological analysis via the geodesic distance (in the N -dimensional PhASE embedding space) induced minimum spanning tree (MST) can then be used to recover the intrinsic geometry of the functional connectome.

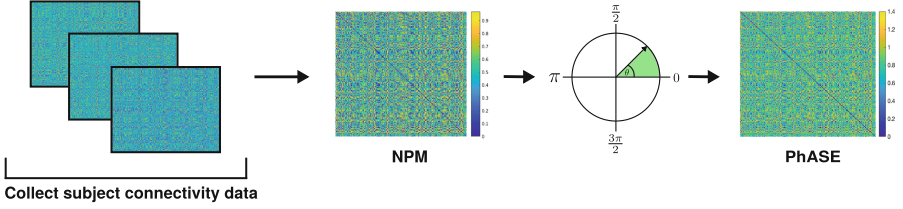


Fig. 1. Computing PhASE from rs-fMRI data. Connectivity data is acquired from multiple subjects, and the frequency of edge negativity is computed for each ROI to form a negative probability matrix (NPM). Then, the phase angle (θ) is computed for each ROI, which encodes each ROI’s “phase” with every other ROI in the network.

PhASE-Induced Modularity: PhASE as a Kernel Two-Sample Test. We further show that PhASE reformulates the connectome modularity problem as a kernel two-sample problem. In its simplest form, the connectome modularity problem seeks to decompose a graph into M modules, or “communities”, and is often computed by maximizing the Q modularity metric using Louvain’s method [10, 11]. Here, we illustrate how PhASE tackles this problem from a machine-learning perspective using the case $M = 2$ as an example of finding two communities.

Following the formulation detailed in Gretton et al. [4], consider the random variables x and y defined on a metric space $\mathcal{X}$ equipped with the metric d , with the corresponding Borel probabilities p and q (i.e. $x \sim p$ and $y \sim q$). Given observations $X := \{x_1, \dots, x_m\}$ and $Y := \{y_1, \dots, y_n\}$ drawn from the probability distributions p and q , $p = q$ if and only if $\mathbf{E}_x(f(x)) = \mathbf{E}_y(f(y)) \forall f \in C(\mathcal{X})$, where $C(\mathcal{X})$ is the space of bounded continuous functions on $\mathcal{X}$.

Given this setup, and a class of functions $\mathcal{F}$ such that $f : \mathcal{X} \rightarrow \mathbb{R}$, the *maximum mean discrepancy* (MMD) between p and q with respect to $\mathcal{F}$ is defined as

$$\text{MMD}[\mathcal{F}, p, q] := \sup_{f \in \mathcal{F}} (\mathbf{E}_x[f(x)] - \mathbf{E}_y[f(y)]). \quad (2)$$

In PhASE modularity, using the same definitions for x, y, p, q, X , and Y , each ROI is assigned to two possible classes of distribution (p and q) such that the

MMD is maximized. Given observations X and Y , an empirical estimate of the MMD can therefore be defined as

$$\text{MMD}[\mathcal{F}, X, Y] := \sup_{f \in \mathcal{F}} \left(\frac{1}{m} \sum_{i=1}^m f(x_i) - \frac{1}{n} \sum_{i=1}^n f(y_i) \right). \quad (3)$$

However, with RKHS it is possible to evaluate the squared MMD using kernel values alone such that

$$\begin{aligned} \text{MMD}^2[\mathcal{F}, X, Y] = & \frac{1}{m(m-1)} \sum_{i=1}^m \sum_{j \neq i}^m k(x_i, x_j) + \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n k(y_i, y_j) \\ & - \frac{2}{mn} \sum_{i=1}^m \sum_{j=1}^n k(x_i, y_j). \end{aligned} \quad (4)$$

A standard choice of kernel defined between two ROIs, each encoded using its PhASE phase angle vector θ_i , is the Radial Basis Function (RBF)

$$k(i, j) = \exp \left\{ -\sigma \sum_{\ell=1}^N |\theta_{i\ell} - \theta_{j\ell}|^2 \right\}. \quad (5)$$

Another kernel choice, which may be more “natural” for phase angle vectors is

$$k(i, j) = \frac{1}{N} \sum_{\ell=1}^N \cos(\theta_{i\ell} - \theta_{j\ell}), \quad (6)$$

where N is the number of ROIs. Thus, this kernel represents each ROI’s phase angle vector similarity in relation to all other ROIs in the network. While these two choices appear quite different, a quick check on their Taylor expansions make it clear that they have very similar leading terms.

This kernel-based approach for representing functional brain connectivity and modularity in a native way can be used in conjunction with topological data analysis on minimum spanning trees, as described in the following section. As a result, these techniques can be complemented together to offer a robust framework for studying differences across group networks.

Inference on Minimum Spanning Trees. Most statistical inference methods on MST rely on existing univariate statistical test procedures on scalar graph theory features such as the average path length [12]. Since the probability distribution of such features are often not known, resampling techniques such as the permutation test are frequently used. The permutation test is known as the exact test procedure in statistics. However, it is not exact in practice and only an approximate method due to the computational bottleneck of generating every possible permutation, which can be astronomically large. Thus, the statistical significance is computed using the small subset of all possible permutations, which

gives approximate p -values. Here, we present a new method for the permutation test, where every possible permutation is enumerated combinatorially.

Let M_1 and M_2 be the minimum spanning trees (MST) corresponding to $N \times N$ PhASE-derived adjacency matrices C_1 and C_2 . We are interested in testing hypotheses

$$H_0 : M_1 = M_2 \text{ vs. } H_1 : M_1 \neq M_2. \quad (7)$$

The statistic for testing H_0 is constructed using Kruskal's algorithm [6]. Kruskal's algorithm is a greedy algorithm with runtime $O(N \log N)$. The algorithm starts with an edge with the smallest weight. Then add an edge with the next smallest weight. This sequential process continues while avoiding a loop and generates a spanning tree with the smallest total edge weights. Thus, the edge weights in MST correspond to the order in which the edges are added in the construction of MST. For a graph with N nodes, MST has $N - 1$ edges.

Let $w_1^1 < w_2^1 < \dots < w_{N-1}^1$ and $w_1^2 < w_2^2 < \dots < w_{N-1}^2$ be the ordered edge weights of two MSTs obtained from Kruskal's algorithm. Let f be a monotone function that maps the edge weights to integers such that $\phi(w_j^1) = \psi(w_j^2) = j$. ϕ, ψ can be the number of edges in the subtree generated in the j -th iteration in Kruskal's algorithm. Under the null hypothesis (7), the monotone functions $\phi(w_j^1)$ and $\psi(w_j^2)$ are interchangeable, thus satisfying the condition for the permutation test. The pseudo-metric $D(M_1, M_2) = \sup_t |\phi(t) - \psi(t)|$ is then used as a test statistic. Under H_0 , $D(M_1, M_2) = 0$. The larger the value of D is, it is more likely to reject H_0 . From Theorem 2 in [3], it can be shown that the probability distribution of D is given by

$$P(D \geq d) = 1 - \frac{A_{N-1, N-1}}{\binom{2N-2}{N-1}}, \quad (8)$$

where $A_{u,v}$ satisfies $A_{u,v} = A_{u-1,v} + A_{u,v-1}$ with the boundary condition $A_{0,N} = A_{N,0} = 1$ within band $|u - v| < d$. (8) is used to compute p -values in the study.

Resting-State fMRI Data Acquisition. For testing PhASE, a publicly available rs-fMRI dataset from the F1000 Project [1] was used. For details regarding data acquisition and processing, see http://fcon_1000.projects.nitrc.org/. The resulting data is a 177×177 ROI correlation matrix for each subject ($Z = 986$, males = 426, females = 560). Then, the negative probability matrix was formed using the simple definition of probability described in (2).

3 Results and Discussion

As a test case for our phase angle spatial embedding (PhASE) method, PhASE was applied to resting-state functional connectivity data (177×177 ROI) [1] to investigate differences between male and female resting-state brain networks. First, we investigated network modularity using MST and anatomical representations. To compare to standard modularity techniques, Louvain's Q modularity

method [11] ($\gamma = 1$) was used, generating three communities (Fig. 2A). After applying the PhASE transformation and computing the MMD kernel matrix via (6), two communities were generated (Fig. 2B). Notably, the PhASE modularity boundaries share similar boundaries with those achieved by Q, both in MST and anatomical space. This demonstrates that PhASE can create reproducible results as those achieved by standard methods.

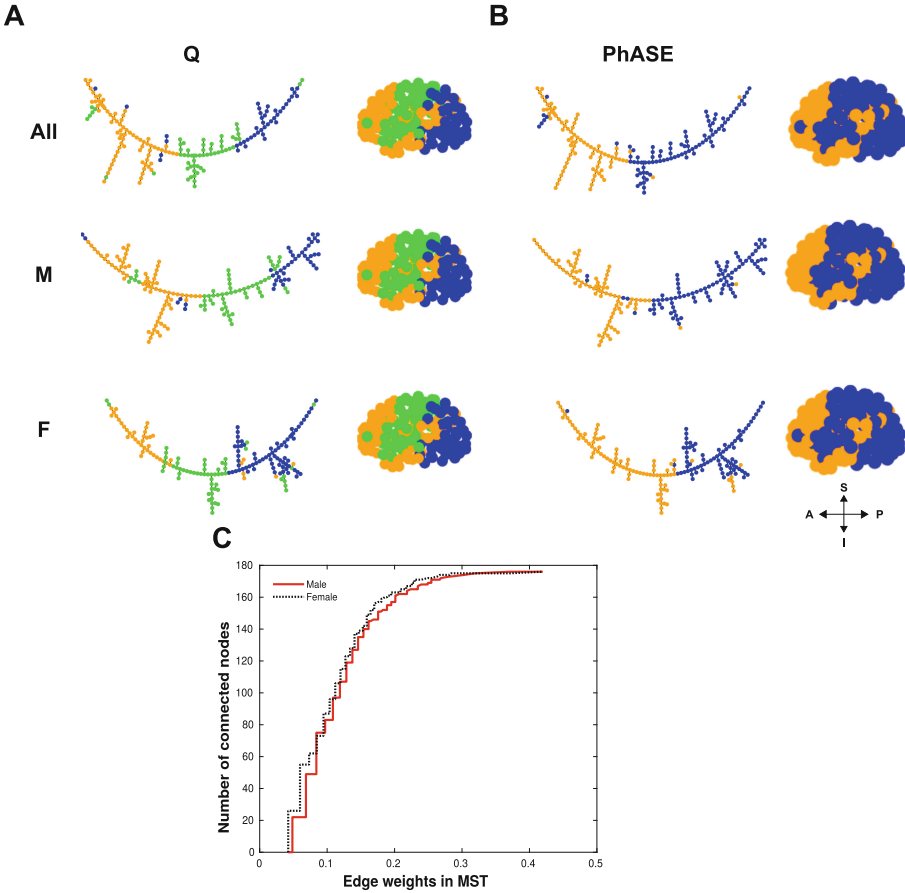


Fig. 2. (A–B) Minimum spanning trees (MSTs) and anatomical representations (MNI coordinates) for all, male (M), and female (F) groups, with associated modularity color using Q ($\gamma = 1$) and PhASE. (C) Exact topological inference comparing male and female brain networks. The distance between networks was $D_{\max}(\text{male}, \text{female}) = 33$, occurring at edge weights 0.0598 and 0.0686, and was statistically significant ($p < 0.001$).

While MSTs can produce informative visual differences between group networks, we sought to quantitatively analyze whether there are statistically sig-

nificant differences between male and female resting-state brain networks using MSTs. As described above, the null hypothesis (7) that the male and female graphs are equivalent was tested using a nonparametric inference method that does not need to rely on the permutation test. Using the male and female PhASE matrices, MST edge weights were simply the phase angle between ROIS (1). Exact topological inference (8) was then used to compute the distance between male and female MSTs, and was statistically significant ($D_{\max}(\text{male}, \text{female}) = 33$, $p < 0.001$). Thus, the null hypothesis can be rejected, suggesting that there are intrinsic differences in resting-state connectivity between sexes. These differences can be seen in Fig. 2C, wherein there is a consistent separation between male and female number of connected nodes as a function of MST edge weights. Although the curves are similar in shape, the results indicate that the female MSTs appear to exhibit a higher number of connected nodes at similar edge weights compared to males, suggesting that female subjects' resting-state backbone networks are more "connected" than male subjects—a salient network finding that merits further investigation, and importantly here demonstrates the capability of PhASE for topological analysis.

4 Conclusions

Representing and computing the modularity of functional connectomes with negative edges is an ongoing challenge in rs-fMRI analysis. To that end, we presented a kernel-based approach for analyzing functional connectivity using a novel graph embedding method, phase angle spatial embedding (PhASE), that respects negative correlations and reformulates the modularity problem as a two-sample problem. By doing so, it is possible to create a more natural representation of functional connectivity that encodes connective phase information in its native space and can be used with advanced topological analyses. As an initial case, sex differences were examined using PhASE with MST inferential statistical methods, revealing statistically significant differences between male and female rs-fMRI networks. Taken together, these results demonstrate the versatility of PhASE as a connectome analysis framework that can yield greater insight into functional connectivity while preserving the intrinsic geometry of the human connectome.

Acknowledgments. This manuscript was supported by NIH T32 AG057468 to ZM, NIH AG056782 to LZ and AL, and NIH EB022856 to MC.

References

1. Biswal, B.B., et al.: Toward discovery science of human brain function. *Proc. Natl. Acad. Sci.* **107**(10), 4734–4739 (2010)
2. Bullmore, E., Sporns, O.: Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* **10**(3), 186–198 (2009)

3. Chung, M.K., Villalta-Gil, V., Lee, H., Rathouz, P.J., Lahey, B.B., Zald, D.H.: Exact topological inference for paired brain networks *via* persistent homology. In: Niethammer, M., et al. (eds.) IPMI 2017. LNCS, vol. 10265, pp. 299–310. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59050-9_24
4. Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B., Smola, A.: A kernel two-sample test. *J. Mach. Learn. Res.* **13**(Mar), 723–773 (2012)
5. Gretton, A., et al.: Optimal kernel choice for large-scale two-sample tests. In: *Advances in Neural Information Processing Systems*, pp. 1205–1213 (2012)
6. Lee, H., Kang, H., Chung, M.K., Kim, B.-N., Lee, D.S.: Persistent brain network homology from the perspective of dendrogram. *IEEE Trans. Med. Imaging* **31**(12), 2267–2277 (2012)
7. Lee, H., Chung, M.K., Kang, H., Kim, B.-N., Lee, D.S.: Computing the shape of brain networks using graph filtration and Gromov-Hausdorff metric. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011. LNCS, vol. 6892, pp. 302–309. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23629-7_37
8. Lv, H.: Resting-state functional MRI: everything that nonexperts have always wanted to know. *Am. J. Neuroradiol.* **39**(8), 1390–1399 (2018). <https://doi.org/10.3174/ajnr.A5527>. ISSN 0195-6108
9. Murphy, K., Fox, M.D.: Towards a consensus regarding global signal regression for resting state functional connectivity MRI. *NeuroImage* **154**, 169–173 (2017)
10. Newman, M.E.: Modularity and community structure in networks. *Proc. Natl. Acad. Sci.* **103**(23), 8577–8582 (2006)
11. Rubinov, M., Sporns, O.: Complex network measures of brain connectivity: uses and interpretations. *NeuroImage* **52**(3), 1059–1069 (2010)
12. Stam, C., Tewarie, P., Van Dellen, E., van Straaten, E., Hillebrand, A., Van Mieghem, P.: The trees and the forest: characterization of complex brain networks with minimum spanning trees. *Int. J. Psychophysiol.* **92**(3), 129–138 (2014)
13. Ye, A.Q., et al.: The intrinsic geometry of the human brain connectome. *Brain Inf.* **2**(4), 197–210 (2015)
14. Zhan, L.: The significance of negative correlations in brain connectivity. *J. Comput. Neurol.* **525**(15), 3251–3265 (2017)