ANGULAR CENTRAL GAUSSIAN AND WATSON MIXTURE MODELS FOR ASSESSING DYNAMIC FUNCTIONAL BRAIN CONNECTIVITY DURING A MOTOR TASK

Anders S. Olsen⋆   Emil Ortvald⋆   Kristoffer H. Madsen⋆†   Mikkel N. Schmidt⋆   Morten Mørup⋆

⋆ Department of Applied Mathematics and Computer Science,
Technical University of Denmark, Kgs. Lyngby, Denmark
† Danish Research Centre for Magnetic Resonance,
Copenhagen University Hospital Amager and Hvidovre, Hvidovre, Denmark

ABSTRACT

The development of appropriate models for dynamic functional connectivity is imperative to gain a better understanding of the brain both during rest and while performing a task. Leading eigenvector dynamics analysis is among the favored methods for assessing frame-wise connectivity, but eigenvectors are distributed on the sign-symmetric unit hypersphere, which is typically disregarded during modeling. Here we develop both mixture model and Hidden Markov model formulations for two sign-symmetric spherical statistical distributions and display their performance on synthetic data and task-fMRI data involving a finger-tapping task.

***Index Terms*—** Dynamic functional connectivity, Leading eigenvector dynamics analysis, Watson, Angular Central Gaussian, Hidden Markov models

1. INTRODUCTION

The brain is a dynamic complex system composed of anatomically and functionally distinct brain regions that integrate their activity to form functional brain networks [1]. There is a great interest in uncovering dynamic functional brain networks using functional neuroimaging modalities such as functional magnetic resonance imaging (fMRI) data, both in terms of the topographical layout of *states* of activity and state-state dynamics [2].

Existing methods for uncovering time-varying brain networks in both resting-state and task fMRI predominantly include assessing interregional correlation in sliding temporal windows and subsequently performing clustering using, e.g., k-means [3]. However, windowed methods require decisions on window length, function, and stride, which greatly influence results [4]. Leading eigenvector dynamics analysis (LEiDA) [5] is a dynamic functional connectivity (dFC) method that computes the leading eigenvector of instantaneous phase coherence maps. By focusing on phase coherence instead of correlation, only the instantaneous synchrony between regional oscillations influences the connectivity measure, while the strength of the correlation is disregarded. As such, the effect of potentially spurious connectivity spikes arising from, e.g., motion is suppressed, while a window-free connectivity measure is established.

Eigenvectors represent axes and are thus, by convention, normalized to unit norm and have arbitrary sign, meaning that they are distributed on the antipodally symmetric unit hypersphere. Olsen et al., (2022) suggested the use of diametrical clustering [6] in conjunction with LEiDA [7], which assumes data to be distributed according to the Watson distribution [8], which models points on the unit hypersphere with sign-symmetric mean cluster *directions*.

The natural probabilistic extension of the K-means type diametrical clustering is a mixture of Watson distributions [9], which in addition models isotropic cluster variance as a scalar, as well as cluster weights indicating their relative importance. The generalization of the Watson distribution is the Bingham distribution [10], which models data using a covariance matrix, thereby allowing for anisotropic covariance structures. However, the Bingham distribution has a normalization constant that is intractable in high dimensions. Instead, the Angular Central Gaussian (ACG) distribution [11], which is an analogue of the zero-mean Gaussian distribution projected onto the hypersphere, has a tractable normalization constant.

Here we model synthetic and task-fMRI data with both Watson and ACG mixture models as well as their equivalent Hidden Markov model (HMM) formulations, leading to a comparison of four models. The models are implemented in PyTorch, which enables gradient-based likelihood optimization using automatic differentiation. Using synthetic data we demonstrate the superiority of the ACG in learning anisotropic covariance structures. Using fMRI data involving a finger-tapping experiment [12] we assess if the models uncover sensible connectivity representations in a constrained setting. We demonstrate that while the Watson model learns a rank one connectivity matrix representation, the ACG may be used to learn any rank $r \in \{1, \ldots, p\}$ representation, where the optimal rank r and number of components K can be found using cross-validation.

2. METHODS

2.1. Leading Eigenvector Dynamics Analysis

LEiDA computes the dominant instantaneous interregional phase coherence pattern (see Fig. 1A). The regional phase series $\phi_l(t)$ is estimated for each region l from the regional input time series $s_l(t)$ via

$$\phi_l(t) = \tan^{-1}\left(\frac{s_l(t) * \frac{1}{\pi t}}{s_l(t)}\right),$$

where $*$ represents the convolution operator. A phase coherence map $\mathbf{A}_t$ is established for regions l and l' for each time point t with elements $A_{l,l',t} = \cos(\phi_l(t) - \phi_{l'}(t))$. Due to the angle difference identity, a matrix with cosine elements has rank 2, and thus, extracting the leading eigenvector from $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ retains at least half of the variance in each phase coherence map while retaining only one latent factor.
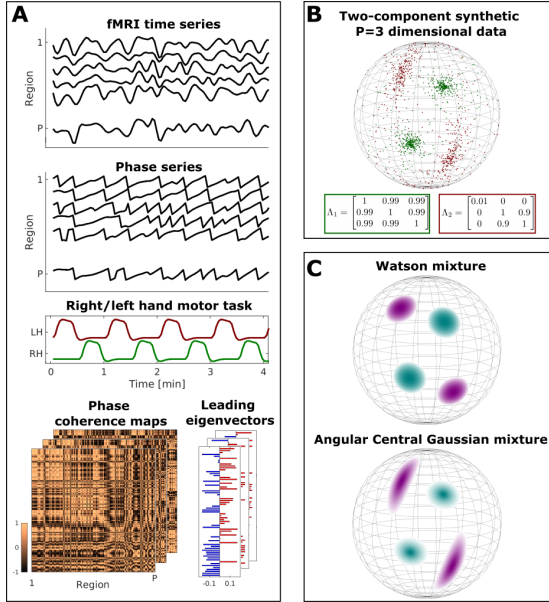

Fig. 1. Methodological pipeline. (A): Leading eigenvector dynamics analysis (LEiDA) constructs leading eigenvectors of instantaneous phase coherence maps. (B): Synthetic data on the sign-symmetric unit hypersphere generated by a two-component angular central Gaussian (ACG) distribution with the displayed covariance matrices. (C): Two-component Watson and ACG mixture model fits on the synthetic data in (B).

2.2. Watson distribution

Since eigenvectors by convention are unit norm and sign-symmetric [7], the LEiDA data $\mathbf{x}$ lie on the antipodally symmetric $(p-1)$-dimensional unit hypersphere, $\pm\mathbf{x} \in \mathbb{S}^{p-1}$. Several statistical distributions model points on this manifold, the simplest being the Watson distribution with density [8]:

$$f_{\mathrm{W}}(\pm\mathbf{x}; \boldsymbol{\mu}, \kappa) = c_p(\kappa)e^{\kappa(\boldsymbol{\mu}^\top\mathbf{x})^2}, \quad \mathbf{x} \in \mathbb{S}^{p-1}, \qquad (1)$$

where $\boldsymbol{\mu} \in \mathbb{S}^{p-1}$ is the mean direction and κ is a scalar precision parameter. $c_p(\kappa)$ is the normalization constant

$$c_p(\kappa) = \frac{\Gamma\left(\frac{p}{2}\right)}{2\pi^{p/2}M\left(\frac{1}{2}, \frac{p}{2}, \kappa\right)},$$

where M is Kummer's confluent hypergeometric function

$$M(a, b, \kappa) = \sum_{n=0}^{\infty} \frac{\Gamma(a+n)\Gamma(b)}{\Gamma(a)\Gamma(b+n)}\frac{\kappa^n}{n!}.$$

2.3. Angular Central Gaussian distribution

Whereas the Watson distribution has equal variance in all dimensions, a more expressive alternative is the ACG distribution with density:

$$f_{\mathrm{ACG}}(\pm\mathbf{x}; \boldsymbol{\Lambda}) = \frac{\Gamma\left(\frac{p}{2}\right)}{2\pi^{p/2}|\boldsymbol{\Lambda}|^{\frac{1}{2}}}\left(\mathbf{x}^\top\boldsymbol{\Lambda}^{-1}\mathbf{x}\right)^{-\frac{p}{2}}, \mathbf{x} \in \mathbb{S}^{p-1}.$$

$$(2)$$

Here $\boldsymbol{\Lambda} \in \mathbb{R}^{p \times p}$ is a symmetric positive-definite matrix identifiable up to a positive scale factor.

2.4. Mixture models and Hidden Markov models

The Watson and ACG distributions can enter as components in a mixture model with weights $\boldsymbol{\pi} = \{\pi_j\}_{j=1}^K$, leading to the likelihood

$$L_{\mathrm{Mix}}(\boldsymbol{\theta}, \boldsymbol{\pi}|\mathbf{X}) = \prod_{i=1}^{N}\sum_{j=1}^{K}\pi_j\, f\left(\mathbf{x}_i; \boldsymbol{\theta}_j\right), \qquad (3)$$

where $\boldsymbol{\theta}_j$ are the parameters of the jth component, i.e., $\boldsymbol{\theta}_j = \{\boldsymbol{\mu}_j, \kappa_j\}$ (Watson) or $\boldsymbol{\theta}_j = \Lambda_j$ (ACG).

While a mixture model does not account for temporal structure, we can model state dynamics using an HMM. This extends the mixture model by a transition probability matrix $\boldsymbol{T}$ with elements $T_{j',j} = P\left(z_t^{(s)} = j|z_{t-1}^{(s)} = j'\right)$, where $z_t^{(s)}$ is the state at time $t \in 1, \ldots, \tau_s$ for the sth sequence. In practice, the HMM is likely more reasonable than the mixture model as it favors more smooth state transitions. Maximum likelihood inference requires evaluation over all possible state sequences $\mathbf{z}$: This can be computed efficiently using the forward algorithm

$$\alpha_{t,j}^{(s)} = \sum_{j'=1}^{K}\alpha_{t-1,j'}^{(s)}T_{j',j}f(\mathbf{x}_t^{(s)}; \boldsymbol{\theta}_j),$$

where $\alpha_{t,j}^{(s)}$ represents the probability of being in state j at time t, and $\alpha_{1,j}^{(s)} = \pi_j f(\mathbf{x}_1^{(s)}; \boldsymbol{\theta}_j)$. Finally, the likelihood to be optimized is

$$L_{\text{HMM}}(\boldsymbol{\theta}, \boldsymbol{\pi}, \boldsymbol{T}|\mathbf{X}) = \prod_{s=1}^{S}\sum_{j=1}^{K} \alpha_{\tau_s,j}^{(s)}, \quad (4)$$

where S is the number of observation sequences.

2.5. Computational implementation

We maximized the likelihood of the mixture model (Eq. 3) and HMM (Eq. 4) using stochastic gradient descent with the ADAM [13] optimizer with a learning rate of 0.1, which we found to be a reasonable trade-off between fast convergence and stochasticity. Calculations were performed in the log domain where possible to avoid numerical instability. Parameter constraints were handled by reparametrization, expressing all parameters in terms of unconstrained variables: Watson components were constrained to unit-norm mean directions $\boldsymbol{\mu} = \frac{\tilde{\boldsymbol{\mu}}}{\|\tilde{\boldsymbol{\mu}}\|}$ and the positive precision parameters were expressed using the softplus function $\kappa = \log(1+e^{\tilde{\kappa}})$. We used a softmax function for the mixture weights $\pi_j = \frac{e^{\tilde{\pi}_j}}{\sum_{k=1}^{K} e^{\tilde{\pi}_k}}$ as well as each row of the transition matrix $\mathbf{T}$. For the ACG components, we developed two estimation schemes: For low-dimensional problems, we expressed $\boldsymbol{\Lambda}^{-1} = \mathbf{L}\mathbf{L}^{\top}$ as a product of lower-triangular matrices. For high dimensions we introduce a low rank parametrization $\boldsymbol{\Lambda} = \mathbf{M}\mathbf{M}^{\top} + \mathbf{I}$, where $\mathbf{M} \in \mathbb{R}^{p \times r}$ is of rank r and $\mathbf{I}$ is the identity matrix. $\mathbf{M}$ starts with one column vector and consecutively adds columns initialized by the previous model until rank r is achieved.

2.6. Experimental data

We analyzed experimental 3T BOLD fMRI data with acquisition parameters (TR/TE = 2490/30 ms, $3 \times 3 \times 3$ mm3 isotropic voxels, 240 volumes). The full data acquisition and preprocessing details are described elsewhere [12]. During scanning, the 29 participants performed a block-design motor task visually cued by a blinking light indicating whether the finger tapping should be performed with the right or left hand (see Fig. 1A). We used the 100-region Schaefer parcellation to spatially downsample the volumes [14]. For model performance investigations we used a split-half cross-validation scheme in which the first 120 volumes for all subjects were part of the training set, while the latter 120 volumes constituted the test set.

Brain networks were visualized with BrainNet Viewer (http://www.nitrc.org/projects/bnv/) [15]. All code produced is available at https://github.com/anders-s-olsen/Task_WMM_ACGMM_HMM.

3. RESULTS

3.1. Synthetic data

Three-dimensional data (on the two-sphere) were generated from a two-component ACG with a time-dependent proba-

bility corresponding to the maximum right/left-hand motor task probability (Fig. 1). One of the components is isotropic and the other has an oval covariance structure. The Watson model learns mean cluster directions corresponding roughly to the ground truth, albeit with a scalar variance parameter for both components (Fig. 1C). On the other hand, the ACG model optimized via the Cholesky decomposition of $\boldsymbol{\Lambda}^{-1}$ learns roughly the correct distributions.

Next, we conducted a noise experiment where we varied the difference of the off-diagonal elements of $\boldsymbol{\Lambda}_1$ from 1 and the first diagonal element of $\boldsymbol{\Lambda}_2$, thereby effectively increasing the isotropic variance of both components while the covariance of latter also gradually loses anisotropy. The normalized mutual information (NMI) reveals that the ACG models generally have a higher information overlap with the ground truth cluster indices than the Watson equivalents. The HMM models work well for low noise levels but converge to single-component models for medium-high noise levels leading to zero NMI.
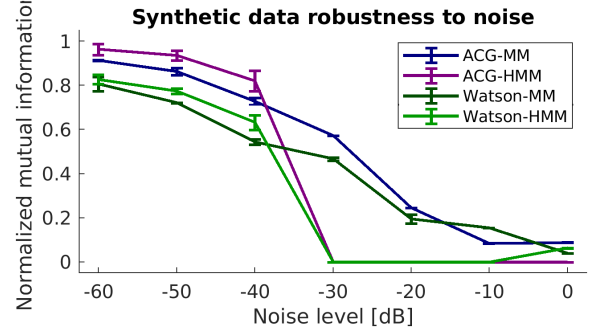

Fig. 2. Normalized mutual information between the two-component fit state probabilities (mixture models (MM)) or state sequence (Hidden Markov models (HMM)) with the true cluster identity. Noise varied as one minus the off-diagonal elements of $\boldsymbol{\Lambda}_1$ and the first diagonal element of $\boldsymbol{\Lambda}_2$ in Figure 1B. Error bars represent the standard deviation over five runs.

3.2. Finger-tapping fMRI data

The Cholesky formulation of the ACG model becomes unstable in higher dimensions and tends to single-component models. Instead we focus on a rank-r approximation $\boldsymbol{\Lambda} = \mathbf{M}\mathbf{M}^{\top} + \mathbf{I}$. First, we investigated whether initializing the rank-1 ACG mixture model with the Watson mixture model output, i.e., $\mathbf{M} = \sqrt{\kappa}\boldsymbol{\mu}$ would lead to superior model performance over a randomly initialized model. As seen in Fig. 3A, the variance decreases slightly for larger K, thus, we proceed with a Watson-initialized ACG model. Fig. 3B shows that the test likelihood as a function of ACG rank does not appear to improve significantly beyond $r \approx 15$.

Finally, we investigated model order for all four models including rank-15 ACG models (see Fig. 4A). While the Watson models decrease in negative test likelihood over all K
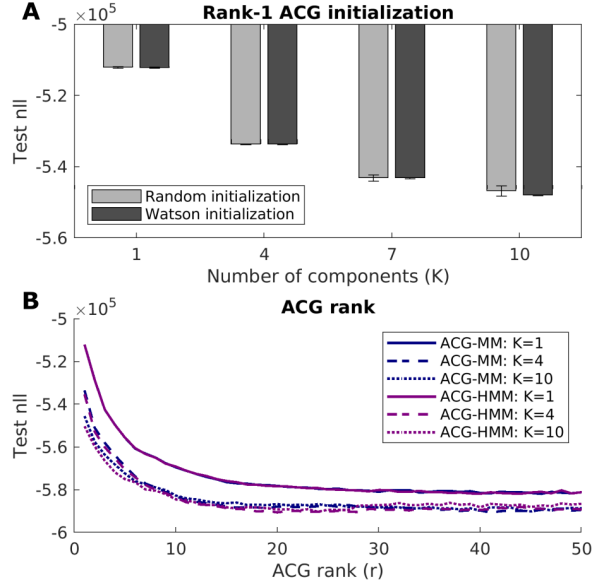
Fig. 3. ACG model specifications. (A): Comparison of the test negative log-likelihood (nll) performance for two initialization methods of the ACG mixture model. Error bars represent the standard deviation over five runs. (B): The evolution of test performance dependent on ACG rank for three different model orders.

there appears to be a slight bend in the likelihood curves at $K = 4$. The ACG models increase slightly in test likelihood from $K = 4$ to $K = 5$. Thus, $K = 4$ seems to be a reasonable tradeoff between model complexity and performance for all models. The components learned on the full dataset for the Watson model for $K = 4$ (Fig. 4B) show a general hypoconnectivity in some subregions of the somatomotor cortex for both the right and left hemisphere for states 2 and 4, while states 1 and 3 appear to be less influenced by these regions though still distinguishable. For the ACG model (Fig. 4C), all four states clearly separate the somatomotor cortex from the rest of the brain, two of the states being hemisphere-specific. Interestingly, the differentiation appears to be stronger for the right hemisphere (state 4) and even covers three of five subsomatomotor areas in the Schaefer-100 atlas.

4. DISCUSSION

Here we implemented Watson and ACG distributions in both mixture model and HMM formulations to model brain dFC in a finger-tapping task-fMRI experiment. We noticed a propensity for the HMM models to converge to single-component models for medium-high noise levels in synthetic data, indicating that these are more susceptible to noise. Using split-half cross-validation with fMRI data we demonstrated a reasonable rank for the ACG models to learn from this data. Future studies could investigate the distribution of the optimal rank depending on data type (e.g., task vs resting-state), num-
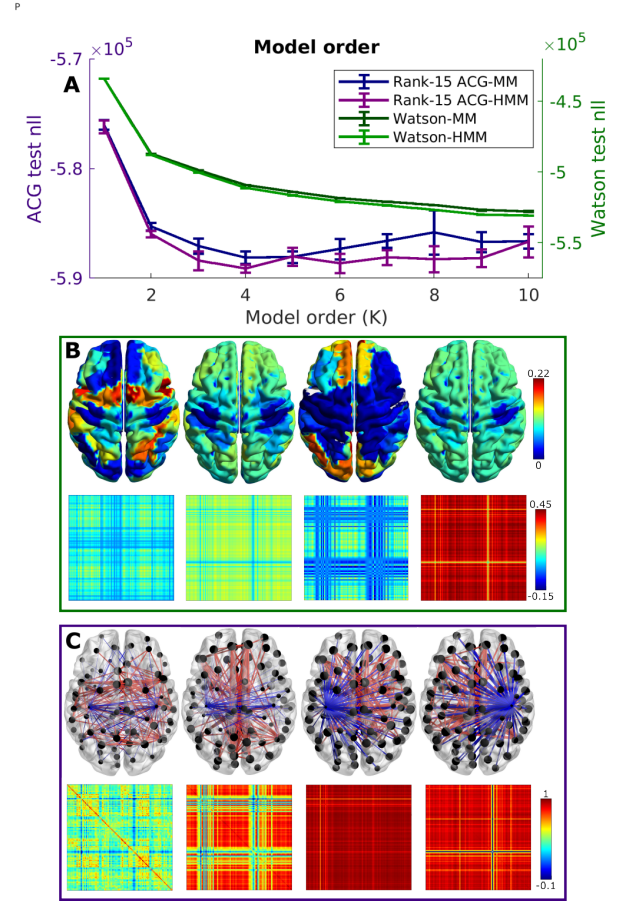

Fig. 4. Model fits to experimental data. (A): Evolution of test performance over model order. (B) Watson mixture model fit for $K = 4$ including connectivity map $\mathbf{\Lambda} = \sqrt{\kappa}\boldsymbol{\mu}\boldsymbol{\mu}^\top$ and a surface rendering of the diagonal of $\mathbf{\Lambda}$. (C) Brain graph rendering showing the top (red) and bottom (blue) 2.5% edges and connectivity map for the ACG mixture model fit for $K = 4$, where node size is the diagonal of $\mathbf{\Lambda}$.

ber of regions, and model order. A converging rank could indicate an optimal LEiDA-determined *complexity* of the human brain. The high rank indicates that the ACG captures more information in the data than the Watson model. In the rank-15 ACG model, we noticed a worsened test likelihood for model orders higher than four, which could be explained by the constrained task-active brain, i.e., we might see higher model orders for resting-state data or more complex tasks. The components fitted on the full data set revealed hypoconnectivity in some somatomotor regions for some Watson components and all ACG components while left/right hemisphere specificity was only present in the latter. Clearly, the complexity of the state maps is higher for ACG than for Watson. Further studies may investigate the temporal alignment of states to the presumed task activation using, e.g., predictive modeling to fully utilize the dynamic aspect of the models.

5. REFERENCES

[1] Gustavo Deco, Giulio Tononi, Melanie Boly, and Morten L. Kringelbach, "Rethinking segregation and integration: contributions of whole-brain modelling," *Nature Reviews Neuroscience 2015 16:7*, vol. 16, no. 7, pp. 430–439, 6 2015.

[2] Maria Giulia Preti, Thomas AW Bolton, and Dimitri Van De Ville, "The dynamic functional connectome: State-of-the-art and perspectives," *NeuroImage*, vol. 160, pp. 41–54, 10 2017.

[3] Elena A. Allen, Eswar Damaraju, Sergey M. Plis, Erik B. Erhardt, Tom Eichele, and Vince D. Calhoun, "Tracking whole-brain connectivity dynamics in the resting state," *Cerebral Cortex*, vol. 24, no. 3, pp. 663–676, 3 2014.

[4] Nora Leonardi and Dimitri Van De Ville∗, "Comments and Controversies On spurious and real fluctuations of dynamic functional connectivity during rest," 2014.

[5] Joana Cabral, Diego Vidaurre, Paulo Marques, Ricardo Magalhães, Pedro Silva Moreira, José Miguel Soares, Gustavo Deco, Nuno Sousa, and Morten L. Kringelbach, "Cognitive performance in healthy older adults relates to spontaneous switching between states of functional connectivity during rest," *Scientific Reports*, vol. 7, no. 1, pp. 1–13, 2017.

[6] Inderjit S. Dhillon, Edward M. Marcotte, and Usman Roshan, "Diametrical clustering for identifying anti-correlated gene clusters," *Bioinformatics*, vol. 19, no. 13, pp. 1612–1619, 2003.

[7] Anders S. Olsen, Anders Lykkebo-Valløe, Brice Ozenne, Martin K. Madsen, Dea S. Stenbæk, Sophia Armand, Morten Mørup, Melanie Ganz, Gitte M. Knudsen, and Patrick M. Fisher, "Psilocybin modulation of time-varying functional connectivity is associated with plasma psilocin and subjective effects," *NeuroImage*, p. 119716, 10 2022.

[8] G. S. Watson, "Equatorial Distributions on a Sphere," *Biometrika*, vol. 52, no. 1/2, pp. 193, 6 1965.

[9] Suvrit Sra and Dmitrii Karp, "The multivariate watson distribution: Maximum-likelihood estimation and other aspects," *Journal of Multivariate Analysis*, vol. 114, no. 1, pp. 256–269, 2013.

[10] Christopher Bingham, "An Antipodally Symmetric Distribution on the Sphere," *The Annals of Statistics*, vol. 2, no. 6, pp. 1201–1225, 11 1974.

[11] David E. Tyler, "Statistical Analysis for the Angular Central Gaussian Distribution on the Sphere," *Biometrika*, vol. 74, no. 3, pp. 579, 9 1987.

[12] Peter M. Rasmussen, Lars K. Hansen, Kristoffer H. Madsen, Nathan W. Churchill, and Stephen C. Strother, "Model sparsity and brain pattern interpretation of classification models in neuroimaging," *Pattern Recognition*, vol. 45, no. 6, pp. 2085–2100, 6 2012.

[13] Diederik P. Kingma and Jimmy Lei Ba, "Adam: A Method for Stochastic Optimization," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 12 2014.

[14] Alexander Schaefer, Ru Kong, Evan M Gordon, Timothy O Laumann, Xi-Nian Zuo, Avram J Holmes, Simon B Eickhoff, and B T Thomas Yeo, "Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI," *Cerebral Cortex*, vol. 28, no. 9, pp. 3095–3114, 9 2018.

[15] Mingrui Xia, Jinhui Wang, and Yong He, "BrainNet Viewer: A Network Visualization Tool for Human Brain Connectomics," *PLoS ONE*, vol. 8, no. 7, pp. e68910, 7 2013.