Graph Kernel Learning for Connectome Analysis

Authors removed for blind review

No Institute Given

Abstract. This paper presents a novel graph-based kernel learning approach for EEG-derived connectome analysis. Specifically, we demonstrate how to leverage the naturally available structure within the graph representation to encode prior knowledge in the kernel. We first proposed a matrix factorization to directly extract structural features from natural symmetric graph representations of EEG-connectome data. We then used them to derive a structure-persevering symmetric graph kernel to be fed into the support vector machine. The proposed approach has the advantage of being clinically interpretable. Quantitative evaluations on challenging emotion recognition tasks with respect to different frequency bands demonstrate the superior performance of our proposed method against the state-of-the-art. Results show that relevant EEG-connectome information is primarily encoded in the alpha band during the emotion regulation task.

Keywords: EEG · connectome · graph kernel · SVM · emotion regulation · brain network

1 Introduction

Connectome analysis has been an appealing research topic for almost a decade in neuroscience. The challenge originates from the so called "connectome", namely the highly organized connectivity matrix of the complex brain network [1]. A great deal of research efforts have been devoted to investigate the connectome, with particular emphasis on the use of electroencephalography (EEG)-derived sources [2,3]. Compared with other brain imaging techniques, EEG exhibits superior time resolution and allows frequency specific analyses to achieve more sensitive response for different tasks. Therefore, it is imperative that robust analytical methods be developed and refined to keep up with the growing volume and quality of EEG-connectome data; in particular, it is desirable to develop accurate predictive methods as a complement to the effort of pathologists in diagnosis process and treatment decision-making.

In recent years, many methods have been established for such purposes, most of them are based on graph-theoretic techniques [4–6], which basically derive feature vectors from the connectivity matrix of the EEG-connectome network (i.e., graph) and feed them to a conventional machine learning algorithm. However, these solutions may not be optimal for leveraging the use of correlation within the connectome. When converting graphs into vectors, some important connectivity
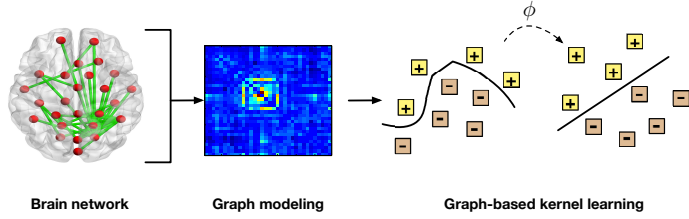
Fig. 1. The framework of graph-based kernel learning.

information will be lost. Ideally, the EEG-connectome data should be viewed as a whole graph and relevant analysis should be derived throughout from the whole graph rather than vectorization. High-dimensional prediction methods have been available for some time [7–9], which aim to better capture multivariate nonlinear relationships in the data. A critical step underlying the success of such methods is effective feature extraction and selection based on factorization techniques. However, most existing matrix factorization techniques are designed in the general matrix setting without taking into account the symmetric characteristics of these data. In this paper, we propose a novel graph-based kernel learning approach inspired by [8] for EEG-connectome predictive analysis and a new matrix factorization scheme to handle the symmetric case as well as sparse layouts. We perform evaluations on a challenging EEG-connectome emotion regulation task. The proposed framework is illustrated in Fig. 1. The contributions of this work are twofold:

– Our methods directly extract features from graph representations of EEG-connectome using a new matrix factorization. The extracted latent features are then used to derive a Structure-preserving Symmetric Graph Kernel (SSGK) to be fed into the Support Vector Machine (SVM). Superior performance on multiclass emotion regulation task with respect to different frequency bands is achieved without requiring vectorization of data.
– Our SSGK learning scheme allows for good interpretability and freedom in choosing the base kernels. It can be easily extended to the study of other types of connectome problems.

2   Preliminaries

In this section we first introduce some notations and basic operations that will be used throughout this paper, then we review some aspects of the kernel learning problem.

   Notations and Basic Operations. Following [10], we denote vectors by lowercase boldface letters, e.g., $\mathbf{x}$; and matrices by uppercase boldface, e.g., $\mathbf{X}$. An index is denoted with a lowercase letter, spanning the range from 1 to the uppercase letter of the index, e.g., $i = 1, 2, \cdots, I$. We denote a matrix as $\mathbf{A} \in \mathbb{R}^{I \times J}$, and their elements by $a_{i,j}$. We will often use calligraphic letters ($\mathcal{A}$, $\mathcal{B}$,

$\mathcal{C}, \cdots$) to denote general spaces. Specifically, the inner product of two matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{I \times J}$ is defined as $\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i=1}^{I} \sum_{j=1}^{J} a_{i,j} b_{i,j}$. The Frobenius norm of a matrix $\mathbf{A}$ is defined as $\|\mathbf{A}\|_F = \sqrt{\langle \mathbf{A}, \mathbf{A} \rangle} = \sqrt{\sum_{i=1}^{I} \sum_{j=1}^{J} a_{i,j}^2}$. The ℓ_1 norm of a vector is defined as the sum of the absolute values of its elements. A rank-one matrix $\mathbf{A}$ equals to the outer product of two vectors: $\mathbf{A} = \mathbf{u} \otimes \mathbf{v}$, where $a_{i,j} = u_i v_j$. Note that for rank-one matrices it holds that

$$\langle \mathbf{a} \otimes \mathbf{b}, \mathbf{u} \otimes \mathbf{v} \rangle = \langle \mathbf{a}, \mathbf{u} \rangle \langle \mathbf{b}, \mathbf{v} \rangle. \tag{1}$$

Kernel Learning. In a typical prediction task, given a collection of n training examples $\{\mathbf{x}_i, y_i\}_{i=1}^{n} \subset \mathcal{X} \times \mathcal{Y}$, where $\mathbf{x}_i \in \mathbb{R}^I$ is the input samples, and y_i is the class label of $\mathbf{x}_i$, the goal is to find a function $f : \mathcal{X} \to \mathcal{Y}$ that accurately predicts the label of an unseen example in $\mathcal{X}$. Support Vector Machines (SVMs) are one of the most popular kernel-based learning algorithms, which are effective on the data by linear boundaries, and in order to extend classifier functionality to classify by non-linear boundaries the kernel functions are used [11]. The kernel function encapsulates the hypothesis language, i.e., how to perform data transformation and knowledge encoding. In general, it maps data from original input feature space to a higher dimensional feature space (known as Hilbert space), a kernel function corresponds to the inner product in this higher dimensional feature space. The computational attractiveness of kernel methods comes from the fact that quite often a closed form of 'feature space inner products' exists [12]. Instead of mapping the data explicitly, the kernel can be calculated directly. According to *Mercer's theorem* [13], we can verify whether a kernel function is valid by the following Theorem [14].

Theorem 1 *A function κ defined on $: \mathcal{X} \times \mathcal{X}$ is a positive definite kernel of $\mathcal{H}$ if and only if there exists a feature mapping function $\phi : \mathcal{X} \mapsto \mathcal{H}$ such that*

$$\kappa(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle \tag{2}$$

for any $(\mathbf{x}, \mathbf{y}) \in \mathfrak{X} \times \mathfrak{X}$.

In particular, an important property of positive definite kernels is that they are closed under sum, multiplication by a scalar and product [15].

3   Methods

Our proposed method uses the EEG-connectome data represented by the mean connectivity over time and frequency domains as input features and model them directly as symmetric matrices (graphs). We propose a new matrix factorization scheme to learn bilinear symmetric bases from these graph input to obtain latent features. We then use the bilinear symmetric bases to define a structure-preserving symmetric graph kernel function (SSGK) for the SVM classifier. We present the key steps of our methods in detail as below.

Feature Extraction. Graph provides a natural representation for connectome data, but there is no guarantee that such representation will be good for kernel learning. Since learning will only be successful if the regularities that underlie the data can be discerned by the kernel. From the characteristics of connectome object, we know that the essential information in the connectome is embedded in the structure of the graph. Thus, one important aspect of kernel learning for such complex objects is to represent them by sets of key structural features which are easier to manipulate. In previous work, it was found that matrix factorization is particularly effective for extracting this structure. It can take the correlation in the graph matrix into account and represent it directly into a sum of rank one matrices (bilinear bases), yielding a more compact representation of connectome data. Motivated by these observations, we use matrix factorization for feature extraction. In particular, given a graph matrix $\mathbf{X} \in \mathbb{R}^{I \times I}$, we solve the following optimization problem:

$$\min_{\mathbf{a}_r} \|\mathbf{X} - \sum_{r=1}^{R} \mathbf{a}_r \otimes \mathbf{a}_r\|_F^2 + \lambda \sum_{r=1}^{R} \|\mathbf{a}_r\|_1. \tag{3}$$

where R is the rank of the matrix $\mathbf{X}$ defined as the smallest number of rank-one matrices in an exact matrix factorization, $\|.\|_F$ is is the Frobenius norm of the matrix, and $\|.\|_1$ is the ℓ_1 norm for sparse solution (known as lasso regularization). Equation (3) can be solved by the tensorlab toolbox [16] in Matlab.

Graph Structure Mapping. Note that although matrix factorization factorizes the graph matrix, we can still preserve the graph structure and recover the original from the factorized results. We show how the above feature extraction results can be exploited to induce a structure-preserving graph kernel. Suppose we are given the matrix factorization of $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{I \times I}$ by $\mathbf{X} = \sum_{r=1}^{R} \mathbf{x}_r \otimes \mathbf{x}_r$ and $\mathbf{Y} = \sum_{r=1}^{R} \mathbf{y}_r \otimes \mathbf{y}_r$ respectively. We assume the graph observations are mapped into the Hilbert space $\mathfrak{H}$ by

$$\phi : \mathbf{X} \to \phi(\mathbf{X}) \in \mathbb{R}^{H \times H}. \tag{4}$$

Importantly, the mapping result $\phi(\mathbf{X})$ is still a symmetric matrix, but its dimension is high, even infinite depending on the feature mapping function $\phi(.)$.

Based on the definition of the kernel function, we know that the feature space is a high-dimensional space of the original space, equipped with the same operations. Thus, we can factorize graph data directly in the feature space in the same way as in the original space. This is formally equivalent to performing the following mapping:

$$\phi : \sum_{r=1}^{R} \mathbf{x}_r \otimes \mathbf{x}_r \to \sum_{r=1}^{R} \phi(\mathbf{x}_r) \otimes \phi(\mathbf{x}_r). \tag{5}$$

In this sense, it corresponds to mapping graphs into high-dimensional graphs that retain the original structure. More precisely, it can be regarded as mapping the original graph matrix to matrix feature space and then conducting the matrix factorization in the feature space, as illustrated in Fig. 2.
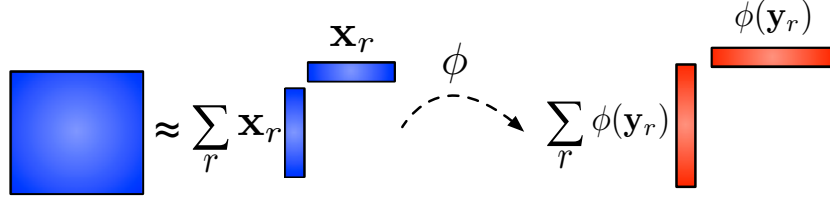
Fig. 2. Schemic diagram of the feature extraction and graph structure mapping

After mapping the matrix factorization into the outer product feature space, the kernel can be defined directly with the inner product in that feature space. Thus, based on equation (1), we can derive our SSGK model:

$$
\begin{aligned}
\kappa(\mathbf{X}, \mathbf{Y}) &= \kappa\Big(\sum_{r=1}^{R} \mathbf{x}_r \otimes \mathbf{x}_r, \sum_{r=1}^{R} \mathbf{y}_r \otimes \mathbf{y}_r\Big) \\
&= \Big\langle \sum_{r=1}^{R} \phi(\mathbf{x}_r) \otimes \phi(\mathbf{x}_r), \sum_{r=1}^{R} \phi(\mathbf{y}_r) \otimes \phi(\mathbf{y}_r) \Big\rangle \\
&= \sum_{p=1}^{R} \sum_{q=1}^{R} \kappa(\mathbf{x}_p, \mathbf{y}_q) \kappa(\mathbf{x}_p, \mathbf{y}_q).
\end{aligned}
\tag{6}
$$

Based on the Theorem 1, it is easy to see that this kernel is 'valid' as it is described as an inner product of two matrices $\sum_{r=1}^{R} \phi(\mathbf{x}_r) \otimes \phi(\mathbf{x}_r)$ and $\sum_{r=1}^{R} \phi(\mathbf{y}_r) \otimes \phi(\mathbf{y}_r)$. From the derivation process, we know that such a kernel can take into account the flexibility of graph structure. In general, SSGK is an extension of the conventional kernels in the vector space to matrix space, and each vector kernel can be used in this framework for EEG-connectome analysis in conjunction with kernel machines. Our positive result can be viewed as saying that designing a good graph kernel function is much like designing a good graph structure in the feature space.

4   Experiments and Discussions

Data. Data were collected from 22 healthy participants at *** and from 11 healthy participants at @@@[1]. Each participant underwent an Emotion Regulation Task (ERT). During the ERT session, participants were instructed to look at pictures displayed on the screen. Emotionally neutral pictures (e.g., landscape, everyday objects) and negative pictures (e.g., car crash, nature disasters) would appear on the screen for seven seconds in random orders. One second after the picture on display, a corresponding auditory guide would instruct the participant to *neutral*: viewing the neutral pictures; to *maintain*: viewing the negative

---

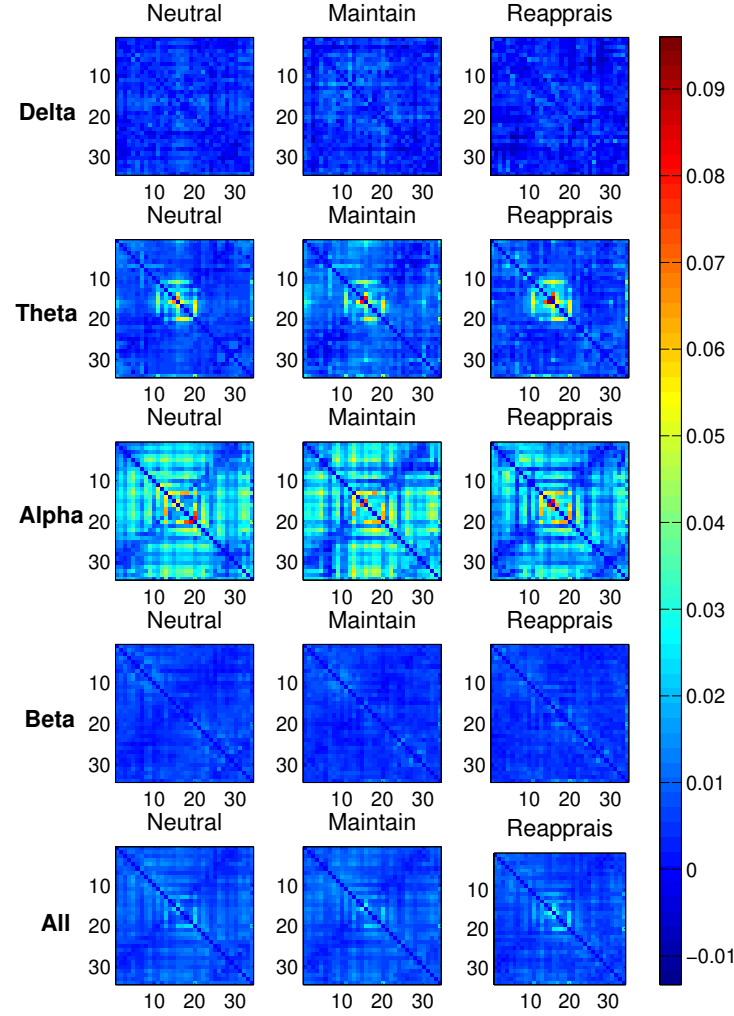[1] *** and @@@ are used for blind review.

Fig. 3. Average EEG-connectome during neutral, maintain and reappraise in the five different frequency bands.

pictures as they normally would; or to *reappraise*: viewing the negative pictures while attempting to reduce their emotion response by re-interpreting the meaning of pictures. All subjects were recorded using the Biosemi system equipped with an elastic cap with 34 scalp channels. The acquisition connectivity matrix

is 34×34 with 130 time points and 50 frequencies ranging from 1Hz to 50Hz in increments of 1Hz.

Tasks. We study multi-class EEG-connectome emotion regulation tasks and analyze the effect of different frequency bands of EEG signals. In emotion regulation, studies have shown that relevant EEG information is primarily encoded in the low frequency bands [17, 18]. Thus, we analyze the EEG-connectome data in 5 frequency bands, and they are: Delta (1–3 Hz), Theta (4–7 Hz), Alpha (8–12 Hz), Beta (13–30 Hz), for relative power, as well as the total power of the EEG (1–30 Hz) [19]. The average EEG-connectome during neutral, maintain and reappraise in the five different frequency bands are shown in Figure 4 where the $x-$ and $y-$ axes represent the vertex id, and the color of the cell represents the strength of the connectivity between vertex x and y. We can see that the connectivity in the alpha band is generally stronger than other frequency bands, and theta band is the second one.

Algorithms. We evaluate eight algorithms in Table 1 on the five tasks above, each of which represents a different strategy: the edge based feature extraction (Edge), where edge values were directly used as features by flatting connectivity matrices of EEG-connectome into vectors; the local clustering coefficients (CC) [20] which measures a network's local segregation; the characteristic path length (CPL) [21], which quantifies the global information integration; the graph-based substructure pattern mining (gSpan) [22], which is a discriminative subgraph selection approach; the dual structure-preserving kernel (DuSK) [8], which accepts general tensors as input. We use second- (i.e., averaged over time and frequency), third- (i.e., averaged over time or frequency) and fourth-order (i.e., all data with dimension $34 \times 34 \times 130 \times x$, where x corresponds to a number of the frequency level) version of this scheme, denoted as $DuSK-2D$, $DuSK-3D_{freq}$, $DuSK-3D_{time}$, $DuSK-4D$ respectively; the proposed method and its variant without sparse-constraint (SSGK and SSGK$_{\text{w/o sparse}}$).

Experimental Settings. We use the subjects collected from *** as the training set, and @@@ as the testing set in all the experiments. Following [8], we use the SVM with the Gaussian RBF kernel as the base classifier for all methods. We use the classification accuracy as the evaluation metric.

Algorithm Settings. All of the methods select the optimal trade-off parameter of SVM and kernel width parameter from $\{2^{-8}, 2^{-7}, \cdots, 2^8\}$. Other parameters for gSpan and DuSK are set following [22] and [8], repectively. For our SSGK and SSGK$_{\text{w/o sparse}}$ methods, the parameter R and λ was automatically selected from the value set of $R = \{1, 2, \cdots, 12\}$ and $\lambda = \{2^{-2}, 2^{-1}, \cdots, 2^8\}$ by the grid search.

Results. Detailed results are listed in Table 1. From Table 1, it can be seen that the proposed SSGK-based methods outperform all compared methods by 10%- 20% on almost all five different frequency bands. The superiority of the proposed methods demonstrate the effectiveness of utilizing the structure within the graph representation during encoding. More specifically, among all five frequency bands, SSGK produces the best performance on Alpha band and second best performance on Theta band, which can also be observed in our visualization

Table 1. The classification accuracy in percentage (%) by six competing methods and the proposed two methods for the five tasks. The best results for each task are highlighted in bold font.

Category	Method	Frequency Band				
		Delta	Theta	Alpha	Beta	All
Traditional	Edge	42.42	54.55	51.52	51.52	45.45
	CC	54.55	54.55	42.42	51.52	42.42
	CPL	48.48	42.42	45.45	48.48	39.39
	gSpan	39.39	51.52	39.39	54.55	48.48
	DuSK–2D	51.52	63.64	51.51	51.52	54.55
	DuSK–3D$_{freq}$	57.58	57.58	57.58	54.55	48.48
	DuSK–3D$_{time}$	57.58	54.55	57.58	54.55	54.55
	DuSK–4D	54.55	54.55	51.52	54.55	57.58
Deep Learning	CNN–2D	51.11 ±11.76	43.71 ±5.13	43.07 ±10.89	42.54 ±6.25	41.48 ±5.13
	CNN–3D$_{freq}$	46.67 ±12.38	45.93 ±5.59	41.48 ±6.79	57.04 ±4.63	44.44 ±10.18
	GCN–2D					
Ours	SSGK$_{w/o\ sparse}$	57.58	66.67	63.64	54.55	57.58
	SSGK	**63.64**	**69.70**	**<u>72.73</u>**	**60.61**	**57.58**

in Figure 3. Furthermore, by comparing SSGK and SSGK$_{w/o\ sparse}$, it is noticed that the proposed SSGK approach with sparse regularization consistently outperforms the same approach without sparse regularization, and the advantage of sparsity characterization indicates the importance of modeling the redundant information of observed frequency bands.

5   Conclusion

In this paper, a graph-based kernel learning approach called Structure-preserving Symmetric Graph Kernel (SSGK) is proposed to deal with EEG-derived connectome classification task. The proposed method mainly follows two consecutive steps: first, a sparse-inducing symmetric matrix factorization strategy is applied to extract structural features from natural symmetric graph representations of EEG-connectome data, then the extracted structural features are directly used to define the SSGK function and further fed into the support vector machine for classification. The proposed method is clinically interpretable and is able to encode prior knowledge in the kernel with structural information in graph representation. Experimental results on challenging emotion recognition task demonstrates the effectiveness of the proposed method for encoding relevant EEG-connectome information.

References

1. J. DeFelipe, "From the connectome to the synaptome: an epic love story," *Science*, vol. 330, no. 6008, pp. 1198–1201, 2010.
2. C. M. Michel and M. M. Murray, "Towards the utilization of eeg as a brain imaging tool," *Neuroimage*, vol. 61, no. 2, pp. 371–385, 2012.
3. F. Vecchio, F. Miraglia, and P. M. Rossini, "Connectome: Graph theory application in functional brain network architecture," *Clinical neurophysiology practice*, vol. 2, pp. 206–213, 2017.
4. F. Deligianni, M. Centeno, D. W. Carmichael, and J. D. Clayden, "Relating resting-state fmri and eeg whole-brain connectomes across frequency bands," *Front Neurosci*, vol. 8, 2014.
5. M. Kaiser, "A tutorial in connectome analysis: topological and spatial features of brain networks," *Neuroimage*, vol. 57, no. 3, pp. 892–907, 2011.
6. S. Ponten, A. Daffertshofer, A. Hillebrand, and C. J. Stam, "The relationship between structural and functional connectivity: graph theoretical analysis of an eeg neural mass model," *Neuroimage*, vol. 52, no. 3, pp. 985–994, 2010.
7. Z. Hao, L. He, B. Chen, and X. Yang, "A linear support higher-order tensor machine for classification," *IEEE Transactions on Image Processing*, vol. 22, no. 7, pp. 2911–2920, 2013.
8. L. He, X. Kong, P. S. Yu, X. Yang, A. B. Ragin, and Z. Hao, "Dusk: A dual structure-preserving kernel for supervised tensor learning with applications to neuroimages," in *SDM*.   SIAM, 2014, pp. 127–135.
9. A. H. Phan and A. Cichocki, "Tensor decompositions for feature extraction and classification of high dimensional datasets," *Nonlinear theory and its applications, IEICE*, vol. 1, no. 1, pp. 37–68, 2010.
10. T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.
11. I. SenGupta, A. Kumar, and R. K. Dwivedi, "A study on handling non linear separation of classes using kernel based supervised noise clustering approach," *International Journal of Computer Applications*, vol. 975, p. 8887.
12. T. Gärtner, "A survey of kernels for structured data," *ACM SIGKDD Explorations Newsletter*, vol. 5, no. 1, pp. 49–58, 2003.
13. V. Vapnik, *The nature of statistical learning theory*.   Springer Science & Business Media, 2013.
14. A. Berlinet and C. Thomas-Agnan, *Reproducing kernel Hilbert spaces in probability and statistics*.   Springer Science & Business Media, 2011.
15. N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*.   Cambridge university press, 2000.
16. L. Sorber, M. V. Barel, and L. D. Lathauwer, "Tensorlab v2.0," http://www. tensorlab.net/, January 2014.
17. M. Balconi, E. Grippa, and M. E. Vanutelli, "What hemodynamic (fnirs), electrophysiological (eeg) and autonomic integrated measures can tell us about emotional processing," *Brain and cognition*, vol. 95, pp. 67–76, 2015.
18. M. Xing, R. Tadayonnejad, A. MacNamara, O. Ajilore, K. L. Phan, H. Klumpp, and A. Leow, "EEG based functional connectivity reflects cognitive load during emotion regulation," in *ISBI*.   IEEE, 2016.
19. A. Lutz, H. A. Slagter, N. B. Rawlings, A. D. Francis, L. L. Greischar, and R. J. Davidson, "Mental training enhances attentional stability: neural and behavioral evidence," *The Journal of Neuroscience*, vol. 29, no. 42, pp. 13 418–13 427, 2009.

20. M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: uses and interpretations," *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, 2010.
21. D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world'networks," *nature*, vol. 393, no. 6684, pp. 440–442, 1998.
22. X. Yan and J. Han, "gspan: Graph-based substructure pattern mining," in *ICDM*. IEEE, 2002, pp. 721–724.