Sub-network Based Kernels for Brain Network Classification

Biao Jie
Dept. of Computer Science and Technology
Anhui Normal University
Wuhu 241003, China
jbiao@nuaa.edu.cn

Mingxia Liu
Dept. of Computer Science and Engineering
Nanjing University of Aeronautics and Astronautics
Nanjing 210016, China
plumlmx@126.com

Xi Jiang
Dept. of Computer Science
The University of Georgia
GA 30602-7404, USA
superjx2318@gmail.com

Daoqiang Zhang*
Dept. of Computer Science and Engineering
Nanjing University of Aeronautics and Astronautics
Nanjing 210016, China
dqzhang@nuaa.edu.cn

ABSTRACT

In brain network analysis, a challenging problem is deciding how to measure the similarity between a pair of networks. Recently, graph kernels have been proposed for measuring the similarity between brain networks. However, existing graph kernels are mainly defined on general graphs that ignores specific characteristics of brain networks, such as the uniqueness of nodes (i.e., each node corresponds to a unique brain region). Accordingly, in this paper, we construct a novel sub-network based kernel for brain networks and apply it for mild cognitive impairment (MCI) classification. Experimental results on a real MCI dataset demonstrate that the proposed method outperform several state-of-the-art graph kernel based methods.

Keywords

Alzheimer's disease; Mild cognitive impairment; Functional connectivity network; Graph Kernel; Classification

1. INTRODUCTION

As a neurodegenerative disorder, Alzheimer's disease (AD) is the most common form of dementia in elderly population worldwide, which usually starts slowly and gets worse over time. The most common early symptom of AD is difficulty in remembering recent events. In general, AD leads to substantial, progressive neuron damage that is irreversible, which eventually causes death. Recently, a prodromal stage of AD called mild cognitive impairment (MCI) has gained increasing attention, due to its high probability of progression to AD. It is reported the patients with MCI will progress to clinical AD at an annual rate of approximately 10% to 15%, while normal controls (NC) will develop dementia at an annual rate of 1% to 2% [14]. Thus, accurate diagnosis of MCI is very important for possible early treatment and delay of AD progression.

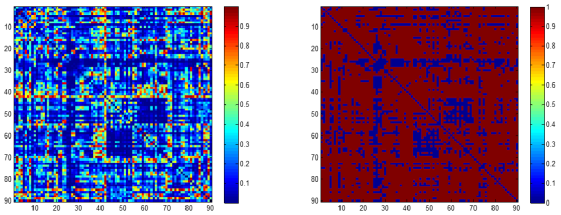
In the literature, a large amount of evidences from both anatomical and physiological studies suggest that cognitive processes depend on interactions among distributed brain regions [20]. These interaction patterns can be characterized as brain network, and, thus, help us to better understand the pathological underpinnings of neurological disorder by studying structural and functional connectivity of networks. Graph theory provides important ways to concisely quantify the connectivity properties of brain networks with each node denoting an anatomical element (e.g., brain region) and each edge corresponding to the relationship between nodes (e.g., connectivity), and has been applied to investigate the brain network of patients and normal controls [11]. Recently, graph theory has been applied to classification of brain disease [1, 27].

Compared with other network approaches, graph theory offers two important advantages [22]. First, it provides quantitative measurements of each node that preserve the connectivity information from the network and thus reflect the segregated and integrated nature of local brain activity. For example, clustering coefficient [15], which quantifies the degree to which nodes in a graph tend to cluster together, is one of the common and simple measures of functional segregation in brain network analysis. Studies have investigated the local clustering property of functional connectivity network and found the local clustering disruption of AD/MCI patients [21], which may suggest that the functional connectivity network structure becomes less modular and thus the efficient organization for information movement is lost for AD/MCI patients [21].

The second advantage of graph theory is that it provides a general framework for comparing all kinds of graphs that describe different types of data (e.g., intra-class and inter-class data, anatomical and functional data) [22]. Therefore, graph theory shows great promise to disentangle how various pathological processes in brain diseases (e.g., how the association between brain functional deficits and the underlying structural disruption related to brain disorders, and why the disease propagates along specific routes), and helps to identify

BCB '16, October 02-05, 2016, Seattle, WA, USA

DOI: http://dx.doi.org/10.1145/2975167.2985687



(a) Left (b) Right

Figure 1: The discrimative power of each connection in brain network between MCI patients and normal controls using the standard t-test on a real MCI dataset from ADNI database. Left: the p-values on connection between all pairs of ROIs. Right:The thresholded p-values on connection between all pairs of ROIs

image-based biomarkers for diagnosis of brain diseases. However, different from traditional data in feature spaces, graph data is not directly represented as feature vectors, which raises one fundamental challenge for graph data that is how to compare two graphs, i.e., how to measure the similarity between a pair of graphs. Motivated by this challenge, computing the similarity of graphs has attracted much attention in the last decade. Among all kinds of methods, kernel methods [16] offer a natural framework to study this problem. In the literature, graph kernels, i.e., the kernels constructed on graphs, have been proposed and applied to brain network analysis. However, existing graph kernels are mainly defined on general graphs and thus ignore some inherent characteristics of brain networks, such as the uniqueness of node that means each node corresponds to a unique brain region, which may affect the performance of the subsequent brain network analysis.

On the other hand, in brain network analysis, the local topological properties are very important for the similarity computation of brain networks. Fig. 1 illustrates the discriminative power of each connection in brain network between MCI patients and normal controls using the standard t-test on a real MCI dataset from Alzheimer disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI), where Fig. 1(a) shows the obtained p-values on all connections, and Fig. 1(b) shows the thresholded p-value (i.e., p-value more than 0.05 is set to 1). From Fig. 1, one can see that lots of discriminative connections (i.e., the corresponding p-value less than 0.05) mainly focus on some specific brain regions. Intuitively, exploring the local connectivity properties of brain regions can construct more effective kernel for measuring the similarity of brain networks. To the best of our knowledge, few previous works has exploited those characteristics of brain networks to construct graph kernels for measuring the similarity of brain networks.

To address these problems, in this paper, motivated by a recent work in [19], we propose a novel sub-network based kernel on brain networks and apply it for MCI classification. Different from traditional graph kernels, in the proposed method we take into account the inherent characteristic of brain network, and capture the local topological properties of brain network that are very important for distinguishing two brain networks. The underlying basic idea of our proposed method is first to construct a group of sub-networks on each node, to reflect the local connectivity properties of brain network, and then define the similarity of a pair of brain networks through calculating the similarity of all pairs of sub-network groups. We evaluate our proposed method on 149 subjects with the baseline resting-state fMRI data from the ADNI database, which contains 99 MCI patients, including 56 early MCI (EMCI) and 43 late MCI (LMCI), and 50 normal controls. The experiment results demonstrate the efficacy of our proposed method.

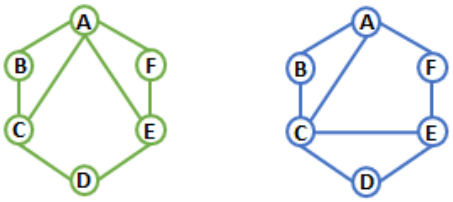
1.1 Related Works

Currently, there are a series of methods to compute the similarity of brain networks. One simple and straightforward ways is to extract some local measures of network, such as edge weight, path length and clustering coefficient, as feature vector for brain network analysis (classification). For example, Chen et al. in [5] used connectivity strengths between the brain region pairs for AD/MCI classificationm. Wee et al. in [26] extracted clustering coefficients from white matter connectivity networks for identification of MCI patients. Zanin et al. in [28] explore 16 topological features from functional connectivity networks as well as their combination to find the optimal network representation via comparing the classification performance based on those extracted features. Tijma et al. [22] investigated 13 graph properties and examined which properties have been consistently reported to be disturbed in AD studies by using group analysis. All of these studies demonstrate the advantages of application of graph theory in neuroimaging data analysis. However, those studies usually explore the local topological measures of brain networks and thus ignore their global topology characteristic.

In recent years, kernel-based methods are also proposed for measuring similarity of brain networks. Informally, a kernel is a function that measures the similarity between a pair of objects, while it corresponds to an inner product in a reproducing kernel Hilbert space [16] mathematically. Once a kernel is defined, many learning algorithms such as support vector machines (SVM) can be used, and thus many applications of classification, regression as well as group analysis is quite simple. Also, kernel method provides a general framework to calculate the similarity of graphs. Graph kernels have been proposed and successfully applied to diverse fields including neuroimaging studies [17, 13, 12], image classification [3] and protein function prediction [30]. For example, Jie et al. in [9] adopted the subtree kernels on functional brain networks for MCI classification. Mokhtari et al. in [12] also used the graph kernel for classifying between attentional cueing task and rest states from functional brain networks.

Graph kernels are instances of the family of so-called R-convolution kernels by Haussler [8]. Most of graph kernels are defined via a comparing small sub-graphs such as walks, paths or graphlets. For example, Gärtner in [7] proposed a random walk kernel that counts common walks in two graphs. Borgwardt and Kriegel in [2] constructed a shortest-path kernel through comparing features of the shortest paths between all pairs of nodes in two graphs. Since searching for structural similarities in a pair of graphs is often computationally expensive, some researchers explore graph kernels with lower computational complexity using computation techniques [6]. Recently, Shrivastava and Li in [19] defined an effective graph kernel based on a new mathematical representation for graph. Johansson et al. in [10] constructed a global graph kernel by using geometric embedding. Compared with feature-based methods, due to take full advantage of kernel method, kernel-based methods usually achieve better performance in brain network analysis.

In general, existing graph kernels can be roughly divided into two categories: 1) kernels defined on unlabeled graphs where each node has no distinct identification except through their interconnectivity, such as graph kernels in [19, 2, 10]; 2) kernels defined on labeled graphs where each node is assigned a label, such as graph kernels in [7, 6]. It is worth noting that some graph kernels (e.g., in [18]) defined on labeled graphs are also defined on unlabeled



(a) Network G (b) Network H

Figure 2: Examples of brain network (i.e., labeled graphs) with six nodes, where each node denotes a brain region, and each connection represents the correlation between a pair of nodes. Note that some connections have been changed in these networks. Since these brain networks are isomorphic if ignoring the label information of each node, graph kernel without considering label information will fail to compute the similarity of them.

graphs.

In the first category, graph kernels defined on unlabeled graphs, which ignore the label information of nodes, are often not able to compute the real similarity for a pair of labeled graphs. For example, for brain networks G and H in Fig. 2 where each node corresponds a brain region (denoted by a letter, called label of node), their connectivities have been changed due to disease affect, while they are isomorphic if we ignore the label information of each node. Therefore, graph kernels without considering label information will fail to compute the similarity for this pair of brain networks.

On the other hand, most of graph kernels in the second category, which ignore the specific information of brain network, may be also fail to compare the real similarity of a pair of brain networks. In this category, some graph kernels are infeasible for connectivity networks due to their computational complexity, e.g., in [7, 25]. Also, some graph kernels cannot be computed on brain networks. For instance, graph kernels in [25] are used to compare graphs with edge labels, and graph kernels in [6] are used to compare graphs with continuous-valued node labels. Other graph kernels such as in [18] are constructed based on Weisfeiler-Lehman test of graph isomorphism. However, there don't exist problem of isomorphism between brain networks when considering uniqueness of each node (see Fig. 2 for an example).

The main contributions of this paper are as follows: 1) we proposed a novel sub-network-based kernel for measuring the similarity between a pair of brain networks. 2) On a real MCI dataset from ADNI database, we investigate the effectiveness of proposed graph kernels via extensive experiments. 3) We provide an implementation for performing inference on brain network data. It is worth noting that our proposed graph kernel is a general similarity measure for brain networks, which can be applied to a lot of brain network analysis tasks.

2. PROPOSED GRAPH KERNEL

In this section, we first briefly introduce the existing graph kernel [19]. Motivated by that work, we propose our graph kernels for brain networks analysis.

Given a graph G (denoted by a matrix $A \in R^{m\times m}$) and a number l, where m is the number of nodes in G. To effectively represent a graph, Shrivastava and Li in [19] defined a symmetric positive semi-definite matrix $C^G \in R^{l\times l}$ as:

$$C^G(i,j) = cov(\frac{mA^ie}{\|A^ie\|_1}, \frac{mA^je}{\|A^je\|_1}) \quad (1)$$

where cov denotes the covariance between two vectors, e denotes the vector of all 1s, A^i e denotes the i-th power iteration of matrix A on a given starting vector e. Here, the number l controls the number of algorithm iterations in this new mathematical representation for graph.

Shrivastava and Li have argued that the matrix C^G can capture critical information of the underlying graph, including the spectrum of adjacency matrix A as well as counts of various sub-structures (e.g., number of triangles and number of small paths, etc.), and owns many good properties, such as graph invariant (i.e., isomorphic graphs have the same representation).

According to the definition in Eq. (1), different kinds of graphs can be represented in a common mathematical space where they can be directly compared. Furthermore, based on this new mathematical representation of graph, Shrivastava and Li in [19] also defined a kernel to compute the similarity between a pair of graphs G and H as the follows:

$$k(G,H) = exp(-\frac{1}{2}log(|\Sigma|/\sqrt{|C^G||C^H|})) \quad (2)$$

where $|\cdot|$ denotes the determinant, $\Sigma = (C^G + C^H)/2$, C^G and C^H are the corresponding covariance matrices defined on graph G and H according to Eq. (1), respectively.

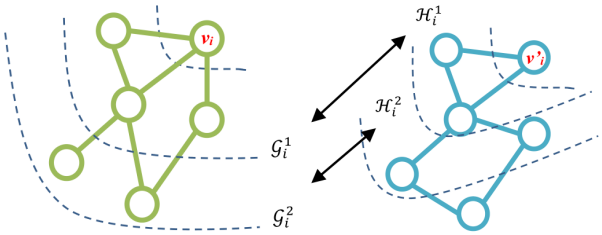
However, the kernel in Eq. (2) is defined on unlabeled graphs and thus also lacks consideration of specific characteristic of brain connectivity networks, i.e., the uniqueness information of each node. Moreover, this kernel mainly considers the global topological properties of graphs and thus ignores local topological properties of brain networks that are also very important for the similarity computation of brain networks. To address that problem, we define a novel sub-network-based kernel to measure similarity between a pair of brain networks. The basic idea of our proposed method is first to construct a group of sub-networks on each node to reflect the local connectivity properties of brain network. Then, considering the uniqueness of each node and one-to-one correspondence of nodes across different brain networks, we define the similarity of brain networks via calculating the similarity of pairs of sub-network groups from the same node across different brain networks.

Specifically, denote $\mathcal{G} = \{V, E\}$ and $\mathcal{H} = \{V', E'\}$ as a pair of brain networks and given a number h. Here, V and V' denote the corresponding node sets, and E and E' denote the corresponding edge sets. First, to reflect the local topological properties of brain networks, we respectively define a group of sub-networks on each node v_i/v_i' for network $\mathcal{G}$ and $\mathcal{H}$, i.e.,

$$\begin{aligned} \mathcal{G}_i^h &= \{G_i^j = (V_i^j, E_i^j)\}_{j=1,2,\ldots,h} \\ \mathcal{H}_i^h &= \{H_i^j = (V'^j_i, E'^j_i)\}_{j=1,2,\ldots,h} \end{aligned} \quad (3)$$

where $V_i^j = \{v \in V | s(v, v_i) \leq j\}$, $E_i^j = \{(u,v) \in E | u \in V_i^j, v \in V_i^j\}$, $V'^j_i = \{v \in V' | s(v, v'_i) \leq j\}$, $E'^j_i = \{(u,v) \in E' | u \in V'^j_i, v \in V'^j_i\}$, and $s(\cdot,\cdot)$ denotes the length of shortest-path between two nodes.

According to this definition, G_i^j and H_i^j respectively denotes a sub-network, where V_i^j in G_i^j (resp. V'^j_i in H_i^j) represents the corresponding set of nodes which consists of node i and those nodes whose shortest-path to the node i is less than or equal to j, and E_i^j in G_i^j (resp. E'^j_i in H_i^j) represents the corresponding set of edges which includes those edges (i.e., connections) occurred in G



(a) Network $\mathcal{G}$ (b) Network $\mathcal{H}$

Figure 3: Example of construction process of a group of sub-networks on the i-th node with $h = 2$ for a pair of brain networks $\mathcal{G}$ and $\mathcal{H}$.

Algorithm 1 Subnetwork-based Kernel defined on Brain Networks

Input:
Two connectivity networks $\mathcal{G}$ and $\mathcal{H}$, the number l and h
Output:
Graph Kernel $k(\mathcal{G}, \mathcal{H})$
1: Construct N sets of sub-networks for each brain connectivity network by Eq. (3) and Eq. (4)
2: Compute the similarity between $\mathcal{G}_i^h$ and $\mathcal{H}_i^h$ by Eq. (6)
3: Compute the kernel $k(\mathcal{G}, \mathcal{H})$ by Eq. (5)

(resp. H). Fig. 3 illustrates an example of construction process of two groups of sub-networks on node i. It is worth noting that the structure of each pair of sub-networks G_i^j and H_i^j on i-th node may be different, i.e., they may include different nodes and edges. Therefore, for each brain network with N nodes, we can obtain N groups of sub-networks, i.e.,

$$\begin{aligned} \mathcal{G} &= \{\mathcal{G}_1^h, \mathcal{G}_2^h, \ldots, \mathcal{G}_N^h\} \\ \mathcal{H} &= \{\mathcal{H}_1^h, \mathcal{H}_2^h, \ldots, \mathcal{H}_N^h\} \end{aligned} \tag{4}$$

Then, motivated by Shrivastava and Li's work [19], and considering the uniqueness of nodes and one-to-one correspondence of nodes across different brain networks, we can define the kernel on brain networks $\mathcal{G}$ and $\mathcal{H}$ by measuring the similarity on each pair of groups of sub-network from the same node, i.e.,

$$k(\mathcal{G}, \mathcal{H}) = \frac{1}{N} \sum_{i=1}^{N} f(\mathcal{G}_i^h, \mathcal{H}_i^h) \tag{5}$$

with

$$f(\mathcal{G}_i^h, \mathcal{H}_i^h) = \frac{1}{h} \sum_{j=1}^{h} exp(-\frac{1}{2} \log(|\Sigma_i^j| / \sqrt{|C^{G_i^j}||C^{H_i^j}|})) \tag{6}$$

Here, $|\cdot|$ denotes the determinant, $\Sigma_i^j = (C^{G_i^j} + C^{H_i^j})/2$, $C^{G_i^j}$ and $C^{H_i^j}$ represent the corresponding covariance matrices defined on sub-networks G_i^j and H_i^j by Eq. (1) through computing the first l terms of power iteration of their adjacency matrix, respectively. The detailed construction procedure of proposed graph kernel is summarized in Algorithm 1.

Theorem 1. Kernel defined in Eq. (5) is a positive semi-definite kernel.

Proof: According to [19], the function defined in Eq. (2) is positive semi-definite and a valid kernel. Therefore, the function defined in Eq. (5) is also a positive semi-definite and valid kernel according to the principle of weighted summation kernel [16].

Thus, our kernels defined in this section are valid kernels and hence can be directly used in existing kernel-based methods such as SVM. We will evaluate classification performance of our kernel on a real MCI dataset in the next section.

Computation complexity. The complexity for constructing sub-networks is $\mathcal{O}(N^2)$, according to [19] the complexity for computing similarity on each sub-graph is $\mathcal{O}(s_i^j l + t_i^j l^2 + l^3)$, where s_i^j and t_i^j denote the number of edges and nodes on each sub-graph G_i^j, respectively. Note that $s_i^j < S$, $t_i^j < N$, where S and N denote the number of edges and nodes on connectivity network, respectively. Thus, the total time complexity of computing the similarity on a pair of connectivity networks is less than $\mathcal{O}(Nh(Sl + Nl^2 + l^3) + N^2)$. Finally, the total time complexity is less than $\mathcal{O}(N(S + N) + N^2)$, since $l \ll N$ and $h \ll N$.

It is worth noting that a group of sub-networks is constructed on each node to reflect the local multi-level topological properties of brain network, and the size of each group of sub-networks is decided by the value of h. Here, multi-level means that the sub-network with larger value of h will contain much more nodes and edges, and $G_i^s \in G_i^j$ if $s < j$. Also, the kernel defined in Eq. (5) computes the similarity on each pair of groups of sub-networks from the same node across different subjects. Therefore, different from the kernel in [19], our graph kernel takes into account not only the uniqueness of nodes but also the local multi-level topological properties of brain networks.

Table 1: Characteristics of the subjects. (MMSE=Mini-Mental State Examination. Here, 16 patients and 6 Normal Controls MMSE unknown.)

GROUP	MCI	NC
NO. OF SUBJECTS (MALE/FEMALE)	47/52	21/29
AGE (MEAN ± SD)	71.7 ± 7.4	75.0 ± 6.9
MMSE (MEAN ± SD)	27.8 ± 1.8	28.9 ± 1.6

3. EXPERIMENTAL SETUP

In this section, we will give the experimental setup, including subjects used in our studies, data preprocessing, classification method and implementation details. The experiment results of proposed graph kernel will be given in the next section.

3.1 Subjects

The dataset used in the preparation of this study were obtained from the ADNI database. In current study we use a total of 149 subjects, which includes 50 normal controls and 99 MCI patients (including 56 early MCI (EMCI) and 43 late MCI (LMCI)), with each subject of MCI or NC being scanned by fMRI. All resting-state fMRI (rs-fMRI) data can be downloaded from the ADNI database. The demographic information of subjects used in our study is presented in Table 1.

3.2 Data Preprocessing

All rs-fMRI dataset were acquired on 3.0 Tesla Philips scanners (varied models/systems) at multiple sites. There is a range for imaging resolution in X and Y dimensions, which is from 2.29 mm to 3.31mm and the slice thickness is 3.31mm. TE (echo time) for all subjects is 30ms and TR (repetition time) is from 2.2s to 3.1s. For each subject, there are 140 volumes (time points).

The pre-processing steps of the rs-fMRI data include brain skull removal, slice time correction, motion correction, spatial smoothing, and temporal prewhitening using FSL FEAT software package (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FEAT). Specifically, the acquired rs-fMRI images are corrected for the acquisition time difference a-

mong all slices. All images are then aligned to the first volume for motion correction and a brain mask is also created from the first volume. At last, the global drift removal and band pass filtering between 0.01Hz 0.1Hz is performed using our in-house tool [31]. The pre-processing steps of the T1-weighted data include brain skull removal and tissue segmentation into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) using FSL FAST software package (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/ FAST). The pre-processed T1 image is then co-registered to the first volume of pre-processed rs-fMRI data of the same subject and the BOLD signals in GM are merely extracted and adopted to avoid the relatively high proportion of noise caused by the cardiac and respiratory cycles in WM and ventricle [24]. Finally, the whole brain of each subject in rs-fMRI space is parcellated into 90 regions of interest (ROI) by warping the automated anatomical labeling [23] template to each subject rs-fMRI image space using the transformation derived from T1 image in MNI_152 template space to rs-fMRI image registration using the FSL FLIRT software package (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FLIRT). For each of the 90 ROIs, the mean rs-fMRI time series was calculated by averaging the GM-masked BOLD signals among all voxels within the specific ROI. For each subject, a functional connectivity network is constructed with the vertices of network corresponding to the ROIs and the weight of edges corresponding to the Pearson correlation coefficients. Fisher's r-to-z transformation is applied on the elements of the functional connectivity network to improve the normality of the correlation coefficients.

3.3 Classification Framework

Before classification using our new graph kernel, we first construct the discriminative sub-network from original brain network by screening out those less discriminative brain regions. Specifically, we first extract local clustering coefficients[15] for constructed brain network $\mathcal{G} = [w_{ij}] \in R^{N \times N}$ as:

$$c_i = \frac{2}{d_i(d_i - 1)} \sum_{i,k} (w_{ij} w_{jk} w_{ki})^{1/3} \tag{7}$$

where d_i is the number of neighboring node around node i. These extracted clustering coefficients were composited as a feature vector and used for subsequent feature selection. Then, we perform a standard paired t-test to screen out those features that are not significant for discrimination between patients and normal controls. For instance, given training subjects, the p-value of each feature is first computed on via using standard t-test and those features with p-value larger than a given threshold is considered as unimportant features and will be omitted. Finally, those surviving features (i.e., ROI) are used to construct the sub-network with corresponding weights of edge in the original connectivity network.

Next, since the functional connectivity networks are intrinsically weighted graphs with full connection, to reflect the topological properties of brain networks, we simultaneously threshold the discriminative sub-network with multiple different predefined values. Here, we select multiple thresholds instead of single threshold since there is no golden standard to select a single threshold, we often need to explore over a range of plausible threshold to select the optimal one. Also, the brain networks with different thresholds may represent different level of topological properties (i.e., the networks with larger threshold often preserve fewer connections and thus are sparser in connection). These properties may be complementary to each other in improving the classification performance.

Moreover, we compute the graph kernels on each thresholded sub-network across different subjects according to Algorithm 1. Therefore, we can get multiple kernels with multiple different predefined thresholds. Finally, we adopt the multi-kernel SVM technique used in [29] for classification, i.e., the following MKL technique is adopted to combine the multiple graph kernels:

$$k(\mathcal{G}, \mathcal{H}) = \sum_{m=1}^{m} \mu_m k_m(\mathcal{G}, \mathcal{H}) \tag{8}$$

where $k_m(\mathcal{G}, \mathcal{H})$ denotes the kernel function over the m-th thresholded sub-networks across subject G and H, μ_m is a no-negative weight parameter with $\sum_{m=1}^{M} \mu_m = 1$, and M is the number of thresholds.

3.4 Implementation Details

In our study, we evaluate the classification performance using the leave-one-out (LOO) cross-validation with a C-support Vector Machine (SVM) classifier. The SVM was implemented based on the LIBSVM library [4] with the default parameter values (i.e., $C = 1$). For simplicity, we adopt five different values (i.e., $T = [0.30, 0.35, 0.40, 0.45, 0.50]$) to threshold the discriminative subnetwork. The corresponding average connection density (i.e., the fraction of present connections to possible connections) of each threshold is located in interval [30% 70%]. It has been reported that the average connection density interval of [25% 75%] demonstrates higher classification performance [28].

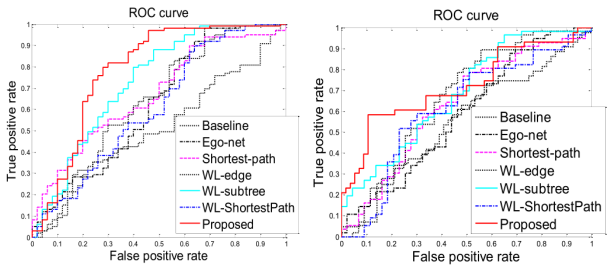
4. EXPERIMENTAL RESULTS

4.1 Classification Performance

To evaluate the effectiveness of our proposed graph kernel, we perform a series of experiments on a real MCI data from the ADNI database. In experiments, two kinds of binary classification tasks are performed, including MCI vs. NC classification, and early MCI (EMCI) vs. late MCI (LMCI) classification. We evaluated the performance of different methods by computing the classification accuracy, balanced accuracy (i.e., the average accuracy obtained on either class), and area under receiver operating characteristic (ROC) curve (AUC).

We compare our proposed subnetwork-based kernels to state-of-the-art kernels, selected so as to represent three major groups of graph kernels, i.e., based on subtrees, shortest paths and edges respectively. Those graph kernels belong to Weisfeiler-Lehman graph kernel framework proposed in [18] (denoted as WL-ST, WL-SP and WL-edge, respectively). Also, we compare the graph kernels defined by Shrivastava and Li for measuring the similarity of ego network [19] (denoted as Ego-net) and shortest-path-based kernels proposed in [2] (denoted as SP). It is worth noting that all graph kernels used for comparison are constructed by ignoring the label information of each node in brain networks. Besides, we also compare the baseline method using only local measures (i.e., weighted clustering coefficients [15]) as features where t-test is used for feature selection and a linear SVM is used to perform classification. Classification results of all methods are summarized in Table 2. Figure 4 plots the ROC curves achieved by different methods. For comparison, in Fig. 5, we also give the classification performance of different methods under different single thresholded sub-networks.

As shown in Table 2 and Fig. 4, the proposed method consistently outperforms the other methods on both classification tasks. Specifically, the proposed method yields a classification accuracy of 82.6% and 74.8% for MCI vs. NC and EMCI vs. LMCI classification, respectively, while the best classification accuracy of other methods are 76.5% and 70.7%, respectively. Also, our proposed graph kernel yields a balanced accuracy of 75.5% and 72.6% for



(a) MCI vs. NC (b) EMCI vs. LMCI

Figure 4: The ROC curves of different methods. (a) classification of MCI vs. NC; (b) classification of EMCI vs. LMCI.

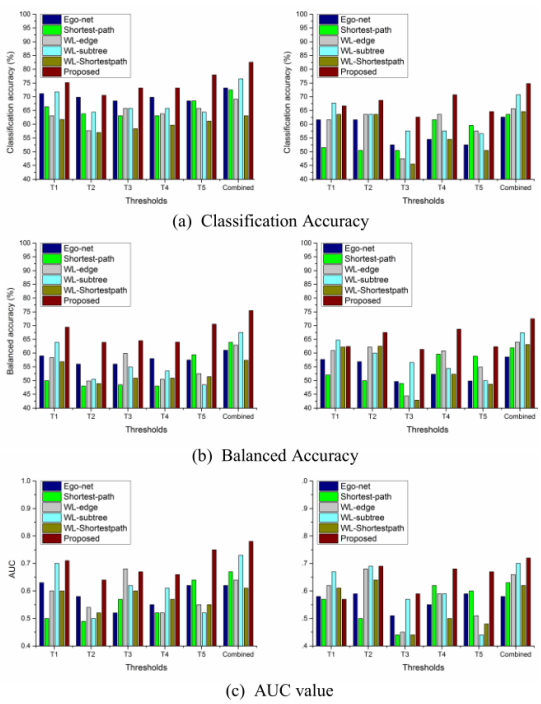
Table 2: Classification performances of different methods

	MCI vs. NC			EMCI vs. LMCI		
KERNELS	ACC	BAC	AUC	ACC	BAC	AUC
BASELINE	65.1	51.5	0.51	55.6	50.0	0.55
EGO-NET	73.2	61.0	0.62	62.6	58.6	0.58
SP	72.5	64.0	0.67	63.6	61.9	0.63
WL-EDGE	69.1	62.9	0.64	65.7	64.0	0.66
WL-ST	76.5	67.5	0.73	70.7	67.4	0.70
WL-SP	63.1	57.4	0.61	64.6	63.1	0.62
PROPOSED	82.6	75.5	0.78	74.8	72.6	0.72

both classification tasks, respectively, while the best balanced accuracy of other methods are 67.5% and 67.4%, respectively. In addition, the AUC of proposed method, respectively, is 0.78 and 0.72 for those classifications, which indicates excellent diagnostic power. Besides, one could observe from Fig. 5 that: 1) the combination of multiple thresholded networks performed significantly better than using any single thresholded network alone, 2) the performance of proposed graph kernel on each thresholded subnetwork outperform that of the state-of-the-art graph kernels, which again shows the efficacy of the proposed graph kernel.

4.2 Discriminative Power of Proposed Graph Kernels

In this subsection, we investigate the discriminative power of our proposed graph kernels. Since the selected ROIs are different in each LOO cross-validation fold, which leads that constructed discriminative sub-networks are different, we first construct the common discriminative sub-network from the original connectivity network of each subject with nodes corresponding the ROIs (i.e., clustering coefficient features) occurring in all LOO cross-validation and weights corresponding weights of edge in the original connectivity network. Then, we use the same thresholds that were used in classification step to simultaneously threshold the common discriminative sub-network. Finally, based on each thresholded common discriminative sub-network, we performed a significance test between two groups of group kernels, i.e., graph kernels defined on intra-class subjects (i.e., subjects with the same class label (patients or normal controls)) and graph kernels constructed on inter-class subjects (i.e., subjects with the different class label), using the standard paired t-test. The obtained p-values for five thresholded discriminative sub-networks are 2.39e-24, 1.22e-08, 2.04e-03, 2.31e-10 and 7.76e-05, respectively. The results show graph kernels constructed on intra-class subjects are significance larger than that defined on inter-class subjects (i.e., the corresponding p-values are very small), which indicates our proposed graph kernels can



(a) Classification Accuracy

(b) Balanced Accuracy

(c) AUC value

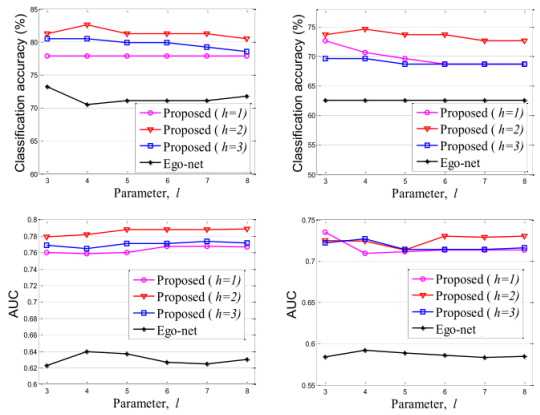
Figure 5: Classification performance of different methods on each thresholed network. Left: classification of MCI and NC; Right: classification of EMCI and LMCI. Here, T_1 , T_2 , T_3 , T_4 and T_5 denote using the individual thesholded network, respectively, while combined denotes using all thesholded networks.

capture the topological properties of brain networks and thus use for the similarity measure and classification of them.

4.3 Effect of Parameters

In our proposed graph kernel, it includes two parameters, i.e., the number l and h. the number l controls the number of algorithm iterations in computing the mathematical representation for graphs. The number h controls the scale of each group of sub-networks. To investigate the effects of the number l and h on classification performance of our proposed graph kernels, we test the different values for the number l from {3, 4, 5, 6, 7, 8}, and number h from {1, 2, 3}. Fig. 6 shows the classification results with respect to different values of the number l and h. For comparison, in Fig. 6 we also give the classification performance of ego-network-based kernel in [19] with different values of l (i.e., l={3, 4, 5, 6, 7, 8}).

As we can see from Fig. 6 and Table 2, for all parameter values, the proposed method significantly outperforms the ego-network-based method [19] on both classification tasks. Also, for most of values of l and h, the classification performance of our defined graph kernel outperforms those of other graph kernels, which further shows the efficacy of our graph kernel. In addition, Fig. 6 shows that, with a fixed h, the varied curves with the value of l are very smooth, which shows that our method is very robust to the parameter l. Moreover, we can observe from Fig. 6 that, given a fixed l, the classification performance is largely affected with different values of h. When $h = 2$, our graph kernel obtains the best classification performance. Those results imply that the selection of h is very important for the constructed graph kernel. This is reasonable since the number h controls the size of each group of sub-networks and thus determines the local similarity measures of



(a) MCI vs. NC (b) EMCI vs. LMCI

Figure 6: The classification accuracy (up) and AUC value (down) of proposed method w.r.t. the selection of l and h.(a) Classification of MCI vs. NC; (b) classification of EMCI vs. LMCI.

brain networks.

5. CONCLUSION

The similarity computation between graphs (or networks) is a fundamental challenge in network-based analysis. In this paper, we have developed a novel sub-network-based kernel for measuring the similarity between a pair of brain networks. Different from existing graph kernels, our new constructed graph kernels can effectively reflect the specific characteristics of brain networks and capture the local multi-level topological properties of brain networks. We apply our proposed graph kernel for brain network based MCI classification, with experimental results demonstrating the efficacy of our proposed method.

6. ACKNOWLEDGMENTS

This work was supported in part by National Natural Science Foundation of China (Nos. 61573023, 61422204, 61473149, 61473190), Natural Science Foundation of Anhui Province (No. 1508085MF125), the Jiangsu Natural Science Foundation for Distinguished Young Scholar (No. BK20130034), the Open Projects Program of National Laboratory of Pattern Recognition (No. 201407361), the Specialized Research Fund for the Doctoral Program of Higher Education (No. 20123218110009), Provincial Natural Science Research Key Project of Anhui Colleges (No. KJ2016A275), the NUAA Fundamental Research Funds (No. NE2013105), the Doctoral Scientific Research Foundation of AHNU (No. 2016XJJ120).

References

[1] M. R. Arbabshirani, K. A. Kiehl, G. D. Pearlson, and V. D. Calhoun. Classification of schizophrenia patients based on resting-state functional network connectivity. *Front. Neurosci*, 7(133):10–3389, 2013.

[2] K. M. Borgwardt and H.-P. Kriegel. Shortest-path kernels on graphs. In *Data Mining, Fifth IEEE International Conference on*, pages 8–pp. IEEE, 2005.

[3] G. Camps-Valls, N. Shervashidze, and K. M. Borgwardt. Spatio-spectral remote sensing image classification with graph kernels. *Geoscience and Remote Sensing Letters, IEEE*, 7(4):741–745, 2010.

[4] C.-C. Chang and C.-J. Lin. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011.

[5] G. Chen, B. D. Ward, C. Xie, W. Li, Z. Wu, J. L. Jones, M. Franczak, P. Antuono, and S.-J. Li. Classification of alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network analysis based on resting-state functional mr imaging. *Radiology*, 259(1):213–221, 2011.

[6] A. Feragen, N. Kasenburg, J. Petersen, M. de Bruijne, and K. Borgwardt. Scalable kernels for graphs with continuous attributes. In *Advances in Neural Information Processing Systems*, pages 216–224, 2013.

[7] T. Gärtner, P. Flach, and S. Wrobel. On graph kernels: Hardness results and efficient alternatives. In *Learning Theory and Kernel Machines*, pages 129–143. Springer, 2003.

[8] D. Haussler. Convolution kernels on discrete structures. Technical report, Citeseer, 1999.

[9] B. Jie, D. Zhang, C.-Y. Wee, and D. Shen. Topological graph kernel on multiple thresholded functional connectivity networks for mild cognitive impairment classification. *Human brain mapping*, 35(7):2876–2897, 2014.

[10] F. Johansson, V. Jethava, D. Dubhashi, and C. Bhattacharyya. Global graph kernels using geometric embeddings. In *Proceedings of the 31st International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*, 2014.

[11] M. Kaiser. A tutorial in connectome analysis: topological and spatial features of brain networks. *Neuroimage*, 57(3):892–907, 2011.

[12] F. Mokhtari, S. K. Bakhtiari, G. A. Hossein-Zadeh, and H. Soltanian-Zadeh. Discriminating between brain rest and attention states using fmri connectivity graphs and subtree svm. In *SPIE Medical Imaging*, pages 83144C–83144C. International Society for Optics and Photonics, 2012.

[13] D. Pachauri, C. Hinrichs, M. K. Chung, S. C. Johnson, and V. Singh. Topology-based kernels with application to inference problems in alzheimer's disease. *Medical Imaging, IEEE Transactions on*, 30(10):1760–1770, 2011.

[14] R. C. Petersen, R. Doody, A. Kurz, R. C. Mohs, J. C. Morris, P. V. Rabins, K. Ritchie, M. Rossor, L. Thal, and B. Winblad. Current concepts in mild cognitive impairment. *Archives of neurology*, 58(12):1985–1992, 2001.

[15] M. Rubinov and O. Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069, 2010.

[16] B. Schölkopf and A. J. Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.

[17] D. Shahnazian, F. Mokhtari, and G.-A. Hossein-Zadeh. A method based on the granger causality and graph kernels for discriminating resting state from attentional task. In *Biomedical Engineering (ICoBE), 2012 International Conference on*, pages 83–88. IEEE, 2012.

[18] N. Shervashidze, P. Schweitzer, E. J. Van Leeuwen, K. Mehlhorn, and K. M. Borgwardt. Weisfeiler-lehman graph kernels. *The Journal of Machine Learning Research*, 12:2539–2561, 2011.

[19] A. Shrivastava and P. Li. A new mathematical space for social networks. In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 62–71, 2014.

[20] O. Sporns. Contributions and challenges for network models in cognitive neuroscience. *Nature neuroscience*, 17(5):652–660, 2014.

[21] K. Supekar, V. Menon, D. Rubin, M. Musen, and M. D. Greicius. Network analysis of intrinsic functional brain connectivity in alzheimer's disease. *PLoS Comput Biol*, 4(6):e1000100, 2008.

[22] B. M. Tijms, A. M. Wink, W. de Haan, W. M. van der Flier, C. J. Stam, P. Scheltens, and F. Barkhof. Alzheimer's disease: connecting findings from graph theoretical studies of brain networks. *Neurobiology of aging*, 34(8):2023–2036, 2013.

[23] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15(1):273–289, 2002.

[24] K. R. Van Dijk, T. Hedden, A. Venkataraman, K. C. Evans, S. W. Lazar, and R. L. Buckner. Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. *Journal of neurophysiology*, 103(1):297–321, 2010.

[25] S. V. N. Vishwanathan, N. N. Schraudolph, R. Kondor, and K. M. Borgwardt. Graph kernels. *The Journal of Machine Learning Research*, 11:1201–1242, 2010.

[26] C.-Y. Wee, P.-T. Yap, W. Li, K. Denny, J. N. Browndyke, G. G. Potter, K. A. Welsh-Bohmer, L. Wang, and D. Shen. Enriched white matter connectivity networks for accurate identification of mci patients. *Neuroimage*, 54(3):1812–1822, 2011.

[27] C.-Y. Wee, Z. Zhao, P.-T. Yap, G. Wu, F. Shi, T. Price, Y. Du, J. Xu, Y. Zhou, and D. Shen. Disrupted brain functional network in internet addiction disorder: a resting-state functional magnetic resonance imaging study. *PloS one*, 9(9):e107306, 2014.

[28] M. Zanin, P. Sousa, D. Papo, R. Bajo, J. García-Prieto, F. del Pozo, E. Menasalvas, and S. Boccaletti. Optimizing functional network representation of multivariate time series. *Scientific reports*, 2, 2012.

[29] D. Zhang, Y. Wang, L. Zhou, H. Yuan, D. Shen, A. D. N. Initiative, et al. Multimodal classification of alzheimer's disease and mild cognitive impairment. *Neuroimage*, 55(3):856–867, 2011.

[30] Y. Zhang, H. Lin, Z. Yang, and Y. Li. Neighborhood hash graph kernel for protein–protein interaction extraction. *Journal of biomedical informatics*, 44(6):1086–1092, 2011.

[31] D. Zhu, K. Li, L. Guo, X. Jiang, T. Zhang, D. Zhang, H. Chen, F. Deng, C. Faraco, C. Jin, et al. Dicccol: dense individualized and common connectivity-based cortical landmarks. *Cerebral cortex*, 23(4):786–800, 2012.