NeuroImage 125 (2016) 587–600

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

# Matched signal detection on graphs: Theory and application to brain imaging data classification

Chenhui Hu [a,b], Jorge Sepulcre [c], Keith A. Johnson [c], Georges E. Fakhri [a], Yue M. Lu [b], Quanzheng Li [a,*]

[a] Center for Advanced Medical Imaging Sciences, NMMI, Radiology, Massachusetts General Hospital, Boston, MA, USA
[b] School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA
[c] NMMI, Radiology, Massachusetts General Hospital, Boston, MA, USA

ARTICLE INFO

*Article history:*
Received 6 April 2015
Accepted 11 October 2015
Available online 19 October 2015

*Keywords:*
Matched signal detection
Graph-structured data
Spectral graph theory
Brain networks
Classification
Alzheimer's disease
PIB-PET
R-fMRI

ABSTRACT

Motivated by recent progress in signal processing on graphs, we have developed a matched signal detection (MSD) theory for signals with intrinsic structures described by weighted graphs. First, we regard graph Laplacian eigenvalues as frequencies of graph-signals and assume that the signal is in a subspace spanned by the first few graph Laplacian eigenvectors associated with lower eigenvalues. The conventional matched subspace detector can be applied to this case. Furthermore, we study signals that may not merely live in a subspace. Concretely, we consider signals with bounded variation on graphs and more general signals that are randomly drawn from a prior distribution. For bounded variation signals, the test is a weighted energy detector. For the random signals, the test statistic is the difference of signal variations on associated graphs, if a degenerate Gaussian distribution specified by the graph Laplacian is adopted. We evaluate the effectiveness of the MSD on graphs both with simulated and real data sets. Specifically, we apply MSD to the brain imaging data classification problem of Alzheimer's disease (AD) based on two independent data sets: 1) positron emission tomography data with Pittsburgh compound-B tracer of 30 AD and 40 normal control (NC) subjects, and 2) resting-state functional magnetic resonance imaging (R-fMRI) data of 30 early mild cognitive impairment and 20 NC subjects. Our results demonstrate that the MSD approach is able to outperform the traditional methods and help detect AD at an early stage, probably due to the success of exploiting the manifold structure of the data.



## Introduction

Matched subspace detection is a classic tool that determines whether a multidimensional signal lies in a given linear subspace or not (Scharf and Friedlander, 1994). It has achieved great success in applications such as radar, hyperspectral imaging (Manolakis and Shaw, 2002) and medical imaging (Li et al., 2009). The subspace is either known from the physical system that generates the signal, or can be inferred from training data. Subspace learning is a natural way of data dimension reduction and can be achieved by principal component analysis (PCA), which projects the original data to a linear subspace spanned by the leading eigenvectors of the covariance matrix (Jolliffe, 2005). While a common assumption of PCA is that the data come from a linear subspace, many real data are lying in or close to a nonlinear *manifold*, which is a topological space that resembles Euclidean space around each point (Belkin and Niyogi, 2003). Examples of the latter case include brain images (Liu et al., 2013), genetic data (Lee et al., 2008), social network records, and sensor network measurements. In this setting, the low-dimensional subspace that best preserves the intrinsic geometry of the data can be effectively learned by graph spectral methods, e.g., isomap, locality linear embedding (LLE), Laplacian eigenmaps (Belkin and Niyogi, 2003; Roweis and Saul, 2000; Saul et al., 2006; Tenenbaum et al., 2000).

In neuroimaging, as more and more nonlinear data are collected by multiple imaging modalities, there is a need for classifying data with complex intrinsic structures. For instance, the analysis and classification of positron emission tomography (PET) images or functional magnetic resonance imaging (fMRI) data may facilitate the prediction and early detection of Alzheimer's disease (AD). Concurrently, an emerging area of signal processing on graphs is developed for handling these challenging data through the combination of algebraic and spectral graph theoretic concepts with computational harmonic analysis (Shuman et al., 2013). Signals are assumed to reside on vertices of weighted graphs which are often naturally defined by the application. The weight associated with a certain edge in the graph represents the similarity between the two vertices joined by the edge. We refer to graph supported data as *graph-signals*, to differentiate them from conventional signals in Euclidean spaces. In the brain imaging classification, we could view the PET/fMRI data as graph-signals on weighted graphs describing the affinity between each pair of brain regions.

Motivated by the above data classification requirement, we are interested in developing a detection framework for graph-signals.

* Corresponding author.
*E-mail address:* Li.Quanzheng@mgh.harvard.edu (Q. Li).

http://dx.doi.org/10.1016/j.neuroimage.2015.10.026
1053-8119/

Specifically, we formulate several hypotheses to decide which graph structure is more likely to match a given signal. Moreover, we exploit the matched subspace detection technique and propose different types of graph-signal models to make our framework generic to deal with a variety of real situations. The subspace for graph-signal is formed by eigenvectors of the *Laplacian matrix* $\boldsymbol{L}$ of the graph. The graph Laplacian matrix encodes the structure of the graph concisely. From a graph signal processing point-of-view, eigenvectors of $\boldsymbol{L}$ could be treated as the generalization of the basis of conventional Fourier transform (Agaskar and Lu, 2013; Sandryhaila and Moura, 2013; Shuman et al., 2013). Based on spectral graph theory, we can define the variation of graph-signals. It follows that the variation of an eigenvector of $\boldsymbol{L}$ is equal to the associated eigenvalue (Hu et al., 2015a). When we decompose a signal containing instantaneous fMRI measurements into linear combination of eigenvectors of Laplacian matrix associated with the brain-region-affinity graph, we might deem that the components in those eigenvectors with larger eigenvalues as being noisier if the true fMRI signal is assumed to be bandlimited on the graph (Gadde et al., 2013; Kim et al., 2013; Meyer and Shen, 2014).

Our first hypothesis test model simply assumes that the signal lies in a subspace spanned by the first few Laplacian eigenvectors corresponding to smaller eigenvalues. The traditional matched subspace detection could be applied directly to this case. Furthermore, we consider two categories of graph-signal models: *deterministic signals* with constraints and *probabilistic signals* with prior distributions. For deterministic signals, we impose a bounded variation on the signal with respect to the graph. The penalized maximum likelihood estimator (MLE) of the true signal is derived by solving a constrained optimization problem. We find that the test is a weighted energy detector. For probabilistic signals, when we choose a certain degenerate Gaussian distribution as the prior of the projection coefficients of the signal onto the graph Laplacian eigenvectors, the decision ends up comparing the signal variations on the two hypothetic graphs in a noise-free case.

We evaluate the effectiveness of the matched signal detection (MSD) theory on both synthetic and real data sets. Simulations based on randomly generated graphs demonstrate the feasibility of our approaches even if we do not know the exact probability distributions of the testing signals. Then, we apply the proposed detection algorithms to brain imaging classification tasks of AD. As one of the most prevalent forms of dementia, AD is believed to be a brain network associated disease (Gomez-Ramirez and Wu, 2014; Raj et al., 2012; Sepulcre et al., 2013), and is characterized by progressive impairment of memory and other cognitive capacity. It affects nearly 36 million people worldwide with an expected number of cases to be 65.7 million by 2030 (Brookmeyer et al., 2007). The development of neuroimaging classification techniques may enable us to monitor the functional and anatomical changes of the brain in vivo and discover reliable biomarkers for identifying AD at an early stage. In this study, we have compared a novel MSD approach with other widely used methods including principle component analysis (PCA), support vector machine (SVM) and linear discriminant analysis (LDA) on two data sets: one is PET imaging of brain amyloid using Pittsburgh compound-B (PIB) tracer of AD and normal control (NC) subjects; the other contains resting-state fMRI (R-fMRI) images of early mild cognitive impairment (EMCI) and NC subjects in the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. For the MSD, we compute the similarity between each of two brain regions with the Gaussian radial basis function (RBF) kernel. This simple way of building brain networks avoids estimating network structures by solving inverse problems, which often requires more data; yet the weighted graphs associated with the networks approximate the data manifolds. Experimental results show that when using the MSD on graphs, we can achieve significantly better classification performance than the compared algorithms. The results indicate that our method provides an effective way for brain imaging classification, probably due to the capability of exploiting the manifold structure of the data.

Our contributions in this paper are three-fold: first, we have developed a matched signal detection theory for graph-signals which are ubiquitous in medical imaging applications; second, we keep the framework generic and simple by proposing a variety of signal models and using simple similarity metrics to construct graphs; third, we demonstrate that the detection theory is particularly suitable for neuroimaging classifications.

## Theory

To formulate the framework of matched signal detection on graphs, we first introduce the concept of graph-signals. We extend the traditional Fourier transform to a graph Fourier transform and define a notion of graph-signal frequency based upon spectral graph theory. Then, to model different real data, we propose three classes of signal models on graphs. Finally, we derive the signal detection criterion under each signal model.

### *Weighted graphs and graph-signals*

Many contemporary applications such as social, power grid, sensor, and brain networks involve high-dimensional data with natural structures defined by weighted graphs. To efficiently process such signals on graphs, an emerging field of signal processing on graphs integrates the graph spectral theory with computational harmonic analysis. Here we present basic concepts of signal processing on graphs in the context of neuroimaging data analysis.

A brain network can be represented by a *weighted graph* $\mathcal{G}(\mathcal{V}, \mathcal{E}, \boldsymbol{W})$ containing a vertex set $\mathcal{V}(|\mathcal{V}| = N)$, an edge set $\mathcal{E}$ and a weighted adjacency matrix $\boldsymbol{W}$. The vertices typically indicate a group of predefined brain regions or a set of image voxels (Stanley et al., 2013; Zalesky et al., 2010). If there is an edge between vertices $i$ and $j$, then $W_{ij}$ denotes the weight of the edge; otherwise, $W_{ij} = 0$. We assume the similarity metric is symmetric and non-negative, namely $W_{ij} = W_{ji} \geq 0$ for all $i$ and $j$. Meanwhile, it is reasonable to assume that no brain region is isolated. Therefore, $\mathcal{G}$ should be undirected and connected. Physiologically, $W_{ij}$ may quantify the similarity of two brain regions in terms of their biochemical measurements (such as the amyloid deposition revealed by PIB-PET) or anatomical properties (such as the number of fiber pathways connecting those regions). The exact formula of the weights could be chosen flexibly based on different applications.

In addition to the adjacency matrix, we introduce the *graph Laplacian* as another important graph associated matrix. We denote by $\boldsymbol{D}$ the degree matrix which is diagonal with $D_{ii} = \sum_{j=1}^{N} W_{ij}$. Then, the graph Laplacian is defined as $\boldsymbol{L} \stackrel{\text{def}}{=} \boldsymbol{D} - \boldsymbol{W}$. Because $\boldsymbol{L}$ is a real symmetric matrix, it has a complete set of orthonormal eigenvectors $\{\boldsymbol{f}_i\}_{i=1,\ldots,N}$. If $\mathcal{G}$ is connected, the associated eigenvalues $\{\lambda_i\}_{i=1,\ldots,N}$ are real non-negative with the unique smallest eigenvalue being zero (Chung, 1997). We assume that the eigenvalues are increasingly sorted as $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_N$. By eigendecomposition, we can decompose the graph Laplacian into $\boldsymbol{L} = \boldsymbol{F}\boldsymbol{\Lambda}\boldsymbol{F}^T$, where $\boldsymbol{\Lambda}$ is diagonal with $\Lambda_{ii} = \lambda_i$ being the $i$th smallest eigenvalue of $\boldsymbol{L}$ and the $i$th column of $\boldsymbol{F}, \boldsymbol{f}_i$, is the associated eigenvector.

A signal $x$ defined on the vertices of graph $\mathcal{G}$ is a function from $\mathcal{V}$ to $\mathbb{R}$. This *graph-signal* could be expressed as a vector in $\mathbb{R}^N$ with the $i$th element of the vector being a real value assigned to the $i$th vertex. We will also denote this vector by $\boldsymbol{x}$ and use $x(i)$ to indicate both the function value at the $i$th vertex and the $i$th element of the vector. Examples of the graph-signal are in Fig. 1. In practice, we could view PET scans or fMRI time series as graph-signals defined on vertices of the brain connectivity network that is constructed by connecting edges between different brain regions or image voxels. For a graph-signal $\boldsymbol{x}$, the graph Laplacian behaves as a difference operator on it:

$$(\boldsymbol{L}\boldsymbol{x})(i) = \sum_{j \sim i} W_{ij}[x(i) - x(j)], \tag{1}$$

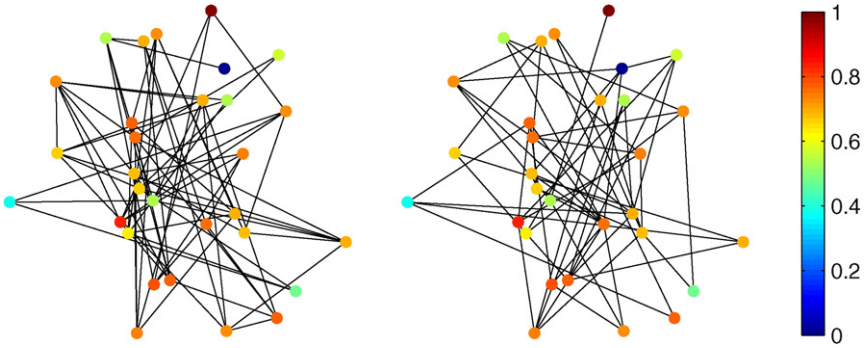


**Fig. 1.** A same graph-signal displayed on two different binary graphs that both have 30 vertices but contain different edges. The color of the vertex denotes the signal value at the corresponding location.

where $j \sim i$ means that vertex $j$ is connected to vertex $i$. Due to the analogy between graph Laplacian and continuous Laplace operator, we are able to extend the classical Fourier transform to a more general form for graph-signals. We elaborate this notion in the next subsection.

*Frequency analysis of graph-signals*

As an efficient way of processing traditional signals, the classical Fourier transform expands a continuous function $x$ of time in terms of the complex exponentials (Oppenheim and Willsky, 1997), namely

$$\tilde{x}(\xi) \overset{\text{def}}{=} \left\langle x, e^{2\pi i\xi t} \right\rangle = \int_{\mathbb{R}} x(t) e^{-2\pi i\xi t} dt, \tag{2}$$

where $\langle \cdot, \cdot \rangle$ denotes inner product. We adopt $\tilde{x}$ to represent the Fourier transform of $x$.

The Fourier basis functions are eigenfunctions of the one-dimensional Laplace operator $\Delta$:

$$-\Delta\left(e^{2\pi i\xi t}\right) = -\frac{\partial^2}{\partial t^2} e^{2\pi i\xi t} = (2\pi\xi)^2 e^{2\pi i\xi t}. \tag{3}$$

Similarly, for a signal $\boldsymbol{x}$ on the vertices of $\mathcal{G}$, we can define its *graph Fourier transform* (GFT) $\tilde{x}$ as the expansion of $\boldsymbol{x}$ in terms of the eigenvectors of the graph Laplacian (Shuman et al., 2013):

$$\tilde{x}(\lambda_l) \overset{\text{def}}{=} \langle \boldsymbol{x}, \boldsymbol{f}_l \rangle = \sum_{i=1}^{N} x(i) f_l^*(i). \tag{4}$$

Accordingly, the *inverse graph Fourier transform* is given by

$$x(i) = \sum_{l=1}^{N} \tilde{x}(\lambda_l) f_l(i). \tag{5}$$

More concisely, the GFT of $\boldsymbol{x}$ could be written as the projection of $\boldsymbol{x}$ onto the eigenvectors of the graph Laplacian in a matrix form

$$\tilde{\boldsymbol{x}} = \boldsymbol{F}^T \boldsymbol{x}, \tag{6}$$

where elements of $\tilde{\boldsymbol{x}}$ and $\boldsymbol{x}$ are indexed by graph Laplacian eigenvalues and vertices, respectively. When the context is clear, we could also write $\tilde{x}(\lambda_i)$ and $x(i)$ as $\tilde{x}_i$ and $x_i$ accordingly. Since $\boldsymbol{F}\boldsymbol{F}^T = \boldsymbol{I}$, the inverse GFT is $\boldsymbol{x} = \boldsymbol{F}\tilde{\boldsymbol{x}}$. The GFT is a generalization of the traditional Fourier transform in the context of graph analysis. For example, when $\mathcal{G}$ is a cycle graph, the eigenvectors of its Laplacian form an exact discrete Fourier transform matrix (Agaskar and Lu, 2013), and thus discrete Fourier transform could be considered as a special case of GFT. Unlike the classical Fourier transform, the GFT depends on the specific graph structure. In addition, because it can adapt to different data with data-dependent graph structures, it yields better results in image compression (Hu et al., 2015b), denoising (Meyer and Shen, 2014) and deblurring (Kheradmand and Milanfar, 2014) as compared to traditional methods. Essentially, we can also consider the GFT as a graph-based orthonormal transform based on Eq. (6). However, to be consistent with the literature of signal processing on graphs, we refer to this orthonormal transform as the GFT.

In classical Fourier analysis, the eigenvalues $\{(2\pi\xi)^2\}_{\xi \in \mathbb{R}}$ in Eq. (3) are also known as frequencies, since when $\xi$ increases the associated complex exponential eigenfunction will oscillate more rapidly. In graph-signal analysis, the eigenvalues and eigenvectors of graph Laplacian provide a similar concept of frequency. The graph Laplacian eigenvectors associated with low frequencies $\lambda_i$ vary slowly across the graph, meaning that if two vertices are linked by an edge with a large weight, the values of the eigenvector at those positions tend to be close. Particularly, for connected graphs, the Laplacian eigenvector $\boldsymbol{f}_1$ associated with $\lambda_1 = 0$ is constant and equal to $N^{-\frac{1}{2}}$ at each vertex. In contrast, the eigenvectors associated with higher eigenvalues change more rapidly and are more likely to have dissimilar values on vertices connected by an edge with a big weight.

To quantify the smoothness of a graph-signal $\boldsymbol{x}$ with respect to the intrinsic structure of the graph, we propose the *graph Laplacian quadratic form*:

$$V_{\mathcal{G}}(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x} = \frac{1}{2} \sum_{i,j} W_{ij} (x(i) - x(j))^2. \tag{7}$$

This quantity is small when $\boldsymbol{x}$ has similar values at neighboring vertices connected by an edge with a large weight, meaning that $\boldsymbol{x}$ is smooth. By the Courant–Fischer Theorem (Horn and Johnson, 1990), we could define the eigenvalues of the Laplacian recursively via the Rayleigh quotient as

$$\lambda_1 = \min_{\|\boldsymbol{x}\|_2 = 1} \{\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}\}, \tag{8}$$

$$\lambda_l = \min_{\substack{\|\boldsymbol{x}\|_2 = 1, \\ \boldsymbol{x} \perp span\{\boldsymbol{f}_1, \cdots, \boldsymbol{f}_{l-1}\}}} \{\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}\}, \; l = 2, 3, \cdots, N, \tag{9}$$

where the minimizer of the $l$th problem is the eigenvector $\boldsymbol{f}_l$. Eqs. (8) and (9) interpret why the graph Laplacian eigenvectors associated with lower eigenvalues are smoother, and give another reason why we can treat the graph Laplacian spectrum as graph-signal frequencies.

Due to the above nature of GFT, we can decompose brain imaging data into linear combinations of many templates which correspond to graph Laplacian eigenvectors of the brain network. After that, image denoising may be performed by suppressing higher frequency components in the image (Gadde et al., 2013; Meyer and Shen, 2014). We may also compare two groups of subjects more robustly by extracting features from the GFT coefficients, similar to the multi-resolution brain network analysis via graph wavelet in (Kim et al., 2013).

*Graph-signal models*

For classifying graph-structured data, it is necessary to introduce statistical signal models on graphs before carrying out hypothesis tests. According to the property of GFT coefficients of real data, we consider three graph-signal models, i.e., bandlimited graph-signals, constrained graph-signals, and probabilistic graph-signals to facilitate hypothesis tests which determine whether a signal is embedded in a given graph structure.

*Bandlimited graph-signals*

In many applications, the GFT coefficient $\tilde{x}_i$ of the signal $\boldsymbol{x}$ is more likely to be close to zero as the associated Laplacian eigenvalue $\lambda_i$ grows, due to the smoothness of $\boldsymbol{x}$ on the graph. We explain the mechanism of signal smoothness for brain imaging data through a network diffusion model in the Discussion section.

Analogous to the DFT frequencies of traditional signals, we call the set $\{\boldsymbol{f}_i\}_{i=1,\cdots,N}$ frequency components of the GFT. We can define bandlimited graph-signals as signals that are only supported on lower frequency components corresponding to the first few smaller Laplacian eigenvalues, namely $\boldsymbol{x} \in \mathbb{R}^N$ with its GFT $\tilde{\boldsymbol{x}}$ satisfying

$$\tilde{x}_i = 0, \text{ forall } i > K, \tag{10}$$

where $\tilde{x}_i \overset{\text{def}}{=} \tilde{x}(\lambda_i)$ is the GFT coefficient associated with the graph Laplacian eigenvalue $\lambda_i$ and $K \in \{1, 2, \cdots, N\}$.

*Constrained graph-signals*

To improve the performance of a hypothesis test, prior information is often applied. Here we incorporate prior knowledge about the signal $\boldsymbol{x}$ via a constraint on its GFT coefficients, namely $C(\tilde{\boldsymbol{x}}) \le r$ for a given $r > 0$. We specify the constraint function as the following normalized quadratic form

$$C(\tilde{x}) = \frac{\sum_{i=1}^{N} \alpha_i \tilde{x}_i^2}{\sum_{i=1}^{N} \tilde{x}_i^2}, \tag{11}$$

where $\{\alpha_i\}_{i=1,\cdots,N}$ is a set of non-negative penalty weights. Since we believe that GFT coefficients associated with higher eigenvalues are less informative, we could impose larger penalty weights on those GFT coefficients.

There is a special case of Eq. (11) when $\alpha_i = \lambda_i$ for $i = 1, \cdots, N$. In this case, because $\boldsymbol{x}^T \mathbf{L} \mathbf{x} = \tilde{\boldsymbol{x}}^T \boldsymbol{\Lambda} \tilde{\boldsymbol{x}} = \sum_{i=1}^{N} \lambda_i \tilde{x}_i^2$, the constraint function can be expressed as

$$C(\tilde{\boldsymbol{x}}) = \frac{\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}, \tag{12}$$

which is a normalized version of the smoothness measure (Eq. (7)). Under this circumstance, the constrained graph-signals are essentially signals with bounded variations on the graph. It is easy to see that this constraint function is bounded below and above by zero and the maximum graph Laplacian eigenvalue, respectively. If $C(\tilde{\boldsymbol{x}}) \le r$, we can rewrite the requirement to $\sum_{i=1}^{N} \lambda_i \tilde{x}_i^2 \le r \sum_{i=1}^{N} \tilde{x}_i^2$, i.e., $\sum_{i=1}^{N} (\lambda_i - r) \tilde{x}_i^2 \le 0$. Typically, $r$ is less than the maximum eigenvalue $\lambda_N$. We may further reshape the condition to $\sum_{\lambda_i > r} (\lambda_i - r) \tilde{x}_i^2 \le \sum_{\lambda_i \le r} (\lambda_i - r) \tilde{x}_i^2$. Hence, to satisfy the constraint of smoothness on graph, we should ensure that the signal has a higher energy concentration (the energy is measured by the square of GFT coefficients) on the lower frequency components compared with the higher frequency components.

*Probabilistic graph-signals*

In the above two signal models, we treat the signal as deterministic but unknown due to noise. From a Bayesian perspective, we can also assume that the signal is randomly generated from a prior distribution. We express the distribution in the transformed domain as the following general form

$$\mathbb{P}(\tilde{\boldsymbol{x}}) \propto \exp\left(-\sum_{i=1}^{N} \alpha_i \tilde{x}_i^2\right), \tag{13}$$

with $\alpha_i > 0$ denoting the penalty on the energy of the $i$th frequency component. Specially, when $\alpha_i = \lambda_i$ for $i = 1, \cdots, N$, the distribution can be rewritten as

$$\mathbb{P}(\tilde{\boldsymbol{x}}) \propto \exp\left(-\tilde{\boldsymbol{x}}^T \boldsymbol{\Lambda} \tilde{\boldsymbol{x}}\right), \tag{14}$$

which is equivalent to $\mathbb{P}(\boldsymbol{x}) \propto \exp(-\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x})$ and can be viewed as a degenerate Gaussian distribution. Noticing the identity (Eq. (7)), we can see that Eq. (14) is a reasonable signal model for brain imaging data, since functionally correlated brain regions should yield similar measurements. We will construct MSD on graphs in the next section based on the above graph-signal models.

*Decision models*

Let $\boldsymbol{y} \in \mathbb{R}^N$ be an observed signal defined on a potential graph. There are two hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, representing that $\boldsymbol{y}$ is embedded in either graph $\mathcal{G}_0$ or $\mathcal{G}_1$, with the associated Laplacian matrix $\boldsymbol{L}_0$ or $\boldsymbol{L}_1$, respectively. The graphs are defined on the same set of vertices but containing different edges. Fig. 1 gives an example of a same graph-signal on two different binary graphs. Our goal is to formulate hypothesis tests to decide which graph fits the signal more accurately. We refer to the procedure as *matched signal detection* (MSD) on graphs.

To model real challenges, we consider $\boldsymbol{y}$ as a contaminated version of $\boldsymbol{x}$ with additive Gaussian noise. We assume $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n}$ with $\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$. Supposing that the eigendecomposition of $\boldsymbol{L}_j$ possesses the form $\boldsymbol{L}_j = \boldsymbol{F}_j \boldsymbol{\Lambda}_j \boldsymbol{F}_j^T$ under $\mathcal{H}_j$ for $j = 0, 1$, from Eq. (6) we can write the hypothesis test as

$$\mathcal{H}_0 : \boldsymbol{y} = \boldsymbol{F}_0 \tilde{\boldsymbol{x}}_{\mathcal{G}_0} + \boldsymbol{n} \text{ vs. } \mathcal{H}_1 : \boldsymbol{y} = \boldsymbol{F}_1 \tilde{\boldsymbol{x}}_{\mathcal{G}_1} + \boldsymbol{n}, \tag{15}$$

where $\tilde{\boldsymbol{x}}_{\mathcal{G}_0}$ and $\tilde{\boldsymbol{x}}_{\mathcal{G}_1}$ indicate the GFT transform on $\mathcal{G}_0$ and $\mathcal{G}_1$ respectively. Given the observation $\boldsymbol{y}$, the general form of the likelihood ratio test (LRT) is

$$\frac{\ell(\Theta_1; \boldsymbol{y})}{\ell(\Theta_0; \boldsymbol{y})} = \frac{\mathbb{P}(\boldsymbol{y}|\mathcal{H}_1)}{\mathbb{P}(\boldsymbol{y}|\mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \frac{\pi_0}{\pi_1} = \eta, \tag{16}$$

where $\ell(\cdot)$ indicates the likelihood function, $\Theta_j$ and $\pi_j$ are the set of parameters and prior probability under $\mathcal{H}_j$ for $j = 0,1$, respectively. In accordance with the signal models presented in the last section, we present more concrete versions of (16) in terms of different types of MSD on graphs.

*Simple-MSD on graphs*

For bandlimited graph-signals, the MSD on graphs follows immediately from the conventional matched subspace detection (Scharf and Friedlander, 1994), where the subspace is spanned by a subset of the eigenvectors of the Laplacian matrix under each hypothesis. For convenience, we refer to this test as SMSD. Denote by $\tilde{y}_{i,0}$ and $\tilde{y}_{i,1}$ as the $i$th GFT coefficient of $\boldsymbol{y}$ on graph $\mathcal{G}_0$ and $\mathcal{G}_1$ respectively. If we assume that the noise variance is unknown but could be estimated by its MLE, we have the test statistic

$$T_{\text{SMSD}} = \sum_{i=1}^{K} \left(\tilde{y}_{i,0}^2 - \tilde{y}_{i,1}^2\right), \tag{17}$$

where $K$ indicates the maximum frequency of the signal. For the sake of completeness, we include the proof of Eq. (17) in Appendix A. In other words, to decide which graph the signal belongs to, one simply

compares the projection energy of the observation on the first few graph Laplacian eigenvectors of each hypothetic graph.

*Constrained-MSD on graphs*

We adopt CMSD as a shorthand for constrained-MSD problem, when the constrained graph-signal model in the last section is considered. By multiplying $\boldsymbol{F}_0^T$ and $\boldsymbol{F}_1^T$ on both sides of each of the two equations in (15) respectively, we obtain

$$\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{x}} + \tilde{\boldsymbol{n}}, \tag{18}$$

where the tilde sign means the GFT on the hypothetic graph. Since $\boldsymbol{F}$ is orthonormal, $\tilde{\boldsymbol{n}}$ is still white Gaussian noise with identical distribution as $\boldsymbol{n}$. For smooth graph-signals satisfying the constraint $C(\tilde{\boldsymbol{x}}) \le r$ with $C(\tilde{\boldsymbol{x}})$ specified by Eq. (11), the test statistic can be written as

$$T_{\mathrm{CMSD}} = \sum_{i=1}^{N} \left[ \frac{\tau_{i,0}^2}{(1+\tau_{i,0})^2} \tilde{y}_{i,0}^2 - \frac{\tau_{i,1}^2}{(1+\tau_{i,1})^2} \tilde{y}_{i,1}^2 \right], \tag{19}$$

where $\tilde{y}_{i,j}$ is the $i$th GFT coefficient of $\boldsymbol{y}$ on graph $\mathcal{G}_j (j = 0, 1)$ and $\tau_j (j = 0, 1)$ is a quantity dependent on the respective GFT of $\boldsymbol{y}$, the noise level, and the parameters of the signal model. Detailed derivations of Eq. (19) are provided in Appendix B. From Eq. (19), we find that the test is a weighted energy detector where $\tau_{i,j}$ embodies the deviation between the $i$th estimated GFT coefficient and the corresponding GFT coefficient of the observation. When there is no noise or the observed signal $\boldsymbol{y}$ satisfies the constraint of the variation on graph $\mathcal{G}_j$, we have $\tau_{i,j} = 0$ for $i = 0, \ldots, N$ meaning that $\boldsymbol{y}$ does conform to the constrained graph-signal model. This is analogous to the case where the observed signal perfectly lies in the subspace of the first $K$ graph Laplacian eigenvectors in the bandlimited graph-signal model.

*Probabilistic-MSD on graphs*

A more general MSD is to consider random graph-signals, which gives rise to the probabilistic-MSD (PMSD) on graph. We assume that $\tilde{\boldsymbol{x}}$ has a probability distribution function (p.d.f.) $h_j(\tilde{\boldsymbol{x}})$ under $\mathcal{H}_j$. Due to the independence of the noise on the signal, the likelihood ratio (LR) is

$$\mathrm{LR} = \frac{\int h_1(\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{n}}) g_1(\tilde{\boldsymbol{n}}) d\tilde{\boldsymbol{n}}}{\int h_0(\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{n}}) g_0(\tilde{\boldsymbol{n}}) d\tilde{\boldsymbol{n}}}, \tag{20}$$

where $g_j(\cdot)$ denotes the p.d.f. of the white Gaussian noise under $\mathcal{H}_j$. If the GFT coefficients of the true signal follow a Gaussian distribution $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_j)$, the test statistic would be $\tilde{\boldsymbol{y}}^T(\Phi_0^{-1} - \Phi_1^{-1})\tilde{\boldsymbol{y}}$, where $\Phi_j = \boldsymbol{\Sigma}_j + \sigma^2 \boldsymbol{I}$ for $j = 0,1$. In particular, if we apply the degenerate Gaussian distribution ((14)) to the true signal, we will have $h_j(\tilde{\boldsymbol{x}}) \propto \exp(-\tilde{\boldsymbol{x}}^T \boldsymbol{\Lambda}_j \tilde{\boldsymbol{x}})$. It turns out that $\tilde{\boldsymbol{y}}$ is distributed as $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Lambda}_j^{\dagger} + \sigma^2 \boldsymbol{I})$, when viewing $g_j$ as a degenerated Gaussian distribution. Note that $\boldsymbol{\Lambda}_j^{\dagger}$ represents the pseudoinverse of $\boldsymbol{\Lambda}_j$. When $\sigma$ is known, the LR test statistic reduces to

$$T_{\mathrm{PMSD}} = \sum_{i=1}^{N} \left( \beta_{i,0} \tilde{y}_{i,0}^2 - \beta_{i,1} \tilde{y}_{i,1}^2 \right), \tag{21}$$

where $\beta_{1,j} = \sigma^{-2}$ and $\beta_{i,j} = (\lambda_{i,j}^{-1} + \sigma^2)^{-1}$ when $i \ge 2$, for $j = 0, 1$. We can also replace $\sigma$ in Eq. (21) with its MLE if $\sigma$ is unknown. In a special noise-free case, $\tilde{\boldsymbol{y}}$ follows a degenerate Gaussian distribution (Eq. (14)) and the test statistic becomes $\boldsymbol{y}^T(\boldsymbol{L}_0 - \boldsymbol{L}_1)\boldsymbol{y}$, which is simply the difference between the signal variations on the two graph structures.

## Results

*Numerical simulations*

We have evaluated the proposed MSD rules on small-world networks, which are simplified yet effective models of brain connectivity structures (Bassett and Bullmore, 2006; He et al., 2007; Stam et al., 2007). A small-world network is characterized by dense local cliques of connections between neighboring vertices and a short path length between any pair of vertices due to the existence of few long-range connections. Brain anatomical and functional networks have a small-world topology that supports both segregated and integrated information processing with a minimized cost (Bassett and Bullmore, 2006). Moreover, certain brain diseases like AD could cause the alternation of this network structure before other notable symptoms emerge (Sanz-Arigita et al., 2010; Supekar et al., 2008). Sensitive biomarkers derived from brain connectivity networks have been used for detecting AD at an early stage successfully (Balthazar et al., 2014; Wee et al., 2012). For a comprehensive review on this direction, refer to (Gomez-Ramirez and Wu, 2014). Our MSD schemes jointly model the alternation of network topology and the distribution of the neuroimaging data in order to achieve better classification performance. We classified data on two small-world networks specified by the Watts–Strogatz model (Watts and Strogatz, 1998), using the proposed schemes in the last section.

As shown in Fig. 2, the networks were constructed from a random rewiring of a ring graph with 40 vertices. First, we created a ring graph in which each vertex was connected to the same number of nearest neighbors on each side. Then, we removed each edge with a uniform independent probability and rewired it to build an edge between a pair of vertices that were chosen uniformly at random. In our tests, we considered two scenarios in terms of the network structure: in the first case, the number of edges associated with each vertex namely the vertex degree $d$ and the rewiring probability $p_r$ were (12, 0.1) under $\mathcal{H}_0$, and (20, 0.4) under $\mathcal{H}_1$, respectively; in the second case, the vertex degree and the rewiring probability were (12, 0.1) under both $\mathcal{H}_0$ and $\mathcal{H}_1$. Meanwhile, we assumed the GFT coefficient distributions of the signals followed either an exponential decay or a step function. In the former case, the GFT coefficients were distributed as $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ with $\sum$ being a diagonal matrix and $\sum_{ii} = \exp(-i/5)$ under both hypotheses; in the latter case, $\sum_{ii} = 1$ for $i = 1, \ldots, 12$ and $\sum_{ii} = 0.1$ for $i = 13, \ldots, 40$. To mimic real challenges, Gaussian noise with different standard deviations $\sigma = 0.3$ or 0.5 was added to the original signals.

Under the above different settings, we compared our methods with traditional classifiers. For the SMSD, we projected the noisy signal on the first 12 graph Laplacian eigenvectors; for the CMSD, we imposed piecewise-constant penalty weights on the GFT amplitudes; for the PMSD, we assumed that the distribution of the signal was known. To account for the randomness of the networks and signals, we ran each experiment 100 times for each algorithm and plotted the mean receiver operating characteristic (ROC) curve in Fig. 3. The average and standard deviation of the area under the curve (AUC) for each algorithm in different tests are given in Table 1. From the figure and the table, we observe that the MSD schemes consistently outperform the compared algorithms namely PCA, SVM, LDA, as well as kernel PCA and kernel SVM, which are denoted by KPCA and KSVM respectively. In the KPCA and KSVM, Gaussian RBF kernel was adopted. We find that the result of the KPCA or KSVM is better than that of the PCA or SVM accordingly, although the variance of the AUC of the KPCA or KSVM is larger than that of the PCA or SVM. The reason that the AUC of every classic approach is significantly less than that of any MSD scheme might be its inefficiency of exploiting the structure of the signals particularly with noisy data. In Figs. 3(a) and (b), we see that when the GFT coefficients of the true signals follow either an exponential decay or a step function, our methods achieve much better performance than the classic classifiers. Furthermore, from Figs. 3(a) and (c), we find that although large noise in the data leads to some decline of the performance of the MSD schemes,

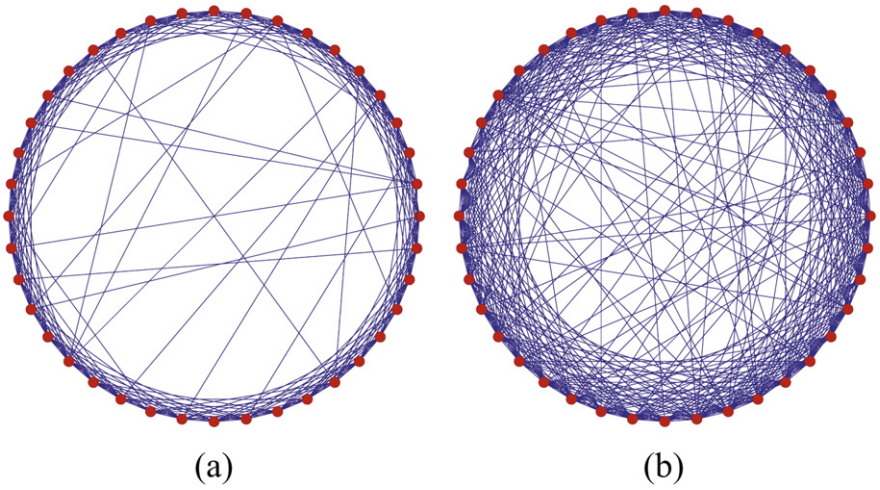


**Fig. 2.** Two random realizations of the small-world networks used in the numerical simulations. Both networks have 40 vertices (red circles) and the weight of every edge (blue line) is one. In (a), the degree of each vertex is $d = 12$ and the rewiring probability of the edges is $p_r = 0.1$; In (b), the vertex degree and rewiring probability are $d = 20$ and $p_r = 0.4$, respectively.

the proposed schemes still outperform the compared methods when the distributions of the network structure and signal GFT coefficients are fixed. From Figs. 3(a) and (d), we observe that when the hypothetic networks have the same distribution, the AUC of each MSD scheme becomes lower than that of the corresponding scheme when the networks have different distributions. In the above simulations, we optimally tuned the parameters of the MSD and the compared algorithms by using grid search. For the SMSD, we chose the maximum frequency

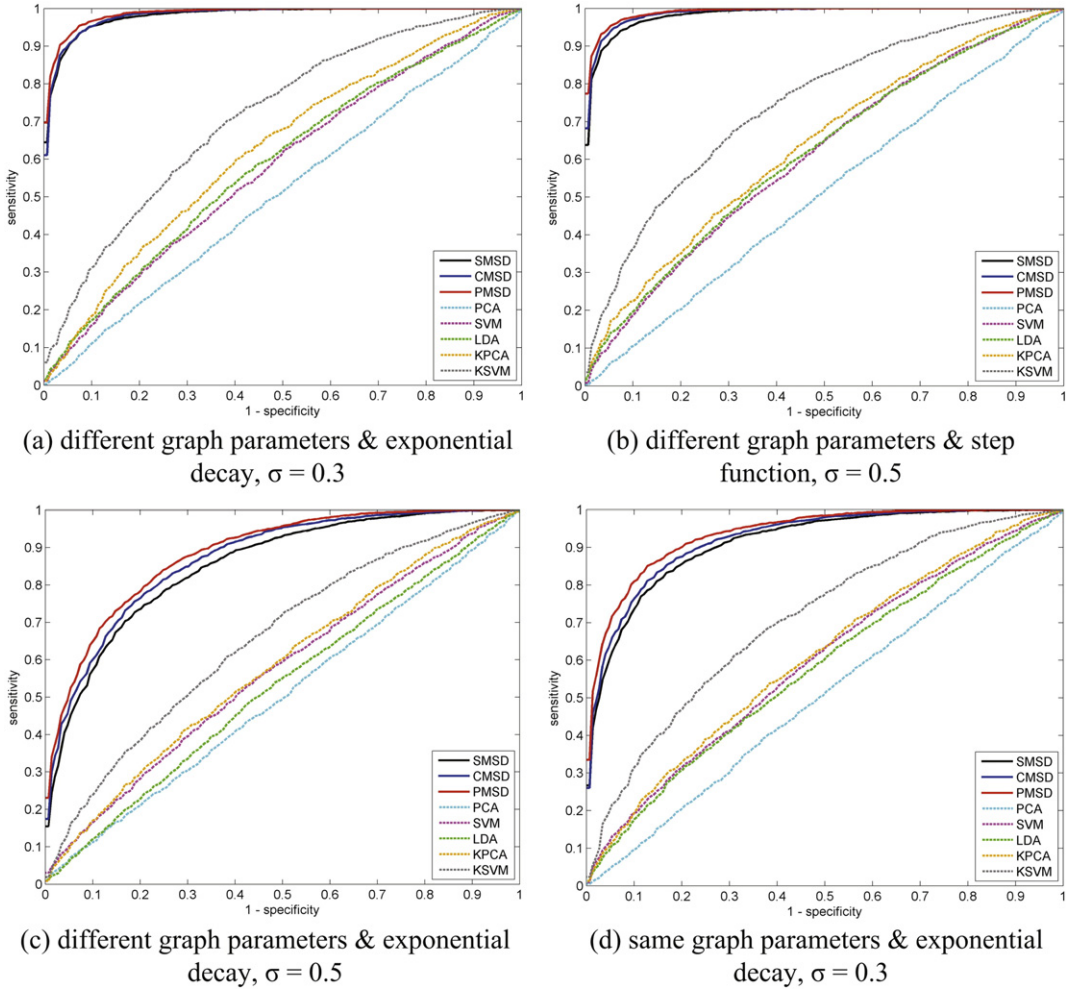


**Fig. 3.** ROC curves for the proposed MSD approaches and other classic classification methods based upon graph-signals on two randomly generated small-world networks composed of 40 vertices. The vertex degree and rewiring probability in (a–c) are (12,0.1) for $\mathcal{G}_0$ and (20, 0.4) for $\mathcal{G}_1$, respectively; the vertex degree and rewiring probability in (d) are (12,0.1) for both $\mathcal{G}_0$ and $\mathcal{G}_1$. Signals in (a) & (c–d) have an exponential decay of GFT coefficients on the corresponding graph, while the GFT coefficients of the signals in (b) follow a step function. KPCA and KSVM denote the kernel PCA and kernel SVM correspondingly; $\sigma$ indicates the standard deviation of the Gaussian noise.

**Table 1**
Area under the curve (AUC) for each classification method in the simulations. Column (a–d) correspond to the results of repetitive running of the experiments with the graph-structural parameters and signal models specified in Figs. 3(a–d), respectively. Mean and standard deviation (showing in parentheses) of the AUC obtained from 100 repetitions of each algorithm are listed.

| | (a) | (b) | (c) | (d) |
|---|---|---|---|---|
| SMSD | 0.9799 (0.0059) | 0.9826 (0.0034) | 0.8500 (0.0169) | 0.9109 (0.0220) |
| CMSD | 0.9814 (0.0057) | 0.9869 (0.0036) | 0.8683 (0.0177) | 0.9209 (0.0180) |
| PMSD | 0.9859 (0.0045) | 0.9901 (0.0027) | 0.8831 (0.0144) | 0.9348 (0.0171) |
| PCA | 0.5091 (0.0425) | 0.5083 (0.0506) | 0.5027 (0.0363) | 0.5063 (0.0233) |
| SVM | 0.5798 (0.0588) | 0.6088 (0.0771) | 0.5705 (0.0909) | 0.5953 (0.0741) |
| LDA | 0.5887 (0.0524) | 0.6132 (0.0347) | 0.5287 (0.0372) | 0.5779 (0.0428) |
| KPCA | 0.6266 (0.0605) | 0.6358 (0.0683) | 0.5821 (0.0492) | 0.6078 (0.0348) |
| KSVM | 0.7146 (0.0663) | 0.7446 (0.0830) | 0.6556 (0.0941) | 0.7095 (0.0797) |

of the bandlimited graph-signal $K$ as 12. For the CMSD, the upper bound $r$ of the constraint function was equal to the 12th eigenvalue of the respective graph Laplacian. In Figs. 3(a) and (c–d), we chose the penalty weights $\{\alpha_i\}_{i=1,\cdots,40}$ in Eq. (11) as $\alpha_i = 1$ for $i = 1, \ldots, 10$; $\alpha_i = 100$ for $i = 11, \ldots, 20$; $\alpha_i = 400$ for $i = 21, \ldots, 30$; and $\alpha_i = 1000$ for $i = 31, \ldots, 40$. In Fig. 3(b), we selected $\alpha_i = 1$ for $i = 1, \ldots, 12$; and $\alpha_i = 100$ for $i = 13, \ldots, 40$. We note that the performance of the CMSD scheme is superior to that of the SMSD and inferior to that of the PMSD in the above four experimental settings. (See Fig. 3.)

*PIB-PET and R-fMRI study*

We then applied our MSD theory to the classification of subjects in two brain imaging data sets: a PIB-PET data set from the neuroimaging studies in aging at the Massachusetts General and Brigham and Women's Hospitals (Sepulcre et al., 2013) and a R-fMRI data set from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu).

As shown in Table 2, the PIB-PET data set consists of 30 AD patients and 40 NC subjects. The detailed criteria for AD and NC subjects as well as the experimental procedures were described in Sepulcre et al. (2013). The PET scans were acquired on a Siemens/CTI ECAT HR$^+$ PET scanner at Massachusetts General Hospital as follows. Amyloid imaging was performed with N-methyl-$^{11}$C-2-(4-methylaminophenyl)-6-hydroxybenzothiazole (PIB). $^{11}$C-PIB was prepared as specified by Mathis et al. (2002). Briefly, after a transmission scan, 10–15 mCi $^{11}$C-PIB was injected as a bolus intravenously, which was followed immediately by a 60-min dynamic PET scan in 3D mode (63 image planes, 15.2 cm axial field of view, 5.6 mm transaxial resolution and 2.4 mm slice interval; 69 frames: 12 × 15 s, 57 × 60 s). PIB-PET data were reconstructed with ordered set expectation maximization, corrected for attenuation, and each frame was evaluated to verify adequate count statistics and absence of head motion. Individual $^{11}$C-PIB-PET scans were spatially normalized using the warping information derived from individual stereotactic normalization of the structural MRI using SPM 2.

PET neuroimaging data were downsampled from the normalized 2 mm isotropic voxels to 8 mm isotropic voxels for computational efficiency. Each image has a dimension of 20 × 24 × 18 with 8 mm × 8 mm × 8 mm voxels. In the data preprocessing step, we first masked out the area outside of the brain. Next, we applied Automated Anatomical Labeling (AAL) (Tzourio-Mazoyera et al., 2002) to map the effective voxels to 90 VOIs. The data were then averaged within each VOI for further analysis. Among all the VOIs, we picked out 42 regions that were considered to be potentially related to AD (Azari et al., 1992). Table 1 of Huang et al. (2009) listed the names of the selected VOIs and their corresponding lobes.

The R-fMRI data set contains 30 EMCI and 20 NC subjects from ADNI2 study which is the latest phase of the ADNI project. The ADNI2 includes many new volunteers in the earliest stages of cognitive impairment, namely EMCI subjects. Compared with MCI in late stages, EMCI does not have dysfunctions in logical memory and delayed recall. The subject inclusion criteria and the data acquisition/pre-processing were documented in Xiang et al. (2013). All the R-fMRI data were collected using 3.0 T MRI equipment (Philips, the Netherlands). During the scanning, subjects were conscious, clear headed and stationary. Scan parameters were as follows: repetition time = 3000; slice thickness = 3.3; field strength = 3.0; echo time = 30.0. Other parameters can be found on the ADNI website (adni.loni.ucla.edu). R-fMRI data were pre-processed as previously described (Chao-Gan and Yu-Feng, 2010). In brief, time-slice correction and motion correction were applied to the data set, so that the head motion was limited to less than 2 mm or 2°. The corrected images underwent spatial standardization, and were mapped to MNI standard space with 3 mm voxels. Finally, the registered images were filtered at low frequency (0.01–0.08 Hz), to reduce low-frequency drift and high-frequency biological noise. Similarly to what we did with the PET data, we mapped the whole-brain to 90 VOIs according to the AAL template.

*AD/NC classification with PIB-PET*

As an initial step, we built a similarity graph over $N = 42$ regions that are potentially related to AD for each subject group. The selection of a small number of VOIs could prevent the weighted adjacency matrix from being too rank deficient and reduce the risk of overfitting the classifier. Suppose we have $M$ subjects in a given group and denote by $X_{ij}$ the average PIB-PET observation in the $i$th brain region of subject $j$. To construct a weighted graph over the $N$ brain volumes, we have assigned a positive weight $W_{ij}$ to the edge between the $i$th and $j$th VOIs as follows

$$W_{ij} = \exp\left(-\frac{\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_2^2}{\rho^2}\right), \tag{22}$$

where $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{iM})$ is the $i$th row of the data matrix $\boldsymbol{X}$ and $\rho > 0$ is a scaling factor. The similarity metric in Eq. (22) is known as a Gaussian RBF kernel (Hofmann et al., 2008). This kernel function allows us to approximate the geodesic distance on the manifold of the brain regions by Euclidean distances with nearby neighbors. The factor $\rho$ controls the notion of proximity in the sense that if $\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_2 \gg \rho$ then $W_{ij} = 0$ and if $\|\boldsymbol{X}_i - \boldsymbol{X}_j\|_2 \ll \rho\rho$ then $W_{ij} = 1$. After building graphs for both groups, we projected a newly observed signal onto the sets of graph Laplacian eigenvectors associated with the two subject groups. The decision was made by comparing the test statistic Eq. (17) of SMSD against zero. For clarity, we present the major steps of our data processing and hypothesis tests in Fig. 4.

**Table 2**
Participant demographics. AD = Alzheimer's disease; NC = normal control subjects; EMCI = early mild cognitive impairment; SD = standard deviation; MMSE = Mini-Mental State Examination; CDR = Clinical Dementia Rating. The age and education are measured in years.

| | Sample size | Avg. age (SD) | Avg. education (SD) | Avg. MMSE (SD) | Avg. CDR |
|---|---|---|---|---|---|
| AD with PIB-PET | 30 (13 male) | 73.03 (7.99) | 17.5 (1) | 25 (1.63) | 1 |
| NC with PIB-PET | 40 (14 male) | 76.15 (8.04) | 16.24 (2.88) | 29.05 (1.10) | 0 |
| EMCI with R-fMRI | 30 (14 male) | 72.15 (7.31) | 17.07 (2.56) | 26.30 (3.08) | 0.5 |
| NC with R-fMRI | 20 (8 male) | 75.16 (6.10) | 16.11 (2.11) | 28.82 (0.82) | 0 |

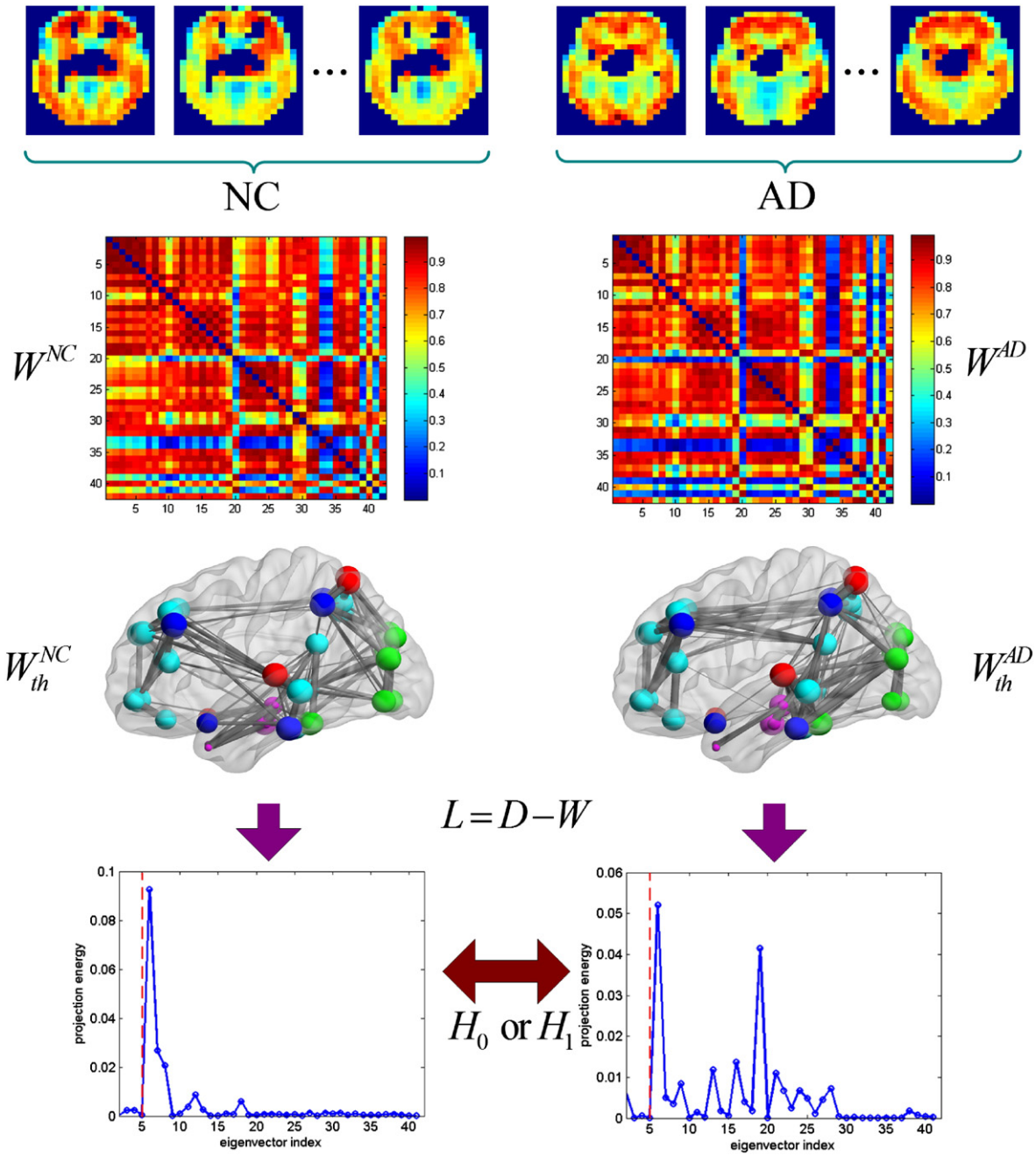


**Fig. 4.** Major steps (from top to bottom) of the MSD implementation for AD network classification. $\mathbf{W}^{NC}$, $\mathbf{W}^{AD}$ are the adjacency matrices of the weighted graphs constructed from the imaging data listed in the first row; while $\mathbf{W}_{th}^{NC}$, $\mathbf{W}_{th}^{AD}$ are their corresponding thresholded versions with 135 edges. The projection energy distributions of a signal on the two sets of eigenvectors of the graph Laplacians are illustrated in the last row, where the graph on the left matches the signal better. Since the first graph Laplacian eigenvectors of the two graphs are the same, we only show the projection energy on the eigenvectors beyond the first one.

We performed leave-one-out cross-validation (LOOCV) and evaluated the classification accuracy of the proposed MSD on graphs for parameter optimization. Specifically, each time we removed one subject to be tested from all the data and presumed that the status of the rest of the subjects were known. The accuracy of the classifier is defined by the proportion of correct decisions among the total number of tests. Following the repeated grid-search cross-validation in Krstajic et al. (2014), Algorithm 1, we tuned the parameters of each classifier as follows: first, we created a grid of points in the parameter vector space with each point indicating a certain combination of all the hyperparameters in the classifier; then, we carried out the LOOCV under each parameter vector and obtained the corresponding accuracy. We chose the parameter combination that yielded the maximum accuracy. If there were multiple parameter vectors for which the accuracy was maximum, then we chose the one with the lowest model complexity. Here the scaling factor $\rho$ in Eq. (22) and the maximum frequency of the bandlimited graph-signal $K$ in Eq. (17) were set to be 1 and 5, respectively. We report the cross-validated results for the classifier. Fig. 5 illustrates the projection errors, namely the residual energy of the PIB-PET data after being projected onto the first 5 graph Laplacian eigenvectors, when the subject belongs to the AD group and NC group respectively. The decision error rates in Figs. 5(a) and (b) are 0/30 and 2/40 accordingly.

In addition, we classified the AD/NC data by using three frequently employed methods: PCA, SVM and LDA. Similarly to the MSD, we tuned the parameters by cross-validation to optimize the performance of each method. For PCA, we used the right singular vectors of the

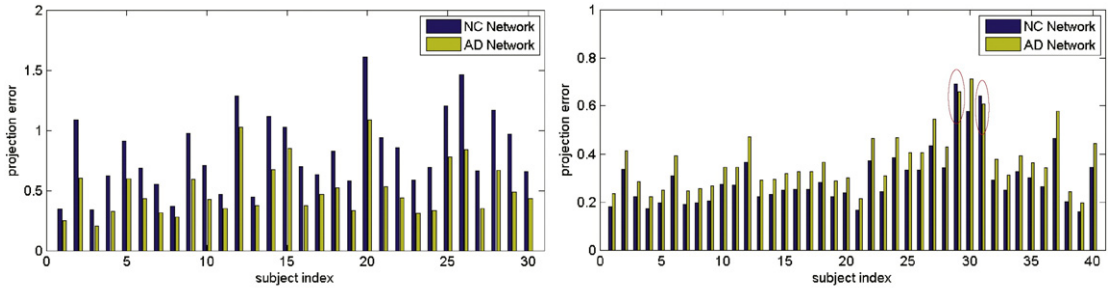


**Fig. 5.** Projection errors in the leave-one-out tests when the true data is from the AD group (left) or NC group (right), with blue/yellow bars showing the residual energy of a certain subject after being projected on AD/NC networks respectively. Red ellipses indicate decision errors made at certain subjects.

data matrix $\boldsymbol{X}$ as an orthogonal basis to carry out the classical matched subspace detection. The associated miss-detection and false-alarm rates of PCA were 4/30 and 7/40 accordingly. For SVM, we constructed a hyperplane in a high-dimensional space so that the training data points were separated by the largest margin. We then decided the status of a new subject by checking the side of the hyperplane on which it falls (Cristianini and Shawe-Taylor, 2000). SVM gave the two types of error rates 3/30 and 5/40, respectively. In machine learning, LDA is another method to find a linear combination of features which characterize or separate two or more classes of objects or events. We implemented it in the following two steps (McLachlan, 2004): first, we found a linear transformation to maximize the ratio of the variance between the classes to the variance within the classes; then, we classified a new data point according to the nearest centroid in the transformed space. With LDA, we obtained 4/30 and 2/40 as miss-detection and false-alarm rates, which were slightly better than those of SVM and PCA. Moreover, since Gaussian RBF kernel was used in the MSD, we also implemented kernel PCA and kernel SVM with Gaussian RBF kernel (Cristianini and Shawe-Taylor, 2000; Hofmann et al., 2008). We refer to these two kernel-based methods as KPCA and KSVM, respectively. The above results were obtained by assuming an equal prior of each hypothesis in (16), namely $\eta = 1$. Subsequently, we changed $\eta$ to get the ROC curve for every classifier in Fig. 6(a). In Table 3(a), we display the accuracy, sensitivity, specificity, and the area under the curve (AUC) of the hypothesis tests. It shows that the AUC of MSD is significantly ($p \le 0.05$) higher than that of the compared algorithms.

To evaluate our methods more comprehensively, we performed additional experiments by using repetitive 10-fold stratified cross-validation (CV). Specifically, we randomly partitioned the subjects into 10 equal-sized folds, making sure that each fold contained the same proportion of subjects from a certain subject group. Out of the 10 folds, a single fold was retained as the validation data for testing the model, and the remaining 9 folds were used as training data. In each CV iteration, we implemented another nested CV on the training data in order to tune the parameters of the classifier. The details of this repeated stratified nested CV are given in Krstajic et al. (2014), Algorithm 2. By repeating the above 10-fold CV with different partitions of the subjects, we can obtain multiple ROC curves for each classifier. Fig. 6(c) displays the averaged ROC curves for MSD and the compared methods after running the 10-fold CV 100 times. From the figure, we find that our method still outperforms the other approaches, although its average AUC is slightly lower than that obtained in the LOOCV. The average AUC for each of the other methods is also lower than the corresponding AUC in the LOOCV. Meanwhile, from Table 3(a), we observe that KPCA has a better performance than PCA in terms of the mean accuracy and mean AUC, while KSVM has a higher mean accuracy but a slightly lower mean AUC when compared with SVM.

We also compared the first few graph Laplacian eigenvectors of the AD and NC networks constructed from the two groups of PIB-PET data. In Figs. 7(a–d), we plotted the second and third eigenvectors of the associated Laplacian matrices both on the brain mesh (above row) and as a sequence against the brain region index (bottom row). Very interestingly, we observe that the patterns of the associated eigenvectors are similar between AD and NC groups, signifying the common features among all the elder subjects. Meanwhile, amplitudes of the eigenvectors in the following five regions are significantly larger than those in other regions: posterior cingulate right, hippocampus left, parahippocampal gyrus left, and middle temporal pole left/right. When comparing the absolute difference between the first few corresponding Laplacian eigenvectors of the AD group and NC group, we find that those five regions are also the most distinctive ones in Figs. 7(e–h). This coincides with the findings that cingulate gyrus, hippocampus area, and temporal areas are the most prominently, and perhaps the earliest, effected sites of AD (Baron et al., 2001; Braak and Braak, 1996; Huang et al., 2002).

*EMCI/NC classification with R-fMRI*

We further applied our MSD method to a more challenging discrimination problem, namely classifying EMCI and NC subjects with R-fMRI data. In order to reduce the dimension of the original data, we divided the whole-brain into 90 regions according to the AAL template (Tzourio-Mazoyera et al., 2002). Then, we averaged the time sequences of all voxels in the same brain region, so that the data reduced to a $90 \times T$ matrix for every subject with $T = 130$ denoting the number of time points for each R-fMRI scan. To construct the graph for each subject group, we concatenated the R-fMRI data into a $90 \times (TM)$ data matrix with $M$ being the number of subjects and normalized the rows of this matrix so that each row had a unit norm.

Next, we assigned the weight between the $i$th and $j$th brain regions according to Eq. (22) with $\rho = 2$, where $\boldsymbol{X}_i$ is the $i$th row of the normalized data matrix $\boldsymbol{X}$.

Using PMSD with the degenerate Gaussian prior Eq. (14), i.e., comparing the variations of newly observed data on two hypothetic graphs, we obtained the ROC curve based on leave-one-out tests in Fig. 6(b). Specifically, we treated the R-fMRI observations on the $N = 90$ brain VOIs at each time step $t = 1, \ldots, T$ as a graph-signal on the brain networks of the EMCI and NC groups. Then, we computed the variations of each graph-signal on the networks of these two groups according to Eq. (7). By comparing the total variations of the $T$ graph-signals from a given subject on the two brain networks, we decided which subject group the testing subject belonged to. When we used PCA, SVM and LDA, we classified the subjects based on the R-fMRI observations on all the VOIs as follows: first, we reshaped the multi-dimensional R-fMRI time series observed on the brain VOIs to a data vector in $\mathbb{R}^{N \times T}$; then, we trained the classifiers with all the training data vectors and classified a new subject according to the criteria that we introduced in the PIB-PET experiments.

In addition, to fully demonstrate the advantage of our method, we have compared it with KPCA/KSVM and PCA/SVM/LDA based on the connectivity measures. As reported in the literature (Chen et al., 2011;

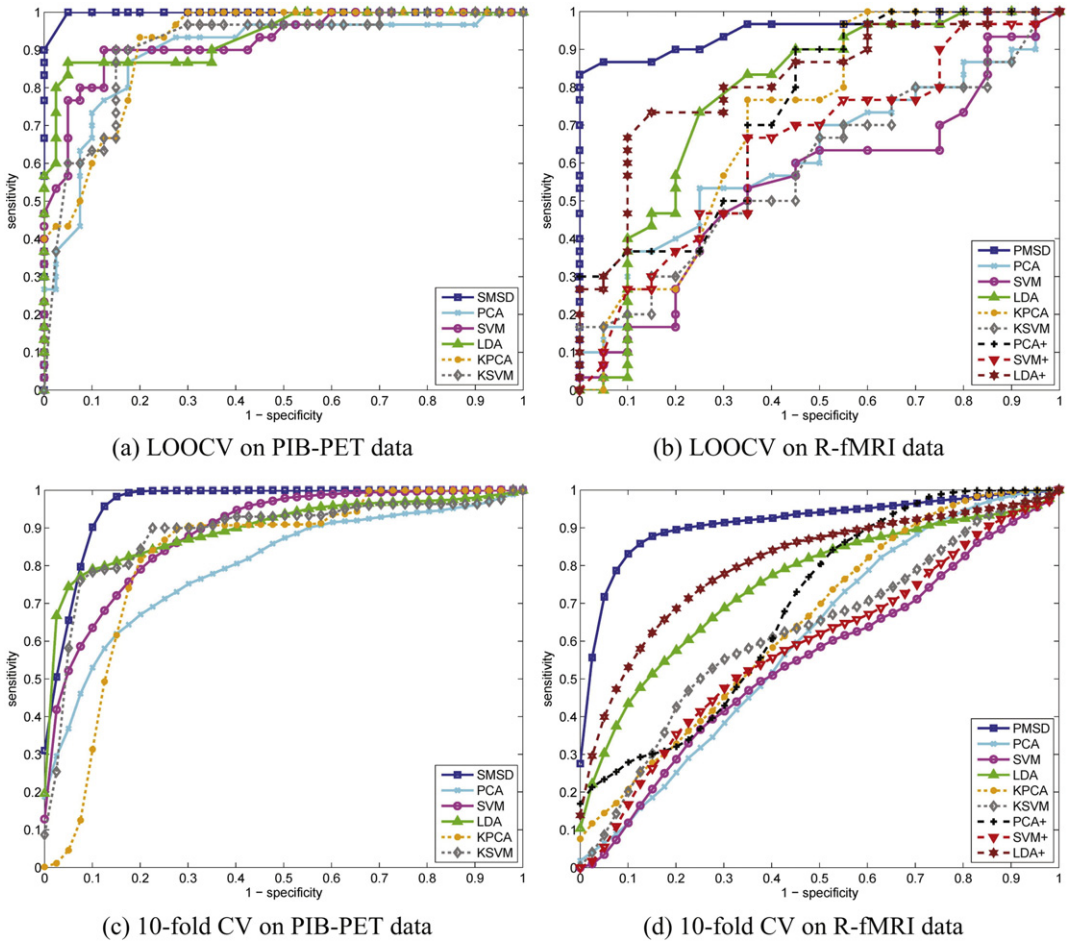


**Fig. 6.** ROC curves for the classifiers based on leave-one-out cross-validation (LOOCV) and 10-fold cross-validation (CV) with PIB-PET data and R-fMRI data. In (a) and (c), we used simple MSD to classify the AD and NC subjects; in (b) and (d), we applied probabilistic MSD to distinguish EMCI and NC subjects. KPCA and KSVM stand for kernel PCA and kernel SVM; PCA+, SVM+, LDA+ indicate applying the associated classifier to the correlation of the R-fMRI data.

**Table 3**
Performance measurements of the classifiers evaluated by both LOOCV and 10-fold CV. In the R-fMRI data classification, PCA+, SVM+, LDA+ indicate applying PCA, SVM, LDA to the correlation of the R-fMRI data. ACR, SEN, SPE, and AUC are the abbreviations for accuracy, sensitivity, specificity, and area under the curve, respectively. For 10-fold CV, mean and standard deviation (in parentheses) of the measurements are listed.

| | LOOCV on PIB-PET data | | | | 10-fold CV on PIB-PET data | | | |
|---|---|---|---|---|---|---|---|---|
| | ACR | SEN | SPE | AUC | ACR | SEN | SPE | AUC |
| (a) PIB-PET data of 30 AD and 40 NC subjects | | | | | | | | |
| MSD | 0.9714 | 1.0000 | 0.9500 | 0.9988 | 0.9245 (0.0153) | 0.9759 (0.0256) | 0.8860 (0.0316) | 0.9612 (0.0142) |
| PCA | 0.8429 | 0.8667 | 0.8250 | 0.8737* | 0.7687 (0.0173) | 0.6537 (0.0315) | 0.8549 (0.0350) | 0.7996 (0.0215) |
| SVM | 0.8857 | 0.9000 | 0.8750 | 0.8975* | 0.8233 (0.0149) | 0.7946 (0.0174) | 0.8448 (0.0224) | 0.8891 (0.0111) |
| LDA | 0.9143 | 0.8667 | 0.9500 | 0.9358* | 0.8731 (0.0177) | 0.7852 (0.0374) | 0.9391 (0.0330) | 0.8994 (0.0123) |
| KPCA | 0.8571 | 0.9333 | 0.8000 | 0.9133* | 0.8199 (0.0239) | 0.8467 (0.0919) | 0.7997 (0.0556) | 0.8242 (0.0284) |
| KSVM | 0.8714 | 0.9000 | 0.8500 | 0.9075* | 0.8597 (0.0234) | 0.7839 (0.0950) | 0.9166 (0.0751) | 0.8841 (0.0226) |
| | LOOCV on R-fMRI data | | | | 10-fold CV on R-fMRI data | | | |
| | ACR | SEN | SPE | AUC | ACR | SEN | SPE | AUC |
| (b) R-fMRI data of 30 EMCI and 20 NC subjects | | | | | | | | |
| MSD | 0.9000 | 0.8667 | 0.9500 | 0.9483 | 0.8862 (0.0297) | 0.8598 (0.0547) | 0.9060 (0.0407) | 0.9147 (0.0236) |
| PCA | 0.5800 | 0.5667 | 0.6000 | 0.6183** | 0.6264 (0.0338) | 0.5534 (0.0829) | 0.6811 (0.1138) | 0.6043 (0.0493) |
| SVM | 0.5800 | 0.6000 | 0.5500 | 0.5466** | 0.6149 (0.0271) | 0.3866 (0.0601) | 0.7860 (0.0426) | 0.5493 (0.0284) |
| LDA | 0.7400 | 0.7333 | 0.7500 | 0.7333* | 0.7369 (0.0332) | 0.6252 (0.0781) | 0.8207 (0.0828) | 0.7510 (0.0393) |
| KPCA | 0.7600 | 0.9667 | 0.4500 | 0.7233* | 0.6528 (0.0352) | 0.5184 (0.1143) | 0.7535 (0.1298) | 0.6512 (0.0610) |
| KSVM | 0.6000 | 0.6667 | 0.5000 | 0.5958** | 0.6651 (0.0279) | 0.5233 (0.0961) | 0.7714 (0.0653) | 0.6267 (0.0321) |
| PCA+ | 0.7600 | 0.9000 | 0.5500 | 0.7525* | 0.6810 (0.0327) | 0.5980 (0.0798) | 0.7432 (0.1059) | 0.6906 (0.0462) |
| SVM+ | 0.6600 | 0.6667 | 0.6500 | 0.6517** | 0.6372 (0.0285) | 0.4580 (0.0567) | 0.7716 (0.0571) | 0.5857 (0.0276) |
| LDA+ | 0.7800 | 0.7333 | 0.8500 | 0.8350* | 0.7776 (0.0314) | 0.7206 (0.0762) | 0.8204 (0.0771) | 0.8054 (0.0360) |

* and ** indicate that the p-value of the improvement of MSD over the associated classical approach is less than 0.05 and 0.001, respectively

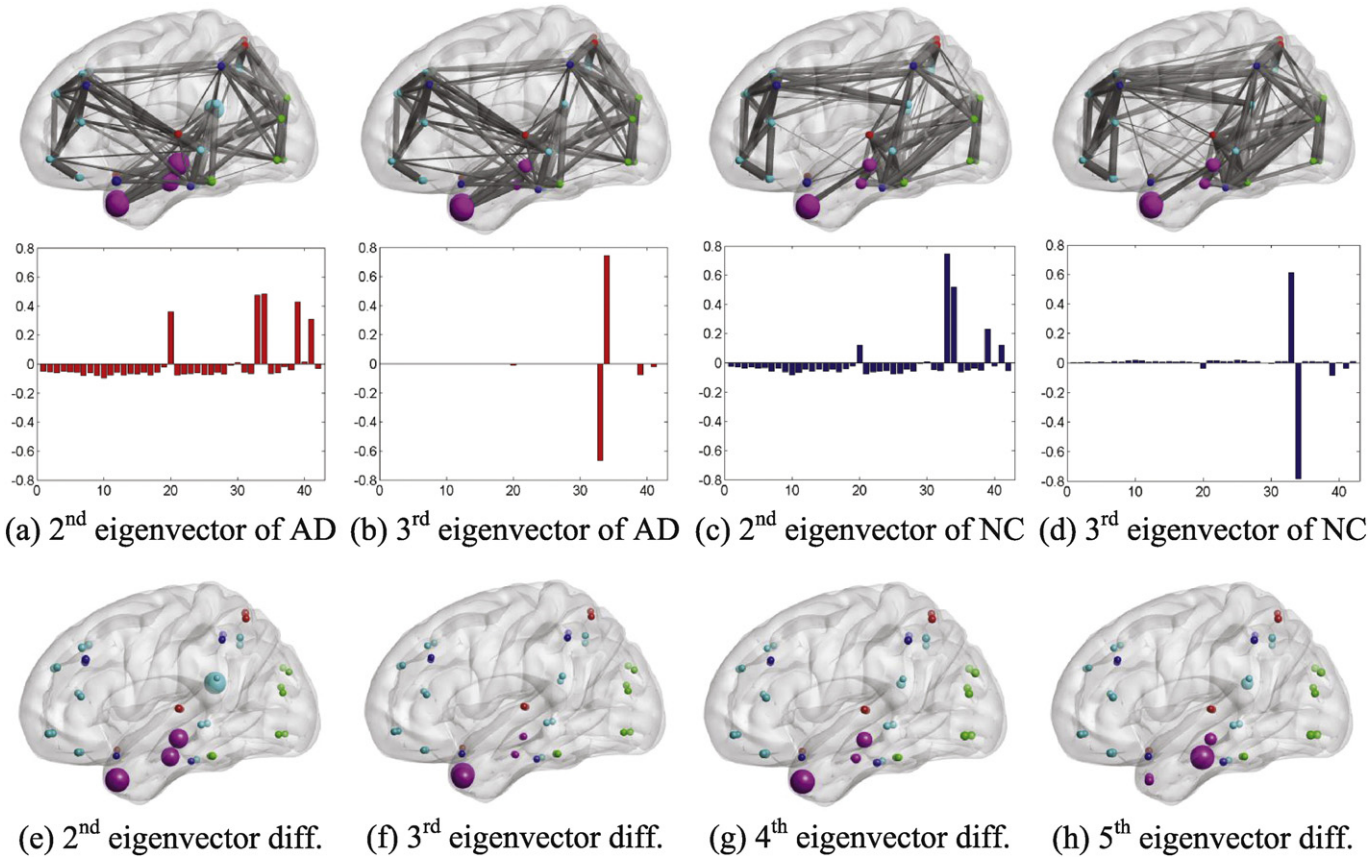


**Fig. 7.** (a)–(d): The 2nd and 3rd eigenvectors of the AD network Laplacian and the NC network Laplacian for the PIB-PET data set. The size and the color of the ball indicate the amplitude of the eigenvector in the given brain region and the modular of that brain region, respectively. (e)–(h): The absolute difference between the AD network Laplacian eigenvector and the NC network Laplacian eigenvector for the PIB-PET data set. The size of the ball denotes the absolute difference in the given brain region.

Gomez-Ramirez and Wu, 2014; Richiardi et al., 2011), functional connectivity measures derived from fMRI might provide better biomarkers for the classification of brain states. Thus, we computed the Pearson correlation coefficient between the R-fMRI time series on every pair of brain VOIs and obtained an $N \times N$ correlation matrix for each subject. Then, we formed a feature vector for each subject by reshaping the upper triangular part of the correlation matrix to a vector. Afterwards, PCA/SVM/LDA were applied to classify the EMCI and NC subjects based on the above features. For convenience, we denote the combination of PCA/SVM/LDA and correlation measures by PCA+/SVM+/LDA+. From Table 3(b), we find that the accuracy achieved by MSD is the best among all the classifiers in either LOOCV or 10-fold CV. In both CV settings, the AUC of the MSD is significantly higher than that of each compared method. We also observe that PCA+/SVM+/LDA+ yield better results than their associated counterparts that directly rely on the time series, and have performance close to that of the corresponding kernel-based methods. For the PIB-PET data, we may also use connectivity measures as features to classify the subjects. However, we have observed that the performance of the classification degenerated when adopting the connectivity measures in this case. Specifically, we computed the similarity between every two VOIs according to the Gaussian RBF kernel in Eq. (22) and classified the PET data based on the similarity measures. Note that the dimension of the PET data of each subject was 42. After computing the similarity, the number of features (i.e., pairwise similarity measures) of each subject was $42 \times 41/2 = 861$. This dramatic growth of the data dimension increased the risk of overfitting when the number of subjects was small and resulted in the performance decay. Thus, we have not included the results of PCA+/SVM+/LDA+ for the PIB-PET data.

## Discussion

### *Assumption of smooth graph-signals*

The introduced MSD schemes were based on the smoothness of the imaging data on brain connectivity networks. We justify this assumption as follows. Through large data analysis, it has been demonstrated that the propagation of disease agents of AD obeys a network diffusion model (Raj et al., 2012; Zhou et al., 2012). By using linear dynamics defined over the brain network, Raj et al. predicted spatially distinct "persistent modes" of different types of dementia accurately. Meanwhile, the smoothness of the signal on graph will increase as the diffusion process continues (Agaskar and Lu, 2013).

Mathematically, the dynamics of many neuroimaging data such as amyloid deposition captured by PIB-PET or blood oxygen measured by fMRI could be modeled as a diffusion process on graph, described by the following differential equation

$$\frac{d\boldsymbol{x}}{dt} = -\boldsymbol{L}\boldsymbol{x}, \tag{23}$$

where $\boldsymbol{L}$ is the graph Laplacian defined before. With the initial condition $\boldsymbol{x}_{\text{init}} = \boldsymbol{\delta}_{u_0}$ representing a unit input at vertex $u_0$, the solution to Eq. (23) is

$$\boldsymbol{x}_t = e^{-t\mathbf{L}}\boldsymbol{\delta}_{u_0} = \sum_{i=1}^{N} e^{-t\lambda_i}\boldsymbol{f}_i\boldsymbol{f}_i^T\boldsymbol{\delta}_{u_0}, \tag{24}$$

where again $\{\lambda_i\}$ are the eigenvalues of $\boldsymbol{L}$ and $\{\boldsymbol{f}_i\}$ are the corresponding eigenvectors. From Eq. (24), as time $t$ increases the contribution of the higher frequencies, namely $\boldsymbol{f}_i$s associated with relatively larger $\lambda_i$s, will decrease quickly. This makes the observed signals at time $t$ smooth on the graph.

*Advantage of MSD and extensions*

Fundamentally, our MSD methods on graphs detect which orthonormal subspace derived from the graph structure of observed data might best capture the properties of a signal, with the subspaces defined by the Laplacian eigenvectors. In this paper, we describe our methods in the framework of graph spectrum analysis. Due to the discrepancies of the eigenvectors of two different graphs, direct comparisons between the GFT coefficients of a graph-signal on these two graphs may not be very meaningful. This is the main difference between the GFT and classical Fourier transforms. However, we can gain insight about how a graph-signal matches up to a certain graph structure by the spectral graph properties. In the Decision Models section, we view the graph Laplacian eigenvectors as an orthonormal basis for each of the two subspaces and detect which subspace fits a graph-signal better. In contrast to the classical matched subspace detectors (Scharf and Friedlander, 1994), we also incorporate prior information about the signal in terms of the signal variation on graphs.

From these experiments, we showed that signal detection techniques on graphs that treat brain imaging data as graph-signals could generate better classification results. PIB-PET imaging of brain amyloid provides a sensitive and consistent biomarker for detecting early AD-related pathological changes (Mintun et al., 2006) and characterizing the network nature of amyloid accumulation in the human brain (Sepulcre et al., 2013). Nevertheless, cognitively normal elderly controls may also have high amyloid binding in the PIB-PET image (Forsberg et al., 2008; Lopresti et al., 2005; Mintun et al., 2006). Similar nonlinearity exists in the R-fMRI data set as well (Friston et al., 2000). Those difficulties hindered the performance of classification with conventional methods. However, we achieved much less classification error as shown in Fig. 5. We found that the two miss-classified subjects (numbers 29 and 31 in Fig. 4) in the figure were both PIB positive,[1] which were more likely to develop AD. Moreover, from Fig. 6 and Table 3, we observe that the MSD outperforms KPCA and KSVM, even if KPCA and KSVM may yield better results than PCA and SVM, accordingly.

The advantage of the MSD on graphs over other approaches might be explained by its capability of modeling the nonlinear data manifold with data-dependent graph structures. Although we could train nonlinear classifiers like kernel SVMs, these nonlinear classifiers are not designed to learn the data manifold. The major difference between the MSD and the kernel-based methods, namely KPCA and KSVM, is that MSD considers the similarity between every pair of brain regions while KPCA and KSVM consider the affinity between every two subjects. With small sample sizes, there may be an overfitting risk in those kernel-base methods as well. In fact, in the PIB-PET data set, we have tried SVM with polynomial or Gaussian kernel which resulted in performance close to that of the linear SVM.

The general MSD framework allows us to select a variety of similarity metrics to build weighted graphs and choose a reasonable signal model to describe real data. While there are more complicated methods to learn the graph structure from the data (Daitch et al., 2009; Hu et al., 2015a), we utilized a simple Gaussian kernel (Eq. (22)) to construct the graph efficiently. Other than the Gaussian kernel which is often used to learn the data manifold (Belkin and Niyogi, 2003; Shuman et al., 2013), we could also use many other kernel functions to build graphs (Hofmann et al., 2008). The kernel function selection offers additional freedom to adapt the graphs to complex neuroimaging data. Meanwhile, we can determine the graph-signal model based on the training data. Basically, we can first try the SMSD and check the distribution of the GFT coefficients. If there are notable patterns in the GFT coefficients, we may further use CMSD and PMSD to classify the data.

In our applications, we have mapped the imaging data to 90 VOIs or selected the AD related VOIs according to prior studies. These procedures decreased the risk of overfitting by effectively reducing the data dimension. In the PIB-PET data, instead of using all the 90 VOIs, choosing the AD related VOIs removed irrelevant information from the data, and therefore increased the signal-to-noise ratio of the data. Our MSD framework can also be generalized to the setting where the sizes of the graphs are large, since it only relies on limited prior information about the data. In many applications, we may need to consider more brain VOIs so as to distinguish more subtle differences between two subject groups. In practice, with the increasing resolution of neuroimaging techniques (De Martino et al., 2011; Dutta et al., 2014) and advancements in brain image parcellation and registration, we are able to construct reliable large-scale brain networks (Bressler and Menon, 2010). When the number of regions increases, one challenge of applying the proposed MSD algorithms is to compute the subspace of the graph Laplacian matrix. However, effective computational methods such as Nyström method have been developed to solve this issue (Gittens and Mahoney, 2013). In addition, when the number of VOIs becomes very large, the number of samples will be much less than the dimension of the sample. As mentioned, overfitting of a classifier might be a concern in this situation. We can greatly avoid this issue by either using the SMSD which has a low model complexity or reducing the complexity of the CMSD or PMSD through imposing sparsity on the penalty weights in Eqs. (11) or (13), respectively.

## Conclusion

In this paper, we formulated the MSD for graph-structure data to classify brain imaging data. We adopted the bandlimited, constrained, and probabilistic graph-signal models to capture the smoothness of the imaging data on brain connectivity networks. We found that GLRT statistics derived under these three signal models were weighted energy detectors in general. The effectiveness of the MSD was demonstrated through simulations and real experiments. Specifically, we applied it to two classification tasks: one is to classify AD and NC subjects with PIB-PET; the other is to discriminate EMCI and NC subjects with R-fMRI. Compared with the widely used conventional classifiers such as PCA, SVM and LDA, the MSD schemes achieved better performance possibly due to the ability of the proposed algorithm to exploit the manifold structure of the neuroimaging data. Our future work will extend the MSD on graphs to the classification of multi-modal and longitudinal brain imaging data.

## Conflict of interest

We confirm that there are no known conflicts of interest associated with this publication.

## Acknowledgments

This research was funded by the National Institutes of Health grants R01 EB013293, 1 K23 EB19023-01, the National Institute on Aging Grants P01 AG036694, R01 AG034556, R01 AG037497, and the Alzheimers Association Zenith Award.

## Appendix A. Derivation of $T_{\mathrm{SMSD}}$ in Eq. (17)

To simplify the presentation, we first consider signal $\mathbf{y} = \mathbf{F}\tilde{\mathbf{x}} + \mathbf{n}$ where $\mathbf{F} = \mathbf{F}_0$ or $\mathbf{F}_1$ and $\tilde{\mathbf{x}} = \tilde{\mathbf{x}}_{\mathcal{G}_0}$ or $\tilde{\mathbf{x}}_{\mathcal{G}_1}$ depending on which hypothesis

[1] Subjects with mean cortical PIB distribution volume ratio greater 1.1.

we make. We can write out the p.d.f. of the observation $\mathbf{y}$ under a certain hypothesis as follows

$$f(\mathbf{y};\tilde{\mathbf{x}},\sigma)=\left(2\pi\sigma^2\right)^{-\frac{N}{2}}\exp\left\{-\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{F}\tilde{\mathbf{x}}\|_2^2\right\}. \tag{A.1}$$

Since $\mathbf{y}=\mathbf{F}\tilde{\mathbf{y}}$, the likelihood function for this multivariate distribution is

$$\ell(\tilde{\mathbf{x}},\sigma;\mathbf{y})=\left(2\pi\sigma^2\right)^{-\frac{N}{2}}\exp\left\{-\frac{1}{2\sigma^2}\|\tilde{\mathbf{y}}-\tilde{\mathbf{x}}\|_2^2\right\}. \tag{A.2}$$

To obtain the MLE of unknown parameters $\tilde{\mathbf{x}}$ and $\sigma$, we take the derivative of the log-likelihood function with respect to $\sigma$:

$$\begin{aligned}\frac{\partial\ell(\tilde{\mathbf{x}},\sigma;\mathbf{y})}{\partial\sigma}&=\frac{\partial}{\partial\sigma}\left[-\frac{N}{2}\log(2\pi\sigma^2)-\frac{1}{2\sigma^2}\|\tilde{\mathbf{y}}-\tilde{\mathbf{x}}\|_2^2\right]\\&=-\frac{N}{\sigma}+\frac{1}{\sigma^3}\|\tilde{\mathbf{y}}-\tilde{\mathbf{x}}\|_2^2.\end{aligned} \tag{A.3}$$

By setting this derivative equal to zero, we achieve the MLE of $\sigma$ in the following formula

$$\sigma^*=N^{-\frac{1}{2}}\|\tilde{\mathbf{y}}-\tilde{\mathbf{x}}\|_2. \tag{A.4}$$

From Eq.(A.2), we know that the MLE of $\tilde{\mathbf{x}}$ will minimize $\|\tilde{\mathbf{y}}-\tilde{\mathbf{x}}\|_2$. Meanwhile, according to the assumption of the bandlimited graph-signal model, the $i$th component of $\mathbf{x}$ satisfies $\tilde{x}_i=0$ for all $i>K$. Thus, the MLE of $\tilde{\mathbf{x}}$ is $\tilde{\mathbf{x}}^*$ given by

$$\tilde{x}_i^*=\begin{cases}\tilde{y}_i, & \text{for } 1\le i\le K,\\ 0, & \text{for } i>K.\end{cases} \tag{A.5}$$

Plugging Eqs. (A.4) and (A.5) to the likelihood function (Eq. (A.2)), we obtain the generalized likelihood:

$$\ell^*\left(\tilde{\mathbf{x}}^*,\sigma^*;\mathbf{y}\right)=\left(2\pi\sigma^2\right)^{-\frac{N}{2}}\exp\left\{-\frac{1}{2\sigma^2}\left(\|\mathbf{y}\|_2^2-\sum_{i=1}^{K}\tilde{y}_i^2\right)\right\}. \tag{A.6}$$

Then, by taking into account the two hypotheses and assuming that the $i$th GFT coefficient of $\mathbf{y}$ on graph $\mathcal{G}_j$ is $\tilde{y}_{i,j}$ for $j=0,1$, we can express the generalized likelihood ratio test (GLRT) as:

$$\frac{\ell^*\left(\tilde{\mathbf{x}}^*_{\mathcal{G}_0},\sigma_0^*;\mathbf{y}\right)}{\ell^*\left(\tilde{\mathbf{x}}^*_{\mathcal{G}_1},\sigma_1^*;\mathbf{y}\right)}=\exp\left\{\frac{1}{2\sigma^2}\left(\sum_{i=1}^{K}\tilde{y}_{i,0}^2-\sum_{i=1}^{K}\tilde{y}_{i,1}^2\right)\right\}. \tag{A.7}$$

Thus, the test statistic under the bandlimited graph-signal model is given by Eq. (17).

## Appendix B. Derivation of $T_{\mathrm{CMSD}}$ in Eq. (19)

We derive the CMSD on graphs by imposing a general constraint $C(\tilde{\mathbf{x}})\le r$ to the graph-signals with the constraint function (Eq. (11)). We assume that the upper bound $r>0$ and the non-negative penalty weights $\{\alpha_i\}_{i=1,\dots,N}$ are specified based on the statistics of the training data. Under this constraint, we could obtain a MLE by solving the following optimization problem

$$\min_{\tilde{\mathbf{x}}}\ N\log\sigma+\frac{1}{2\sigma^2}\|\tilde{\mathbf{x}}-\tilde{\mathbf{y}}\|_2^2, \tag{B.1}$$

$$\text{s.t.}\sum_{i=1}^{N}(\alpha_i-r)\tilde{x}_i^2\le 0. \tag{B.2}$$

By the Karush–Kuhn–Tucker (KKT) condition (Boyd and Vandenberghe, 2004), we have the optimal solutions of $\tilde{\mathbf{x}}$ satisfying

$$\frac{\partial}{\partial\tilde{\mathbf{x}}}\left\{N\log\sigma+\frac{1}{2\sigma^2}\|\tilde{\mathbf{x}}-\tilde{\mathbf{y}}\|_2^2+\frac{1}{2}\gamma\|\mathbf{R}\tilde{x}\|_2^2\right\}=0, \tag{B.3}$$

$$\gamma\sum_{i=1}^{N}(\alpha_i-r)\tilde{x}_i^2=0,\gamma\ge 0. \tag{B.4}$$

Here $\mathbf{R}$ is a diagonal matrix with $R_{ii}=(\alpha_i-r)^{1/2}$ and $\gamma$ is called KKT multiplier. Following Eqs. (B.3), we get

$$\frac{1}{\sigma^2}(\tilde{\mathbf{x}}-\tilde{\mathbf{y}})+\gamma\mathbf{R}^2\tilde{\mathbf{x}}=0, \tag{B.5}$$

which leads to $\tilde{x}_i=\frac{\tilde{y}_i}{1+\gamma\sigma^2(\alpha_i-r)}$. Replacing $\tilde{x}_i$ in Eqs. (B.4) with this, we can write

$$\gamma\sum_{i=1}^{N}\frac{(\alpha_i-r)\tilde{y}_i^2}{\left(1+\gamma\sigma^2(\alpha_i-r)\right)^2}=0, \tag{B.6}$$

from which we could determine the KKT multiplier $\gamma$ if we assume the noise level $\sigma$ is known. If $\sigma$ is unknown, we may estimate it from the average residual energy of the projection as follows

$$\sigma=\sqrt{\frac{\sum_{i=N'+1}^{N}\tilde{y}_i^2}{N-N'}}, \tag{B.7}$$

with $N'<N$ being an integer threshold. After solving for $\gamma$, we can obtain a MLE of $\tilde{x}$ by solving Eq. (B.5).

For clarity, we denote by $\gamma_j^*$ the solution of the KKT multiplier and $\tilde{x}_{i,j}^*$ the estimation of the $i$th GFT coefficient of $\tilde{\mathbf{x}}$ under hypothesis $\mathcal{H}_j, j=0,1$. In addition, we assume that the upper bound of the constraint function is $r_j, j=0,1$ under each hypothesis. Then, by plugging the solution of $\tilde{\mathbf{x}}$ into the likelihood expression in Eq. (16), we reach the GLRT with the test statistic given by

$$\begin{aligned}T_{\mathrm{CMSD}}&=\sum_{i=1}^{N}\left[\tilde{y}_{i,0}-\tilde{x}_{i,0}^*)^2-\left(\tilde{y}_{i,1}-\tilde{x}_{i,1}^*\right)^2\right]\\&=\sum_{i=1}^{N}\left[\frac{\tau_{i,0}^2}{\left(1+\tau_{i,0}\right)^2}\tilde{y}_{i,0}^2-\frac{\tau_{i,1}^2}{\left(1+\tau_{i,1}\right)^2}\tilde{y}_{i,1}^2\right]\end{aligned} \tag{B.8}$$

where $\tau_{i,j}=\gamma_j^*\sigma_j^2(\alpha_i-r_j)$ for $j=0,1$ with the noise level $\sigma_j$ under $\mathcal{H}_j$ being known or estimated by (B.7). Notice that $\tau_{i,j}=0$ for $i=1,\dots,N$ when the constraint (Eq. (B.2)) is met, since $\gamma_j^*=0$ in this situation. In addition, $\tau_{i,j}=0$ if the noise level $\sigma_j=0$.

## References


Agaskar, A., Lu, Y.M., 2013. A spectral graph uncertainty principle. IEEE Trans. Inf. Theory 59, 4338–4356.

Azari, N.P., Rapoport, S.I., Grady, C.L., Schapiro, M.B., Salerno, J.A., Gonzalez-Aviles, A., Horwitz, B., 1992. Patterns of interregional correlations of cerebral glucose metabolic rates in patients with dementia of the Alzheimer type. Neurodegeneration 1, 101–111.

Balthazar, M.L.F., de Campos, B.M., Franco, A.R., Damasceno, B.P., Cendes, F., 2014. Whole cortical and default mode network mean functional connectivity as potential biomarkers for mild Alzheimer's disease. Psychiatry Res. Neuroimaging 221, 37–42.

Baron, J., Chetelat, G., Desgranges, B., Perchey, G., Landeau, B., De La Sayette, V., Eustache, F., 2001. In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease. NeuroImage 14, 298–309.

Bassett, D.S., Bullmore, E., 2006. Small-world brain networks. Neuroscientist 12, 512–523.

Belkin, M., Niyogi, P., 2003. Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput. 15, 1373–1396.

Boyd, S., Vandenberghe, L., 2004. Convex Optimization. Cambridge university press.

Braak, H., Braak, E., 1996. Development of Alzheimer-related neurofibrillary changes in the neocortex inversely recapitulates cortical myelogenesis. Acta Neuropathol. 92, 197–201.

Bressler, S.L., Menon, V., 2010. Large-scale brain networks in cognition: emerging methods and principles. Trends Cogn. Sci. 14, 277–290.

Brookmeyer, R., Johnsona, E., Ziegler-Grahamb, K., Arrighic, H.M., 2007. Forecasting the global burden of Alzheimers disease. Alzheimers Dement. 3, 186–191.
Chao-Gan, Y., Yu-Feng, Z., 2010. DPARSF: a MATLAB toolbox for "pipeline" data analysis of resting-state fMRI. Front. Syst. Neurosci. 4.
Chen, G., Ward, B.D., Xie, C., Li, W., Wu, Z., Jones, J.L., Franczak, M., Antuono, P., Li, S.J., 2011. 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, 213–221.
Chung, F.R., 1997. Spectral graph theory. Am. Math. Soc. 92.
Cristianini, N., Shawe-Taylor, J., 2000. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press.
Daitch, Samuel I., Kelner, Jonathan A., Spielman, Daniel A., 2009. Fitting a graph to vector data. Proceedings of the 26th Annual International Conference on Machine Learning. ACM.
De Martino, F., Esposito, F., van de Moortele, P.F., Harel, N., Formisano, E., Goebel, R., Ugurbil, K., Yacoub, E., 2011. Whole brain high-resolution functional imaging at ultra high magnetic fields: an application to the analysis of resting state networks. NeuroImage 57, 1031–1044.
Dutta, J., Li, Q., Johnson, K., Zhu, X., El Fakhri, G., 2014. High resolution pet imaging of tau using an MR-based information theoretic anatomical prior. J. Nucl. Med. 55, 642-642.
Forsberg, A., Engler, H., Almkvist, O., Blomquist, G., Hagman, G., Wall, A., Ringheim, A., Langstrom, B., Nordberg, A., 2008. PET imaging of amyloid deposition in patients with mild cognitive impairment. Neurobiol. Aging 29 (10), 1456–1465.
Friston, K.J., Mechelli, A., Turner, R., Price, C.J., 2000. Nonlinear responses in fMRI: the balloon model, volterra kernels, and other hemodynamics. NeuroImage 12, 466–477.
Gadde, A., Narang, S.K., Ortega, A., 2013. Bilateral Filter: Graph Spectral Interpretation and Extensions (arXiv, preprint arXiv:1303.2685).
Gittens, A., Mahoney, M.W., 2013. Revisiting the Nystrom Method for Improved Large-scale Machine Learning (arXiv, preprint arXiv:1303.1849).
Gomez-Ramirez, J., Wu, J., 2014. Network-based biomarkers in Alzheimer's disease: review and future directions. Front. Aging Neurosci. 6.
He, Y., Chen, Z.J., Evans, A.C., 2007. Small-world anatomical networks in the human brain revealed by cortical thickness from MRI. Cereb. Cortex 17, 2407–2419.
Hofmann, T., Schölkopf, B., Smola, A.J., 2008. Kernel methods in machine learning. Ann. Stat. 1171–1220.
Horn, R.A., Johnson, C.R., 1990. Matrix Analysis. Cambridge university press.
Hu, C., Cheng, L., Sepulcre, J., Johnson, K.A., Fakhri, G.E., Lu, Y.M., Li, Q., 2015a. A spectral graph regression model for learning brain connectivity of Alzheimer's disease. PLoS One 10 (5), e0128136.
Hu, W., Cheung, G., Ortega, A., Au, O.C., 2015 bb. Multiresolution graph Fourier transform for compression of piecewise smooth images. Image Process., IEEE Trans. 24, 419–433.
Huang, C., Wahlund, L.O., Svensson, L., Winblad, B., Julin, P., 2002. Cingulate cortex hypoperfusion predicts Alzheimer's disease in mild cognitive impairment. BMC Neurol. 2, 9.
Huang, S., Li, J., Sun, L., Liu, J., Wu, T., Chen, K., Fleisher, A., Reiman, E., Ye, J., 2009. Learning brain connectivity of Alzheimer's disease from neuroimaging data. Advances in Neural Information Processing Systems, pp. 808–816.
Jolliffe, I., 2005. Principal Component Analysis. John Wiley & Sons, Ltd.
Kheradmand, A., Milanfar, P., 2014. A general framework for regularized, similarity-based image restoration. IEEE Trans. Image Process. 23, 5136–5151.
Kim, W.H., Adluru, N., Chung, M.K., Charchut, S., GadElkarim, J.J., Altshuler, L., Moody, T., Kumar, A., Singh, V., Leow, A.D., 2013. Multi-resolutional brain network filtering and analysis via wavelets on non-Euclidean space. Medical Image Computing and Computer-Assisted Intervention — MICCAI. Springer, pp. 643–651.
Krstajic, D., Buturovic, L.J., Leahy, D.E., Thomas, S., 2014. Cross-validation pitfalls when selecting and assessing regression and classification models. J.Chem. Inf. 6, 1–15.
Lee, G., Rodriguez, C., Madabhushi, A., 2008. Investigating the efficacy of nonlinear dimensionality reduction schemes in classifying gene and protein expression studies. IEEE/ACM Trans. Comput. Biol. Bioinform. 5, 368–384.
Li, Z., Li, Q., Yu, X., Conti, P.S., Leahy, R.M., 2009. Lesion detection in dynamic FDG–PEG using matched subspace detection. IEEE Trans. Med. Imaging 28 (2), 230–240.
Liu, X., Tosun, D., Weiner, M.W., Schuff, N., Initiative, A.D.N., et al., 2013. Locally linear embedding (LLE) for MRI based Alzheimer's disease classification. NeuroImage 83, 148–157.
Lopresti, B.J., Klunk, W.E., Mathis, C.A., Hoge, J.A., Ziolko, S.K., Lu, X., Meltzer, C.C., Schimmel, K., Tsopelas, N.D., DeKosky, S.T., Price, J.C., 2005. Simplified quantification of Pittsburgh compound B amyloid imaging PET studies: a comparative analysis. J. Nucl. Med. 46 (12), 1959–1972.
Manolakis, D., Shaw, G., 2002. Detection algorithms for hyperspectral imaging applications. IEEE Signal Process. Mag. 19 (1), 29–43.
Mathis, C.A., Bacskai, B.J., Kajdasz, S.T., McLellan, M.E., Frosch, M.P., Hyman, B.T., Holt, D.P., Wang, Y., Huang, G.F., Debnath, M.L., et al., 2002. A lipophilic thioflavin-T derivative for positron emission tomography (PET) imaging of amyloid in brain. Bioorg. Med. Chem. Lett. 12 (3), 295–298.
McLachlan, G.J., 2004. Discriminant Analysis and Statistical Pattern Recognition. Wiley-Interscience.
Meyer, F.G., Shen, X., 2014. Perturbation of the eigenvectors of the graph laplacian: application to image denoising. Appl. Comput. Harmon. Anal. 36, 326–334.
Mintun, M.A., Larossa, G.N., Sheline, Y.I., Dence, C.S., Lee, S.Y., Mach, R.H., Klunk, W.E., Mathis, C.A., DeKosky, S.T., Morris, J.C., 2006. [11C] PIB in a nondemented population potential antecedent marker of Alzheimer disease. Neurology 67, 446–452.
Oppenheim, A.V., Willsky, A.S., 1997. Signals and Systems. Prentice-Hall.
Raj, A., Kuceyeski, A., Weiner, M., 2012. A network diffusion model of disease progression in dementia. Neuron 73, 1204–1215.
Richiardi, J., Eryilmaz, H., Schwartz, S., Vuilleumier, P., Van De Ville, D., 2011. Decoding brain states from fMRI connectivity graphs. NeuroImage 56, 616–626.
Roweis, S.T., Saul, L.K., 2000. Nonlinear dimensionality reduction by locally linear embedding. Science 290, 2323–2326.
Sandryhaila, A., Moura, J.M.F., 2013. Discrete signal processing on graphs. IEEE Trans. Signal Process. 61 (7), 1644–1656.
Sanz-Arigita, E.J., Schoonheim, M.M., Damoiseaux, J.S., Rombouts, S.A., Maris, E., Barkhof, F., Scheltens, P., Stam, C.J., 2010. Loss of 'small-world' networks in Alzheimer's disease: graph analysis of fMRI resting-state functional connectivity. PLoS One 5, e13788.
Saul, L.K., Weinberger, K.Q., Ham, J.H., Sha, F., Lee, D.D., 2006. Spectral Methods for Dimensionality Reduction. Semisupervised learning, MIT Press, Cambridge, MA, pp. 293–308.
Scharf, L.L., Friedlander, B., 1994. Matched subspace detectors. IEEE Trans. Signal Process. 42 (8), 2146–2157.
Sepulcre, J., Sabuncu, M.R., Becker, A., Sperling, R., Johnson, K.A., 2013. In vivo characterization of the early states of the amyloid-beta network. Brain 136, 2239–2252.
Shuman, D., Narang, S., Frossard, P., Ortega, A., Vandergheynst, P., 2013. Signal processing on graphs: extending high-dimensional data analysis to networks and other irregular data domains. IEEE Signal Process. Mag. 30, 83–98.
Stam, C., Jones, B., Nolte, G., Breakspear, M., Scheltens, P., 2007. Small-world networks and functional connectivity in Alzheimer's disease. Cereb. Cortex 17, 92–99.
Stanley, M.L., Moussa, M.N., Paolini, B.M., Lyday, R.G., Burdette, J.H., Laurienti, P.J., 2013. Defining nodes in complex brain networks. Front. Comput. Neurosci. 7.
Supekar, K., Menon, V., Rubin, D., Musen, M., Greicius, M.D., 2008. Network analysis of intrinsic functional brain connectivity in Alzheimer's disease. PLoS Comput. Biol. 4 (6), 1–11.
Tenenbaum, J.B., De Silva, V., Langford, J.C., 2000. A global geometric framework for nonlinear dimensionality reduction. Science 290, 2319–2323.
Tzourio-Mazoyera, N., Landeaub, B., Papathanassioua, D., Crivelloa, F., Etarda, O., Delcroixa, N., Mazoyerc, B., Joliota, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage 15 (1), 273–289.
Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of "small-world" networks. Nature 2, 393–440.
Wee, C.Y., Yap, P.T., Denny, K., Browndyke, J.N., Potter, G.G., Welsh-Bohmer, K.A., Wang, L., Shen, D., 2012. Resting-state multi-spectrum functional connectivity networks for identification of MCI patients. PLoS One 7, e37828.
Xiang, J., Guo, H., Cao, R., Liang, H., Chen, J., 2013. An abnormal resting-state functional brain network indicates progression towards Alzheimer's disease. Neural Regener. Res. 8, 2789.
Zalesky, A., Fornito, A., Harding, I.H., Cocchi, L., Yücel, M., Pantelis, C., Bullmore, E.T., 2010. Whole-brain anatomical networks: does the choice of nodes matter? NeuroImage 50, 970–983.
Zhou, J., Gennatas, E.D., Kramer, J.H., Miller, B.L., Seeley, W.W., 2012. Predicting regional neurodegeneration from the healthy brain functional connectome. Neuron 73, 1216–1227.