Published in final edited form as:

Proc IEEE Int Conf Data Min. 2016 December ; 2016: 301–310. doi:10.1109/ICDM.2016.0041.

New Probabilistic Multi-Graph Decomposition Model to Identify Consistent Human Brain Network Modules

Dijun Luo[*], **Zhouyuan Huo**[*], **Yang Wang**[†], **Andrew J. Saykin**[‡], **Li Shen**[‡], and **Heng Huang**[*]

[*]Department of Computer Science and Engineering, University of Texas at Arlington, USA

[†]Department of Radiology, Medical College of Wisconsin, USA

[‡]Department of Radiology and Imaging Sciences, Indiana University School of Medicine, USA

Abstract

Many recent scientific efforts have been devoted to constructing the human connectome using Diffusion Tensor Imaging (DTI) data for understanding large-scale brain networks that underlie higher-level cognition in human. However, suitable network analysis computational tools are still lacking in human brain connectivity research. To address this problem, we propose a novel probabilistic multi-graph decomposition model to identify consistent network modules from the brain connectivity networks of the studied subjects. At first, we propose a new probabilistic graph decomposition model to address the high computational complexity issue in existing stochastic block models. After that, we further extend our new probabilistic graph decomposition model for multiple networks/graphs to identify the shared modules cross multiple brain networks by simultaneously incorporating multiple networks and predicting the hidden block state variables. We also derive an efficient optimization algorithm to solve the proposed objective and estimate the model parameters. We validate our method by analyzing both the weighted fiber connectivity networks constructed from DTI images and the standard human face image clustering benchmark data sets. The promising empirical results demonstrate the superior performance of our proposed method.

Index Terms

Probabilistic Graph Decomposition; Multi-Graph Decomposition; Human Connectome

I. Introduction

Advent of diffusion MRI technology has made tremendous progress over the last decade [1] and enables us to use Diffusion Tensor Imaging (DTI) for non-invasive in vivo white matter mapping of the human brain by the inference of axonal fiber pathways from local water diffusion [2]. DTI combined with tractography allows reconstruction of the major fiber bundles in the brain and also permits the mapping of white matter cortico-cortical and cortico-subcortical projections at high spatial resolution. These studies enable the analysis of the human connectome as organizational principle of the central nervous system.

Understanding the structural basis of functional connectivity patterns requires a comprehensive map of structural connection of the human brain, which has been conceptualized as the human connectome [3]. A connectome is a comprehensive description of the network elements and connections that form the brain. Such clear and comprehensive knowledge of anatomical connections lies at the basis of understanding network functions. The connectome can be represented as a large interconnected graph, in which nodes are neuroanatomical regions and synapses are bundles of white matter tracts. The resultant networks exhibit important topological properties such as small-worldness and highly connected hubs regions in the posterior medial cortical regions. These studies have accelerated our understandings of human connectome [4], [5], [6].

Although many network and graph analysis tools have been applied to human connectome studies, most of them focus on analyzing the connectome of each subject individually. How to find the consistent network modules from a group of subjects under the same condition (*e.g.* normal or Alzheimer) is important to understand the underlying brain structural and functional mechanisms. To solve this challenging problem, we propose a Bayesian inference based approach to identify consistent network modules from brain connectivity networks of multiple subjects. We explore a new graphical model (probabilistic multi-graph decomposition) to incorporate multiple networks and inference hidden block state variables, by which we identify local cliques among the graphs. The common connectome modules are then pruned from the cliques. By analyzing the weighted fiber connectivity network from 24 young male adults, we identify 4 consistent network modules which consistently carry high connectivity among all the subjects. To show the superior clustering capability of our new model, we also evaluate our method using the human face image clustering benchmark data sets.

We will organize the rest of this paper by the following order. First, we will introduce the previous Mixed Membership Block Model, which is powerful to group data points through graph but not computationally efficient (use $O(n^2)$ latent valuables) and also not for multi-graph situation. Second, we will propose a new probabilistic graph decomposition model to address the computational efficiency problem for single graph case. Third, we will further introduce the new Probabilistic Multi-Graph Decomposition (PMGD) method with the optimization algorithm for multi-graph problem. After that, we will provide the data description and the details of brain connectivity construction. At last, we will show the empirical results on both medical image analysis and human face image clustering tasks to support the proposed algorithms.

II. Probabilistic Multi-Graph Learning Model

A. Problem Description and Related Work

The brain connectome of each subject can be represented as a graph **G**, in which each nodes is an ROI (region of interest) in human brain and the weight of each edge is the density of the nerve fibers connecting a pair of nodes. In later section, we will describe the details of brain network construction. In this section, we will focus on the new probabilistic graphic model to formulate the multiple brain connectivity networks and identify the consistent network modules.

For m subjects with n ROIs, we can denote their connectivity networks as $\mathbf{G}^1, \mathbf{G}^2, \cdots, \mathbf{G}^m$, where $\mathbf{G}^k \in \mathbb{R}^{n \times n}$ and $\mathbf{G}^k_{ij}$ denotes the connectivity of the i-th ROI and the j-th ROI in the k-th subject, $k = 1, 2, \cdots, m$, $1 \leq i, j \leq n$. Given these m networks, we hope to discover the consistent network modules, *i.e.* common structures of connectivity, which are shared by all subjects. Given S as the set of all nodes, we say a subset of $S: c \subset S$ is a "consistent network module" if all the $\mathbf{G}^k_c$ are highly connected module, for $k = 1, 2, \cdots, m$, where

$$\left(\mathbf{G}^k_c\right)_{ij} = \mathbf{G}^k_{c_i, c_j}, 1 \leq i, j \leq |c|.$$

Related work on pattern analysis of graphs falls into two folds. The first fold is *spectral graph partitioning* which clusters objects into groups on the spectral embedding space [7]. The second category is *stochastic block modeling*, in which the graphs are assumed to be the observations of a pair-dependent stochastic block model [8].

However, all these models are not applicable in our problem, since these models only accept single graph as input and no trivial extensions of these methods are available to handle multiple graphs. Meanwhile, the existing stochastic block models require high computational complexity, which limits their practical applications. To address these challenging problems, in this paper we will propose a new and efficient graphical model to capture the hidden generative dependency among the ROIs in structural brain activities from multiple graphs. We will develop the likelihood function for the model and present an EM-like algorithm to estimate the model parameters by maximizing the likelihood.

B. Previous Mixed Membership Block Model

We first provide a brief review of previous Mixed Membership Block Model (MMB) [8]. MMB extends the mixed membership models, such as latent Dirichlet allocation [9], which have emerged in recent years as a flexible modeling tool for data in which the single group assumption is violated by the heterogeneity within a unit of analysis. They have been successfully applied in many domains, *e.g.* natural scene categories learning [10].

Mixed membership models associate each node of graph with multiple groups rather than a single group, via a membership probability-like vector. More specifically MMB models assume that a random graph is generated by the following model:

- For each node i, sample $\vec{\pi}_i \sim \mathbf{Dir}(\vec{\theta})$.
- For each node pair (i, j),
  - Sample $z_{ij} \sim \mathbf{Mul}(\vec{\pi}_i)$
  - Sample $\vec{z}_{ji} \sim \mathbf{Mul}(\vec{\pi}_j)$
  - Sample $\mathbf{G}_{ij} \sim \mathbf{Ber}\left(\overleftarrow{z}^T_{ij} B \vec{z}_{ji}\right)$

[1] http://www.fmrib.ox.ac.uk/fsl.html

where $z\leftarrow_{ij}$ and $\vec{z}_{ji}$ are $K \times 1$ vectors, indicating which group the nodes belong to, *i.e.*, if the node belongs to group t, the t position is 1, and all other positions are 0. **B** is a $K \times K$ matrix where K is the number of blocks (*e.g.* the number of topics or clusters). Here we denote the **Dirichlet, Multinomial**, and **Bernouli** distributions by **Dir**, **Mul** and **Ber**.

This generative model resamples the membership indicator $z\leftarrow_{ij}$ and $\vec{z}_{ji}$ for every node pair. Notice that θ, B are constant quantities to be estimated, and while $\vec{\pi}_1$, $\vec{\pi}_2$, $\cdots$, $\vec{\pi}_n$, $z\leftarrow_{11}$, $z\leftarrow_{12}$, $\cdots$, $z\leftarrow_{nn}$, $\vec{z}_{11}$, $\vec{z}_{12}$, $\cdots$, $\vec{z}_{nn}$ are unknown variable quantities whose posterior distribution needs to be determined. They employ the variational EM [11] procedure to carry out approximate estimation and inference approximately. This model is successfully applied in relational data modeling [8].

C. New Efficient Probabilistic Graph Decomposition Model

The mixed membership formalism is a particularly natural idea for relational data, where the objects can bear multiple latent roles or cluster-memberships that influence their relationships to others. However, from point of view of clustering, this assumption is not natural. In most data mining applications, each node usually belongs to a unique cluster. For example, in image segmentation by clustering pixels, it is possible that pixels from different objects (segments) might have connectivity (similar in color and texture, or close in space), but we always assume each pixel belongs to a unique object. Another examples is human face image clustering. It is not natural to allow a single image to belong to different persons.

Moreover, the number of latent valuables is $O(n^2)$, where n is the number of nodes, which leads to prohibitively computational complexity in most of computer vision applications.

In order to address these issues, we propose Probabilistic Graph Decomposition in which the membership indicators are sampled once (instead of n times) for each node. To simplify the problem, we use undirect graph as example (one can easily generate it into directed graph). We assume the observation data are generated by the following model:

- For each node i
  - Sample $\vec{\pi}_i \sim \mathbf{Dir}(\vec{\theta})$.
  - Sample $\vec{z}_i \sim \mathbf{Mul}(\vec{\pi}_i)$.
- For each node pair (i, j),
  - Sample $G_{ij} \sim \mathbf{Ber}\left(\vec{z}_i^{\,T} B \vec{z}_j\right)$

Since each $\vec{\pi}_i$ is sampled independently, we further reduce the generative model by ignoring the distribution of $\vec{z}_i$ and consider $\vec{z}_i$ as a free parameter, see Figure 1.

D. New Probabilistic Multi-Graph Decomposition Model

In this paper, we target to identify the consistent network modules, thus we need model the multiple graph block structures. The above methods are only designed for single network or graph. Thus, we propose a new Probabilistic Multi-Graph Decomposition (PMGD) model to formulate the multiple structural brain connectivity networks and identify hidden consistent

network modules. The reason of using graphical model is that all the observed graphs are naturally integrated in the model and principled to learn the model by fitting the real anatomical data. The mixed membership formalism is a particularly natural idea for relational data, where the objects can bear multiple latent roles or cluster-memberships that influence their relationships to others given multiple graphs.

Based our above probabilistic graph decomposition model, we assume the observation data are drawn by the following generative model:

- For each node i, $i = 1, 2, \cdots, n$
  - Sample $\vec{\pi}_i \sim \mathbf{Dir}(\vec{\theta})$, where $\vec{\theta}, \vec{\pi}_i \in \mathbb{R}^K$,
  - Sample $\vec{z}_i \sim \mathbf{Mul}(\vec{\pi}_i)$, where $\vec{z}_i \in \{0, 1\}^K$.
- For each node pair (i, j), $1 \le i, j \le n$,
  - For each graph k, Sample $\mathbf{G}^k_{ij} \sim \mathbf{Ber}\left(\vec{z}_i^T \mathbf{B} \vec{z}_j\right)$,

where $\mathbf{B} \in \mathbb{R}^{K \times K}$ and $\vec{\theta} \in \mathbb{R}^{K \times 1}$ are the model parameters, K is the number of blocks. The dependency diagram of our model is illustrated in Figure 1. In this model, we assume the ROIs belong to K groups. If ROI i belongs to the p-th group and ROI j belongs to q-th group, then the observation of $\mathbf{G}_{ij}$ has a probability of $\mathbf{B}_{pq}$ to be 1 and $1 - \mathbf{B}_{pq}$ to be zero. Then a reasonable $\mathbf{B}$ should have a diagonal structure, where the diagonal elements have large value and off-diagonal elements have values close to zero. We will show this property in the experimental section.

Since each $\vec{\pi}_i$ is sampled independently, we further reduce the generative model by ignoring the distribution of $\vec{z}_i$ and consider $\vec{z}_i$ as a free parameter. In order to balance among the individual difference between subjects, we discretize the weighted graph to binary graph by thresholding. We use binary graph $\mathbf{G}^k$ as input in our algorithm.

III. Optimization Algorithm for PMGD Model

We are going to derive the algorithm to inference the model parameters to fit the observations of m connectivity graphs $\mathbb{G} = \{\mathbf{G}^1, \mathbf{G}^2, \cdots, \mathbf{G}^m\}$. For convenience, we denote $\mathbf{Z} = [\vec{z}_1, \vec{z}_2, \cdots, \vec{z}_n]^T \in \mathbb{R}^{n \times K}$.

The distribution of $\mathbb{G}$ given $\mathbf{Z}$ and $\mathbf{B}$ is,

$$P(\mathbb{G} \mid \mathbf{B}, \mathbf{Z}) = \prod_{k=1}^{m} \prod_{ij} \left(\vec{z}_i^T \mathbf{B} \vec{z}_j\right)^{\mathbf{G}^k_{ij}} \left(1 - \vec{z}_i^T \mathbf{B} \vec{z}_j\right)^{1 - \mathbf{G}^k_{ij}}. \quad (1)$$

We construct an **indicator vector c** by $c_i = \mathrm{argmax}_k Z_{ik}$, and Eq. (1) can be written as

$$P(\mathbb{G} \mid \mathbf{B}, \mathbf{c}) = \prod_{k=1}^{m} \prod_{ij} \left(\mathbf{B}_{c_i c_j}\right)^{\mathbf{G}_{ij}^k} \left(1 - \mathbf{B}_{c_i c_j}\right)^{1 - \mathbf{G}_{ij}^k}. \quad (2)$$

We will use **c** the represent the membership indicator in the rest of this paper. To estimate the parameters of the PMGD model, we solve the following optimization problem:

$$\max_{\mathbf{B}, \mathbf{c}} L(\mathbf{B}, \mathbf{c}) = \sum_{ij} \left\{ \sum_{k=1}^{m} \mathbf{G}_{ij}^k \log \mathbf{B}_{c_i c_j} + (m - \sum_{k=1}^{m} \mathbf{G}_{ij}^k) \log(1 - \mathbf{B}_{c_i c_j}) \right\} \quad s.t. \quad \mathbf{0} \le \mathbf{B} \le \mathbf{1}. \quad (3)$$

Here we set $\mathbf{0} \le \mathbf{B} \le \mathbf{1}$ to restrict elements in **B** with the probability constraint. We will show that our solution automatically satisfies the constraint. We solve Eq. (3) as following. Initialize **c** and then iteratively: (1) solve **B** while fixing **c** and (2) solve **c** while fixing **B** until **c** does not change.

A. Estimation of B

Denote $C_p = \{i : c_i = p\}$, $p = 1, 2, \ldots, K$. Here C_p serves as the group set, *i.e.* C_p is the set of nodes which belong to group p. For any group pair (p, q), any pair nodes $(i, j) : c_i = p, c_j = q$ equally contribute to the log likelihood function defined in Eq. (3). Thus we can rewrite Eq. (3) in terms of group index p, q instead of node index i, j as:

$$L(\mathbf{B}) = \sum_{pq} s_{pq} \log \mathbf{B}_{pq} + (n_p n_q - s_{pq}) \log(1 - \mathbf{B}_{pq}), \quad (4)$$

where n_p and n_q are the cardinalities of sets C_p and C_q, respectively, and $s_{pq} = \sum_{k=1}^{m} \sum_{i \in C_p j \in C_q} \mathbf{G}_{ij}^k$ is the total number of edges between group p and q (cross-cut between the two groups). Thus,

$$\frac{\partial L(\mathbf{B})}{\mathbf{B}_{pq}} = \frac{s_{pq}}{\mathbf{B}_{pq}} + \frac{n_p n_q - s_{pq}}{1 - \mathbf{B}_{pq}}. \quad (5)$$

We set

$$\frac{\partial L(\mathbf{B})}{\mathbf{B}_{pq}} = = 0, \quad (6)$$

and get the estimation

$$\frac{s_{pq}}{\mathbf{B}_{pq}} + \frac{n_p n_q - s_{pq}}{1 - \mathbf{B}_{pq}} = 0, \quad (7)$$

or

$$\widehat{\mathbf{B}}_{pq} = \frac{s_{pq}}{n_p n_q}. \quad (8)$$

Since s_{pq} is the total number of edges between group p and q, $s_{pq} \geq 0$ and $s_{pq} \leq n_p n_q$. Thus $0 \leq \frac{s_{pq}}{n_p n_q} \leq 1$, indicating the constraint in Eq. (3) is automatically satisfied. One can easily see that this solution is equivalent to solve maximum likelihood estimation along all the Bernouli distributions over groups p and q independently. This is similar with the estimation of $\mathbf{B}$ in [8].

B. Estimation of Indicator Vector c

One of the advantages of the PMGD model is that it reduces the number of latent valuables from m^2 to n, thus the estimation of the membership indicators is dramatically simplified. As inspired from on-line updating algorithm of K-means method, we solve the indicator one node by one node. For Eq. (3), considering node i, we rewrite the likelihood as a function of $c_i = t$:

$$L^i(t) = \sum_{j \neq i} \log \frac{\mathbf{B}_{tc_j}}{1 - \mathbf{B}_{tc_j}} + \sum_{k=1}^{m} \mathbf{G}_{it}^k \log \frac{\mathbf{B}_{tt}}{1 - \mathbf{B}_{tt}} + \sum_{j \neq i} \log(1 - \mathbf{B}_{tc_j}) + \log(1 - \mathbf{B}_{tt}). \quad (9)$$

By denoting a $n \times K$ matrix $\mathbf{U}_{it} = L^i(t)$, we have the maximum likelihood estimation of node i:

$$\hat{c}_i = \arg \max_t \mathbf{U}_{it}. \quad (10)$$

C. PMGD Algorithm

We summarize the algorithm of our new PMGD Algorithm as follows:

D. Consistent Network Module Recovery

In the previous model, we can interpret $\mathbf{B}_{tt}$ as the cliqueness among the objects in the block t, $t = 1, 2, \cdots, K$. Thus if $\mathbf{B}_{tt}$ is high, we consider the t-th block is a module. In our study, we use 0.5 as a threshold,, *i.e.* if $\mathbf{B}_{tt} \geq 0.5$ we consider the t-th block is a common module. We

can also see the connectivity of block t and s from $\mathbf{B}_{st}$, which will be discussed in the experimental section later.

IV. Theoretical Analysis of Probabilistic Graph Decomposition

Here we explore the relationship between Probabilistic Graph Decomposition and Ratio Cut spectral clustering. The Ratio Cut objective [12] is defined as following:

$$J_{\mathbf{rc}}(\mathbf{c}) = \sum_{p \neq q} \frac{s_{qp}}{n_p} + \frac{s_{qp}}{n_q}, \quad (11)$$

where s_{pq}, n_p, n_q are the cross cuts, and cardinality of group p and q, respectively.

Algorithm 1

The proposed PMGD algorithm.

Input: Brain connectivity networks $\mathbb{G}$, the number of groups K
Output: Clustering indicator vector $\mathbf{c}$
Initialize $\mathbf{c}$ with $\mathbf{c}_0$: $\mathbf{c} \leftarrow \mathbf{c}_0$ **while true do**
  foreach p, q **do**
    $\mathbf{B}_{pq} \leftarrow \frac{s_{pq}}{n_p n_q}$, as defined in Eq. (8).
  end
  foreach $i = 1, 2, \cdots, n$ **do**
    Construct $U_{it} = L^i(t)$ as defined in Eq. (9)
    $\mathbf{c}_i \leftarrow \arg\max_t U_{it}$
  end
  if $\mathbf{c}$ *never changes* **then**
    break.
  end
end

In this section, we show that this objective function is an approximation of log likelihood of Probabilistic Graph Decomposition, with a negative coefficient.

A. Diagonal of B

From Eq. (8), we have

$$\sum_k B_{kk} = \sum_k \frac{s_{kk}}{n_k^2} \approx \frac{K^2 s_{kk}}{n}. \quad (12)$$

If we assume the data is balanced, *i.e.* each group has close number of nodes, we have $n_k \approx n/K$.

$$\sum_k B_{kk} \approx \frac{K^3}{n^2} s_{kk} \quad \text{and} \quad J_{\mathbf{rc}} \approx \sum_{p \neq q} \frac{s_{pq}}{\frac{n}{K}} = \sum_{p \neq q} \frac{K}{n} s_{pq}. \quad (13)$$

Thus

$$\frac{K^2}{n} J_{\mathbf{rc}} + \sum_k B_{kk} = \frac{K^3}{n^2} \sum_{pq} s_{pq} = \frac{K^3}{n^2} E, \quad (14)$$

or

$$J_{\mathbf{rc}} = \frac{KE}{n} - \frac{n \sum_k B_{kk}}{K^2}, \quad (15)$$

where E is the number of edges in the graph. Eq. (15) indicates that minimizing the Ratio Cut objective is equivalent to maximizing the diagonal of B.

B. Ratio Cut versus Probabilistic Graph Decomposition

By substituting Eq. (8) into Eq. (4), we get

$$
\begin{aligned}
L(\mathbf{c}) &= \sum_{pq} s_{pq}\log s_{pq} - s_{pq}\log n_p n_q + (n_p n_q - s_{pq})\log(1 - \frac{s_{pq}}{n_p n_q}) \quad (16)\\
&= \sum_{pq} n_p n_q \log\frac{n_p n_q - s_{pq}}{n_p n_q} - s_{pq}\left\{\log\frac{(n_p n_q)^2}{n_p n_q - s_{pq}} - \log s_{pq}\right\}\\
&\approx \sum_{pq} \frac{n^2}{K^2}\log\frac{n^2 - K^2 s_{pq}}{n^2} - s_{pq}\left\{\log\frac{(n_p n_q)^2}{n_p n_q - s_{pq}} - \log s_{pq}\right\}\\
&\approx \alpha - \beta J_{\mathbf{c}},
\end{aligned}
$$

where

$$\alpha = \sum_{pq}\frac{n^2}{K^2}\log\frac{n^2 - K^2 s_{pq}}{n^2} + \sum_k s_{kk}\left\{\log\frac{(n)^4}{s_{kk}(n_k^2 - s_{kk})}\right\},$$

and

$$\beta = \frac{n}{K}\log\frac{n^2}{K^2}.$$

Here we assume that the off-diagonal of B is relative small compared to the diagonal. This is supported by the argument in [8] and our experimental results (see experiment section for more details).

In order to verify Eq. (16), we use four datasets AT&T, JAFFE, PIE, and YALEB to construct the relational graph, (see experiment section for more details) and we randomly generate membership indicators around the solution of our algorithm $\mathbf{c}^*$, *i.e.* randomly pick some position of $\mathbf{c}^*$ and randomly assign the node to other clusters. In Figure 2, we plot $J_{\mathbf{rc}}(\mathbf{c})$ versus the log likelihood $L(\hat{B}, \mathbf{c})$ with randomly generated membership indicator $\mathbf{c}$ (red dots) and the the curve $A - BJ_{\mathbf{c}}$ (blue solid line). Here $\hat{B}$ is computed by Eq. (8). From the figures, we can see that in most cases, the true log likelihood of Probabilistic Graph Decomposition is linear to the Ratio Cut objective with a coefficient $-\frac{n}{K}\log\frac{n^2}{K^2}$.

V. Construction of Structural Brain Connectivity Networks

In this section, we will describe how did we collect and construct the human connectome, which are used to identify the consistent network modules. In our project, participants included 24 healthy young male adults (age: 24.0 ± 3.2) with no history of neurological or psychiatric disorder. The MRI scans were acquired on a Siemens 3T TIM Trio (Erlangen, Germany) using a 12-channel receive only phased array head coil in combination with a

body coil for radio frequency transmission. A SE-EPI DTI sequence was applied using parameters: matrix= 128 × 128; FOV= 256 × 256mm; TE/TR=77/8300 ms; 68 transversal slices with 2mm thickness; 48 diffusion directions with gradients b=1000s/mm2, and 8 samplings at b=0. Each session also included a high resolution T1-weighted MP-RAGE imaging as anatomical reference for subsequent parcellation and co-registration. Our processing pipeline includes three major steps: (1) DTI tractography, (2) ROI generation from T1-weighted MRI (MP-RAGE or SPGR), and (3) connectivity network construction.

The DTI data are analyzed in FSL[1]. DTI preprocessing includes correction for motion and eddy current effects in DTI images. The processed DTI images are then output to Diffusion Toolkit (http://trackvis.org/) for fiber tracking, using the streamline tractography algorithm called FACT (fiber assignment by continuous tracking). The FACT algorithm initializes tracks from many seed points and propagates these tracks along the vector of the largest principle axis within each voxel until certain termination criteria are met. In our study, stop angle threshold is set to 35 degree, which means if the angle change between two voxels is greater than 35 degree, the tracking process stops. A spline filtering is then applied to smooth the tracks.

Anatomical parcellation is performed using FreeSurfer 5.1 [13], [14], [15] on the high-resolution T1-weighted anatomical MRI scan acquired with MP-RAGE sequence. The parcellation is an automated operation on each subject to obtain 68 gyral-based ROIs, with 34 cortical ROIs in each hemisphere. The T1-weighted MRI image is registered to the low resolution b0 image of DTI data using the FLIRT toolbox in FSL, and the warping parameters are applied to the ROIs so that a new set of ROIs in the DTI image space are created. These new ROIs are used for constructing the structural network.

The topological representation of a network is a collection of nodes and edges between pairs of nodes. In constructing the weighted, undirected network, the nodes are chosen to be the 68 registered ROIs obtained from FreeSurfer parcellation. The weight of the edge is defined as the density of the fibers connecting a pair of nodes, which is the number of tracks between two ROIs divided by the mean volume of the two ROIs [16], [17]. A fiber is considered to connect two ROIs if and only if its end points fall in the two ROIs respectively. The weighted network can be represented by a matrix. The rows and columns correspond to the nodes, and the elements of the matrix correspond to the weights.

VI. Experimental Results

In our experiments, we first perform the proposed PMGD method on the connectivity networks to identify the consistent modules. Because the biomedical image application lacks the ground truth, we also evaluate the proposed model using the image clustering tasks on four human face benchmark data sets.

A. Brain Connectivity Network Module Finding

We employ the PMGD model on the connectivity network data described in the above section. Notice that there is only one hyper-parameter (the number of groups K, which is set to be 8 in all experiments) in our model no parameters in the its optimization algorithm. We

first compare three connectivity networks measurements: fiber number (FN), fiber length (LL), and the weighted network (W) [6]. Since we have no ground truth for the consistent modules of the human brain structure modules, we test the quality of connectivity measurements by comparing with the random background. For each connectivity measurement (FN, LL, or W), we run our algorithm and obtain consistent network modules (we use 0.01 as the threshold to discretize the weighted graph). Then we randomly permute all graph. With the random permuted graph, we apply our algorithm again. Presumably, results on such random graph should be very poor and thus serve as a background to compare.

We compare the significance of the differences between each connectivity graph and its permutation and use the P-value to measure the quality of the connectivity graph (100 permutations). We show the results in Table I. For the weighted, fiber number, and fiber length networks, we discover 4, 6, and 7 consistent network modules, respectively. We show the P-values for the all the consistent network modules for weighted and the first 4 for FN and LL (sorted by the cliqueness $\mathbf{B}_{tt}$). One can observe from Table I that our method performs much better than the background for the weighted network. For example, the significance of difference between the first and the second consistent network modules and the random background is 2.14×10^{-54} and 7.25×10^{-53}, respectively, while other networks achieve much lower level of significance. We also visualize the locations of the consistent modules in Figure 4 in top, bottom, left and right views.

We visualize the weighted connectivity network and highlight the consistent modules discovered by PMGD algorithm in Figure 3. Module 1 includes 5 ROIs: **RINS**, **RPOC**, **RST**, **RSMG**,, and **RTRT**. Module 2 includes **LINS**, **LPOC**, **LST**, **LSMG**, and **LTRT**. They are symmetric. Module 3 includes 7 ROIS **LCMF**, **LLOF**, **LPCS**, **LPOB**, **LPAG**, **LPRC**, and **LRMF**. Module 4 includes 10 ROIs: **LCNS**, **LISC**, **LLIN**, **LPEC**, **LPCN**, **RCNS**, **RISC**, **RLIN**, **RPEC**, and **RPCN**. Module 4 itself is symmetric.

Since the result of block matrix B represents the cliqueness within modules and the connectivity between modules, we are also interested in the representation capability of block matrix B, which is show in **(a)** in Figure 5. In (**b**), we demonstrate how consistently they ROIs belong to the corresponding consistent network modules among the 24 subjects. One can observe that **LPOC, LST, LSMG, LTRT**, **RPOC, RST, RSMG, RTRT**, **LCNS, LPCN,** and **RPCN** are always consistent among the 24 subjects with high probability (≥ 0.99) of belonging to the corresponding consistent network modules.

B. Evaluations Using Clustering Benchmark Data

We also evaluate our model using the human face benchmark data sets. Because existing clustering methods are usually for single data graph clustering, to compare different methods, we apply the probabilistic graph decomposition method in §2.3 to data graph and perform clustering task. Using this set up, we can compare the clustering performance of our new model with other related clustering approaches.

We use ten benchmark data sets to evaluate the performance of the proposed model and algorithm, including five image data sets: AT&T [2], JAFFE (The Japanese Female Facial

Expression)[3], CMU PIE (Face Pose, Illumination, and Expression) [18], YALEB[4], MNIST; and five data sets: Zoo, Wine, Iris, Soybean, Dermatology, from UCI machine learning repository[5].

In all experiments, we construct the pairwise relation using K-nearest neighbor graph, *i.e.* $\mathbf{G}_{ij} = 1$ if node i is the K-nearest neighbor of j or node j is the K-nearest neighbor of i, $\mathbf{G}_{ij} = 0$ otherwise. We set $K = 5$ in all the experiments.

We first compare our algorithm with standard K-means [19] and spectral clustering (normalized cut) [20], [21]. For K-means, we use the gray level values of pixels as feature, and the distance and means are taken in Euclidean space. For spectral clustering, we first calculate the eigenvector of graph Laplacian $L = D - G$, where $D = \mathbf{diag}(d_1, d_2, \ldots, d_n)$ and $d_i = \Sigma_j G_{ij}$. And a standard K-means algorithm is applied on the K eigenvectors associated with the K least eigenvalues of graph Laplacian L. Here K is set to the number of person according to ground truth.

The clustering accuracy is computed as follows. Suppose we have $N = n_1 + n_2 + \cdots + n_K$ data objects (n_1 are known/observed to belong to class F_1, etc.). They are clustered into K clusters. with $m_k = |C_k|$. This forms a contingency table $T = (T_{kl})$, where T_{kl} denotes the number of objects from class F_k and have been clustered into cluster C_l. Clearly, $\Sigma_l T_{kl} = n_k$ and $\Sigma_k T_{kl} = m_l$. The clustering accuracy is the percentage of objects been correctly clustered: $\rho = \Sigma_k T_{kk}/N$. In practice, matching F_k to C_l is obtained by running the Hungarian algorithm for the optimal bipartite matching.

Since the results of all clustering methods depend on the initializations, we run multiple random trials to approximate the optimal results. We perform 128 trials for all three methods with random initializations. We evaluate the performance as following. Define **Best**(N) to be the highest accuracy among N random trials for all four approaches. Clearly, **Best**(N) improves as we increase N. We plot **Best**(N) versus N in Figure 6 for all four data sets. In experimental results, our method consistently outperforms other related methods. For example, in JAFFE data set, the best our clustering accuracy is quite close to 1, about 95%, which is far better than the other approaches. Please not ice that these approaches are totally unsupervised.

In Figure 7, we visualize the block structures found by our algorithm on JAFFE data set (other data sets have more clusters and cannot be clearly plotted). We plot the faces according the membership indicators **c** of the solution of probabilistic graph decomposition. The edges within the groups are also plotted. We can see that the structure we find is consistent with human understanding.

Soft Membership Indicator—One of the advantage of the Probabilistic Graph Decomposition is the soft clustering capability.

[2] http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
[3] http://www.kasrl.org/jaffe.html
[4] http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html
[5] http://archive.ics.uci.edu/ml/

The probability of node i belonging to cluster k (soft membership indicator) is calculated as following:

$$P(i,k) = \frac{e^{U_{ik}}}{\sum_k e^{U_{ik}}}, \quad (17)$$

where U is defined near Eq. (9).

Here we select four groups from each face image data set (other non-image data cannot be meaningfully visualized) and visualize the data with membership probabilities in Figure 8.

In Figure 8, we plot the original faces and the corresponding soft membership above each them. Notice that the selected images are sorted by the probability obtained by the our algorithm. One can see that the soft membership indicator is consistent with human perception. For example, in the first group in AT&T dataset (left top panel of Figure 8), even though two last 6 faces come from different person, they are visually similar with the first 9 face images. And for the fist 9 images, they have a high probability to belong to the group.

VII. Conclusion

In this paper, we proposed a novel brain connectivity network analysis method by employing the new probabilistic multi-graph decomposition model to identify the consistent network modules (common pseudo-cliques) cross multiple brain connectivity networks, which are potentially associated to cognitive functions of humans. We first proposed a new probabilistic graph decomposition method to reduce the high computational complexity which appears in previous stochastic block models. After that, we introduced the probabilistic multi-graph decomposition model to solve the multi-graph problem. Meanwhile, we derived an efficient optimization algorithm to solve the proposed objective and estimate the model parameters. The real DTI data were used to construct the brain connectivity networks to validate our methods. We also evaluate the proposed models via the humane face image clustering task on benchmark data.

Acknowledgments

This work was partially supported by NSF-IIS 1302675, NSF-IIS 1344152, NSF-DBI 1356628, NSF-IIS 1619308, NSF-IIS 1633753, NIH AG049371.

References

1. Catani M, Howard R, Pajevic S, Jones D. Virtual in vivo interactive dissection of white matter fasciculi in the human brain. Neuroimage. 2002; 17(1):77–94. [PubMed: 12482069]
2. Ciccarelli O, Toosy A, Parker G, Wheeler-Kingshott C, Barker G, Miller D, Thompson A. Diffusion tractography based group mapping of major white-matter pathways in the human brain. Neuroimage. 2003; 19(4):1545–1555. [PubMed: 12948710]
3. Sporns O, Tononi G, Kotter R. The human connectome: a structural description of the human brain. PLoS Comput Biol. 2005; 1(4):e42. [PubMed: 16201007]
4. Bassett D, Bullmore E. Small-world brain networks. Neuroscientist. 2006; 18:512–23.

5. Bassett D, Greenfield D, Meyer-Lindenberg A, Weinberger D, Moore S, Bullmore E. Efficient physical embedding of topologically complex information processing networks in brains and computer circuits. PLoS Comput Biol. 2010; 6:e1000748. [PubMed: 20421990]
6. Rubinov M, Sporns O. Complex network measures of brain connectivity: uses and interpretations. Neuroimage. 2010; 52:1059–69. [PubMed: 19819337]
7. O'Donnell L, Kubicki M, Shenton M, Dreusicke M, Grimson W, Westin C. A method for clustering white matter fiber tracts. American Journal of Neuroradiology. 2006; 27(5):1032–1036. [PubMed: 16687538]
8. Airoldi EM, Blei DM, Fienberg SE, Xing EP. Mixed membership stochastic blockmodels. Journal of Machine Learning Research. 2008; 9:1981–2014. [PubMed: 21701698]
9. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. Journal of Machine Learning Research. 2003; 3:993–1022.
10. Li FF, Perona P. A bayesian hierarchical model for learning natural scene categories. CVPR. 2005:II: 524–531.
11. Jordan MI, Ghahramani Z, Jaakkola T, Saul LK. An introduction to variational methods for graphical models. Machine Learning. 1999; 37(2):183–233.
12. Chan, PK., Schlag, MDF., Zien, JY. Spectral K -way ratio-cut partitioning and clustering; DAC. 1993. p. 749-754.[Online]. Available: http://doi.acm.org/10.1145/157485.165117
13. Fischl B, Sereno M, Dale A. Cortical surface-based analysis. ii: Inflation, flattening, and a surface-based coordinate system. Neuroimage. 1999; 9(2):195–207. [PubMed: 9931269]
14. Dale A, Fischl B, Sereno M. Cortical surface-based analysis. i. segmentation and surface reconstruction. Neuroimage. 1999; 9(2):179–94. [PubMed: 9931268]
15. Fischl B, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron. 2002; 33(3):341–55. [PubMed: 11832223]
16. Hagmann P, Kurant M, Gigandet X, Thiran P, Wedeen VJ, Meuli R, Thiran JP. Mapping human whole-brain structural networks with diffusion MRI. PLoS One. 2007; 2(7):e597. [PubMed: 17611629]
17. Cheng H, et al. Optimization of seed density in dti tractography for structural networks. J Neurosci Methods. 2012; 203(1):264–72. [PubMed: 21978486]
18. Bsat, M., Baker, S., Sim, T. The CMU pose, illumination, and expression (PIE) database of human faces. CMU Robotics Institute; 2001.
19. MacQueen, JB. Some methods for classification and analysis of multivariate observations. Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability; Berkeley: University of California Press; 1967. p. 281-297.
20. Ng AY, Jordan MI, Weiss Y. On spectral clustering: Analysis and an algorithm. Advances in Neural Information Processing Systems (NIPS). 2002; 14:849–856.
21. Shi J, Malik J. Normalized cuts and image segmentation. IEEE Transactions on PAMI. 2000; 22(8):888–905.

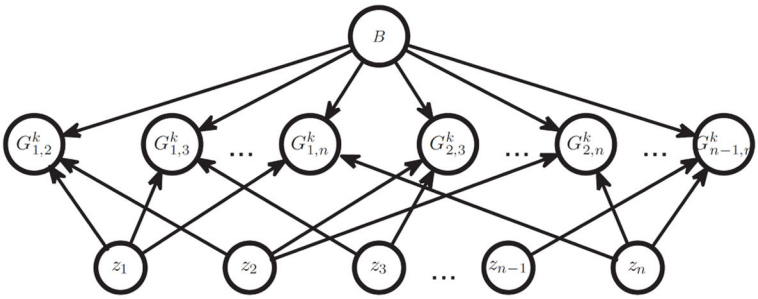
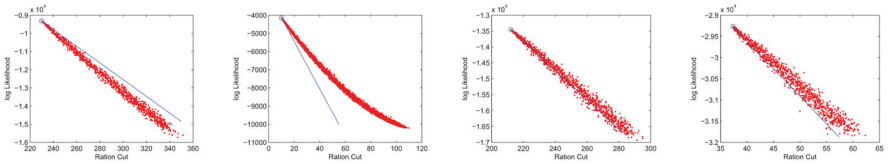


Fig. 1.

A graphical dependency diagram of the proposed Probabilistic Multi-Graph Decomposition (PMGD) model.



(a) AT&T (b) JAFFE (c) PIE (d) YALEB

Fig. 2.

The relationship between Ratio Cut objective and log likelihood of Probabilistic Graph Decomposition on 4 datasets. Blue circles are for the solutions $\hat{\mathbf{c}}$ of our method.

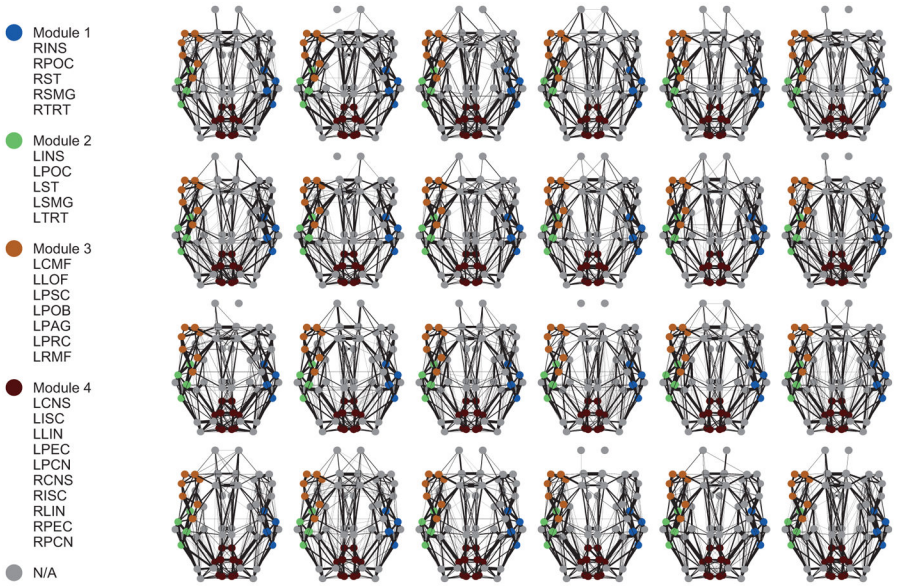


Fig. 3.

The weighted connectivity networks of 24 young males and their consistent network modules discovered by PMGD algorithm. There are 5, 5, 7, and 10 ROIs in Modules 1, 2, 3, and 4 respectively. Module 1 is symmetric with Module 2 and Module 4 itself is symmetric.

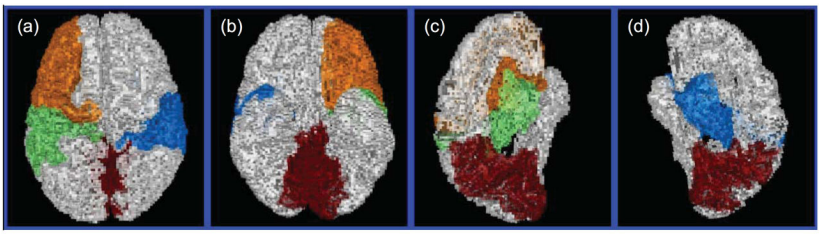


Fig. 4.

Location visualization of 4 consistent network modules discovered by PMGD model from (a) top, (b) bottom, (c) right, and (d) left views.

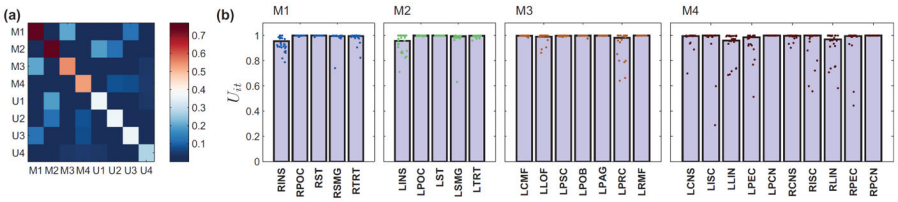


Fig. 5.

The optimization results for PMGD algorithm on the networks of 24 connectivity young males. (**a**): the visualization of block matrix **B**. M1, M2, M3, and M4 are four modules discovered by our method. (**b**): The probability ($\mathbf{U}_{it}$) of each ROI i belonging to the corresponding consistent network module (t) for M1, M2, M3, and M4.

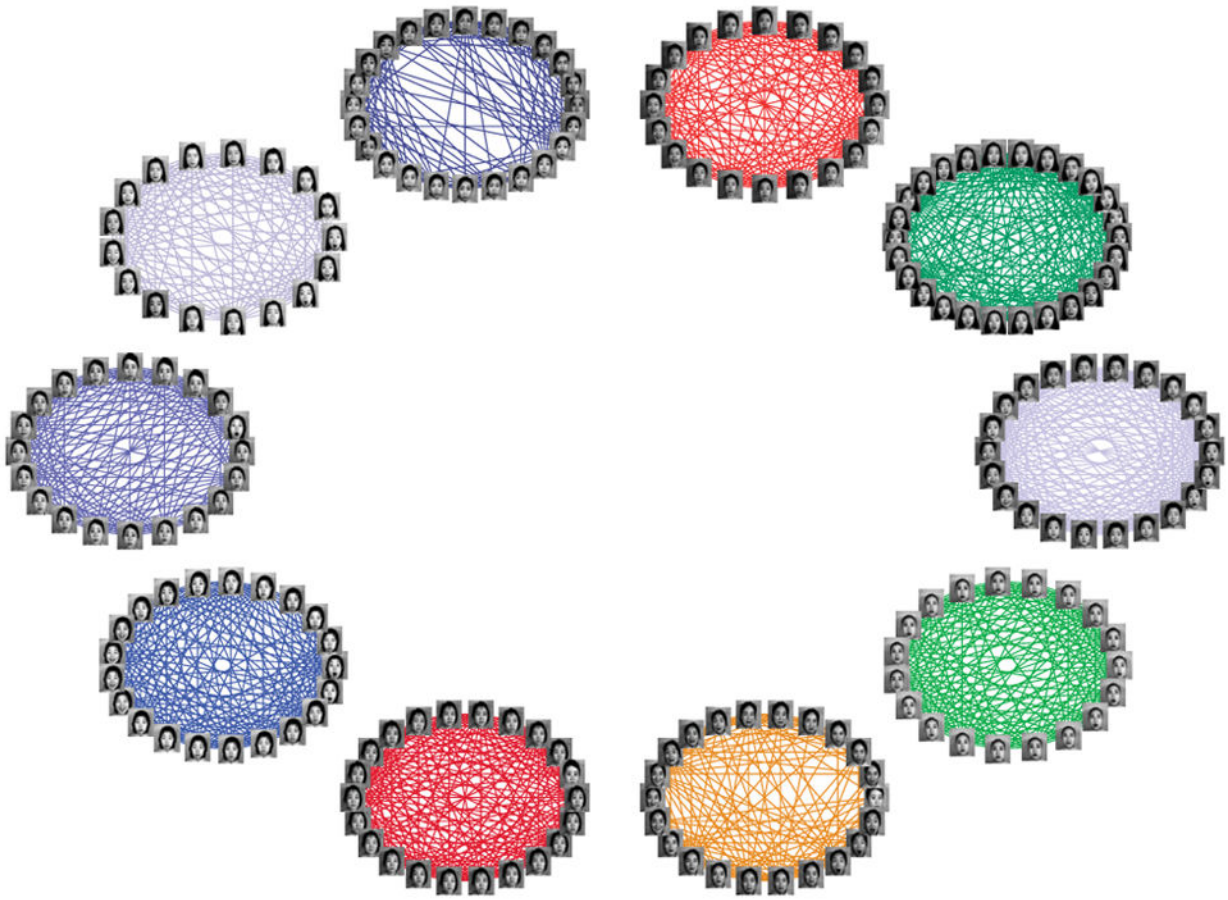


Fig. 6.

Clustering accuracy and Ratio Cut objective on ten data sets. For Ratio Cut objective, the lower the better.

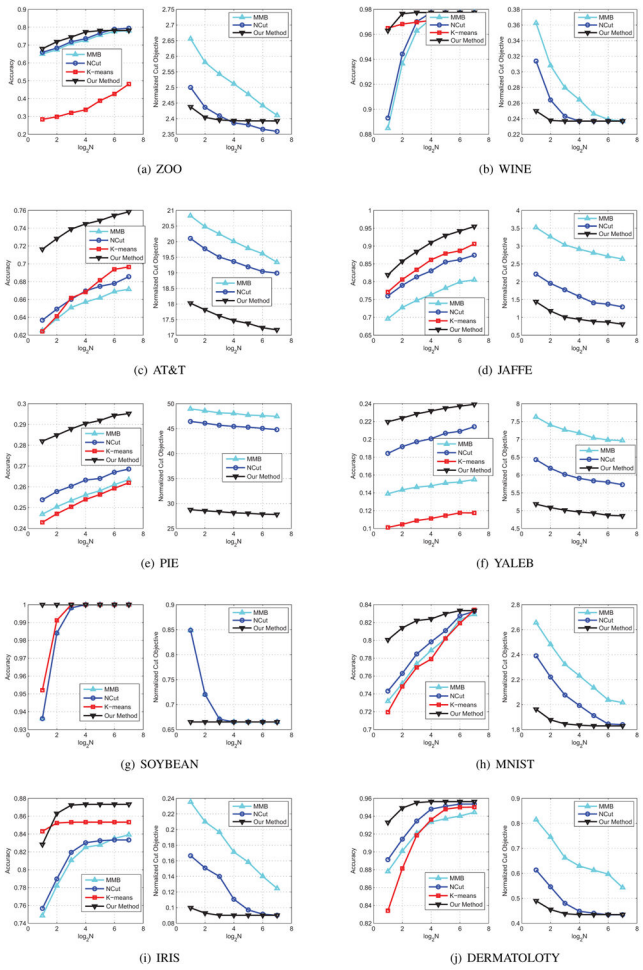


Fig. 7.

Block structures found by our method on JAFFE. Our probabilistic graph decomposition model can correctly identify categories of different face images.

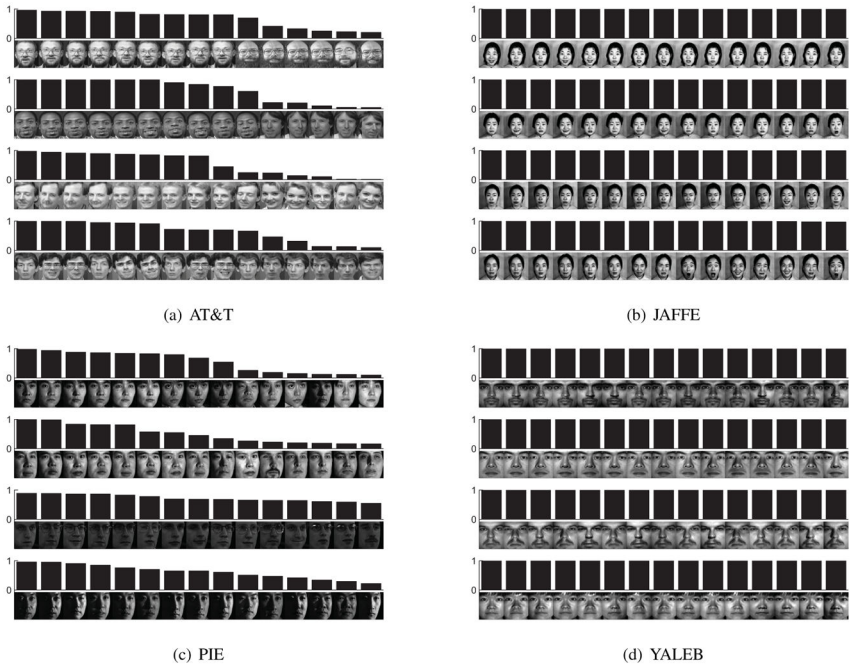


Fig. 8.

Probability output of our method for **AT&T, JAFFE, PIE** and **YALEB** data. For each class, the first row is the probability of a face belong the class, and the image below it is the corresponding face image.

TABLE I

A permutation test of our method on weighted, fiber number, and fiber length connectivity.

	Weighted		Fiber Number		Fiber Length	
Module	B_{tt}	**P-value**	B_{tt}	**P-value**	B_{tt}	**P-value**
1	0.7692	2.14×10^{-54}	0.7306	2.87×10^{-25}	0.7778	1.19×10^{-4}
2	0.7569	7.25×10^{-53}	0.6781	3.52×10^{-12}	0.6427	6.28×10^{-4}
3	0.5490	1.21×10^{-7}	0.6560	1.69×10^{-6}	0.5616	4.11×10^{-3}
4	0.5289	7.90×10^{-5}	0.6420	1.41×10^{-4}	0.5566	7.21×10^{-2}

Proc IEEE Int Conf Data Min. Author manuscript; available in PMC 2018 April 23.