Multi-graph Clustering Based on Interior-Node Topology with Applications to Brain Networks

Guixiang Ma[1], Lifang He[2(✉)], Bokai Cao[1], Jiawei Zhang[1], Philip S. Yu[1,3], and Ann B. Ragin[4]

[1] Department of Computer Science, University of Illinois at Chicago, Chicago, IL, USA
{gma4,caobokai,jzhan9,psyu}@uic.edu
[2] School of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China
lifanghescut@gmail.com
[3] Institute for Data Science, Tsinghua University, Beijing, China
[4] Department of Radiology, Northwestern University, Chicago, IL, USA
ann-ragin@northwestern.edu

Abstract. Learning from graph data has been attracting much attention recently due to its importance in many scientific applications, where objects are represented as graphs. In this paper, we study the problem of multi-graph clustering (*i.e.*, clustering multiple graphs). We propose a multi-graph clustering approach (MGCT) based on the interior-node topology of graphs. Specifically, we extract the interior-node topological structure of each graph through a sparsity-inducing interior-node clustering. We merge the interior-node clustering stage and the multi-graph clustering stage into a unified iterative framework, where the multi-graph clustering will influence the interior-node clustering and the updated interior-node clustering results will be further exerted on multi-graph clustering. We apply MGCT on two real brain network data sets (*i.e.*, ADHD and HIV). Experimental results demonstrate the superior performance of the proposed model on multi-graph clustering.

Keywords: Multi-graph clustering · Interior-node topology · Brain network

1 Introduction

In recent years, graph mining has been a popular research area because of numerous applications in social network analysis, computational biology and computer networking. In addition, many new kinds of data can be represented as graphs. For example, from common brain images such as the functional magnetic resonance imaging (fMRI) data of multiple subjects, we can construct a brain connectivity network for each of them, where each node represents a brain region, and each link represents the functional/structural connectivity between two brain regions [12]. These multiple brain networks provide us with an unprecedented

P. Frasconi et al. (Eds.): ECML PKDD 2016, Part I, LNAI 9851, pp. 476–492, 2016.
DOI: 10.1007/978-3-319-46128-1_30

opportunity to explore the inner structure and activity of the human brain, serving as valuable supportive information for clinical diagnosis of neurological disorders [18]. Therefore, mining on graphs becomes a crucial task and may benefit various real-world applications.

Among the existing works on graph learning, quite a few of them fall into supervised learning, which usually aim to select frequent substructures such as connected subgraph patterns in a database of graphs and then feed these subgraph features into classifiers [6,11]. These methods typically work well when the graph database is very large or the access to side information is assumed. However, the number of subgraphs is exponential to the size of graphs, thus the subgraph enumeration process is both time and memory consuming which makes it infeasible to explore the complete subgraph space. Moreover, in many real-world cases, only a small number of labeled graphs are available. Therefore, finding discriminative subgraph patterns from a large number of candidate patterns based on such limited instances is not reliable. While supervised methods focus on training classifiers, unsupervised clustering could provide exploratory techniques for finding hidden patterns in multiple graphs. In this paper, we investigate the unsupervised scenarios by exploring the multi-graph clustering based on the interior-node topology of graphs. Topology is the mathematics of neighborhood relationships in space, which is independent of the distance metric, thus the interior-node topology of graphs could provide complementary local structure information for the original linkage, which can only characterize the global structure information of graph. Despite its value and significance, to our best knowledge, the interior-node topology of graphs has not been studied in the problem of multi-graph clustering so far. There are two major challenges in this multi-graph clustering problem:

- How to capture the interior-node topology of each graph? Conventional approaches extract graph-theoretical measures, *e.g.*, clustering coefficients, to quantify the prevalence of clustered connectivity [10,23]. However, assigning a predefined measure to specific nodes in a graph might not fully characterize the subtle local topological structure of the graph.
- How to effectively leverage the extracted topological structure information to facilitate the process of multi-graph clustering? The original linkage metric describes the global connectivity structure in the graph, while the topological structure depicts the local neighborhood relationships. An effective multi-graph clustering model should fuse these two complementary structural information together.

To address the above challenges, we propose a framework of multi-graph clustering with interior-node topology. The contributions of this work are twofold:

- We propose to consider both the global structure and the local topological structure of graphs for the multi-graph clustering task. Specifically, we utilize interior-node clustering to capture local topological structure of graphs.
- Considering the fact that graphs with a high similarity tend to have a similar interior-node topology, we propose to merge the multi-graph clustering stage

and interior-node clustering process into a unified iterative framework called MGCT, where the results of interior-node clustering are exerted on multi-graph clustering and the multi-graph clustering will in turn improve interior-node clustering of each graph, thus achieving a mutual reinforcement.

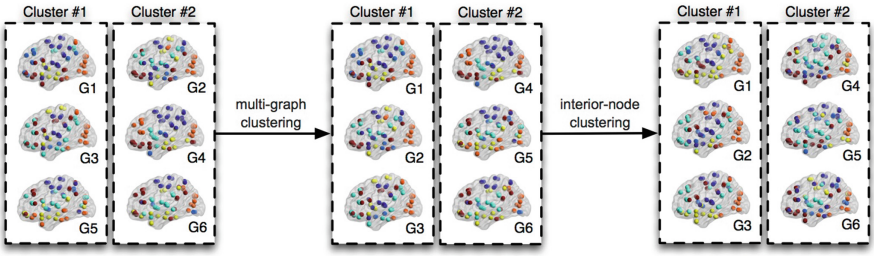


Fig. 1. The framework of the proposed model.

In the scenario of brain network analysis for multiple subjects, the proposed framework of multi-graph clustering can be illustrated with the example shown in Fig. 1. There are two stages in each iteration of the framework: multi-graph clustering and interior-node clustering. In the multi-graph clustering stage, the given six brain networks are clustered into two clusters, and then in the second stage, the interior-node clustering of each graph will be updated with a weighted influence from their neighbor graphs in the same cluster, after which the new interior-node clustering results will be utilized for the multi-graph clustering in the next iteration. After the model converges, we will obtain the final optimal multi-graph clustering results, which can be used for further analysis, for example, the neurological disorder identification.

We evaluate the proposed method on two real brain network data sets (ADHD and HIV). Experimental results illustrate the superior performance of the proposed approach for multi-graph clustering in brain network analysis.

2 Preliminaries

In this section we establish key definitions and notational conventions that simplify the exposition in later sections.

Throughout this paper, matrices are written as boldface capital letters and vectors are denoted as boldface lowercase letters. For a matrix $\mathbf{M} \in \mathbb{R}^{n\times m}$, its elements are denoted by m_{ij}, and its i-th row, j-th column are denoted by $\mathbf{m}^i$, $\mathbf{m}_j$ respectively. The Frobenius norm of $\mathbf{M}$ is defined as $\|\mathbf{M}\|_F = \sqrt{\sum_{i=1}^{n} \|\mathbf{m}^i\|_2^2}$, and the $\ell_{2,1}$ norm of $\mathbf{M}$ is defined as $\|\mathbf{M}\|_{2,1} = \sum_{i=1}^{n} \left\|\mathbf{m}^i\right\|_2$. For any vector $\mathbf{u} \in \mathbb{R}^n$, $Diag(\mathbf{u}) \in \mathbb{R}^{n\times n}$ is the diagonal matrix whose diagonal elements are u_i. $\mathbf{I}_n$ denotes an identity matrix with size n. $\|\mathbf{u}\|_0$ is the ℓ_0 norm, which counts the number of nonzero elements in the vector $\mathbf{u}$.

Definition 1 (Multi-graph Clustering). An undirected graph can be formally represented as $G = (V, E, \mathbf{A})$, where V is the set of vertices, $E \subset V \times V$ is the set of edges, and $\mathbf{A}$ is the weighted affinity matrix whose entry denotes the affinity between a pair of nodes. Given a set of such graphs $D = \{G_1, G_2, \cdots, G_n\}$, the goal of multi-graph clustering is to cluster the graphs in D into c subsets.

Definition 2 (Interior-node Clustering). Given an undirected graph $G = (V, E, \mathbf{A})$, the goal of interior-node clustering is to group the nodes of the graph into k clusters $C = \{C_1, \cdots, C_k\}$, with $V = C_1 \cup \cdots \cup C_k$ and $C_i \cap C_j = \varnothing$ for every pair i, j with $i \neq j$.

Definition 3 (Topology). Topology is the mathematics of neighborhood relationships in space independent of metric distance. In the context of graph structures, such neighborhood relationships often correspond to the *connectivity* of nodes, *i.e.,* how nodes are connected to each other.

3 Methodology

In this section, we first introduce the proposed multi-graph clustering framework MGCT, where we formulate the multi-graph clustering stage and the interior-node clustering stage, both of which can be formulated as optimization problems. We then present an iterative algorithm based on half-quadratic optimization to solve this minimization problem.

3.1 An Iterative Framework: MGCT

In the literature of multi-graph clustering, the pairwise distance is mainly measured based on the structure of each graph, and graphs with highly similar structures tend to be clustered into the same group. In other words, the graphs that are clustered into the same group tend to have highly similar topological structure [3]. Following these observations, we propose an iterative framework called MGCT for multiple-graph clustering based on interior-node topology. In each iteration, there are two stages: the interior-node clustering and the multi-graph clustering, where the interior-node clustering results which imply local topological structure are used together with the global structure of graph for clustering multiple graphs, and then the multi-graph clustering results will be utilized in turn to improve the interior-node clustering. Through this iterative mutual reinforcement of interior-node clustering and multi-graph clustering, we can finally achieve a refined multi-graph clustering result.

Multi-graph Clustering. In this part, we focus on the formulation of the multi-graph clustering stage. Since the multi-graph clustering and interior-node clustering depend on each other and are performed alternatively, here we assume we have obtained the interior-node clustering results of the graphs, which can be used for the multi-graph clustering. The formulation of the interior-node clustering problem and the overall iterative process will be discussed later.

Given a set of graphs $D = \{G_1, G_2, \cdots, G_n\}$, with the corresponding set of affinity matrices $A = \{\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_n\}$, where $\mathbf{A}_i \in \mathbb{R}^{m \times m}$ is the weighted affinity matrix of G_i, and its entry denotes the pairwise affinity between nodes in G_i, suppose we have performed interior-node clustering on each of these graphs and obtained a set of clustering indicators $F = \{\mathbf{F}_1, \mathbf{F}_2, \cdots, \mathbf{F}_n\}$, where $\mathbf{F}_i \in \mathbb{R}^{m \times k}$ is the interior-node clustering indicator of G_i, we build a similarity matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$, where s_{ij} denotes the similarity between the two graphs G_i and G_j, and we define it as:

$$s_{ij} = \delta(-\|\mathbf{A}_i - \mathbf{A}_j\|_F^2) + (1-\delta)(-\|\mathbf{F}_i - \mathbf{F}_j\|_F^2) \tag{1}$$

which is a weighted combination of the similarity based on the original affinity matrix of each graph and the similarity based on interior-node clustering results, where δ is the weight parameter balancing the two parts. In this way, the interior-node topology characterized by the interior-node clustering indicator matrix can be incorporated for multi-graph clustering. With this similarity matrix, we can formulate the clustering of graphs in D as a spectral clustering problem, where graphs with a higher pairwise similarity tend to be grouped into the same cluster. Let $\mathbf{H} \in \mathbb{R}^{n \times c}$ be the multi-graph clustering indicator matrix, then the optimal $\mathbf{H}$ can be obtained by solving the following objective function [22]:

$$\begin{aligned} &\min_{\mathbf{H}} \operatorname{Tr}\left(\mathbf{H}^{\mathrm{T}}\mathbf{L}\mathbf{H}\right) \\ &\text{s.t.} \quad \mathbf{H}^{\mathrm{T}}\mathbf{H} = \mathbf{I}_c \end{aligned} \tag{2}$$

where $\mathbf{L} = \mathbf{D}^{-\frac{1}{2}}(\mathbf{D}-\mathbf{S})\mathbf{D}^{-\frac{1}{2}}$ is the symmetric normalized Laplacian matrix, and $\mathbf{D}$ is a diagonal matrix with $d_{ii} = \sum_{j=1}^{n} s_{ij}$.

Interior-Node Clustering. We now study the problem of interior-node clustering of graph in the context of multi-graph clustering.

In graph theory, a cluster is described as a set of nodes more densely connected with each other than with the rest nodes of the graph. Given a graph G with m nodes and the weighted affinity matrix $\mathbf{A} \in \mathbb{R}^{m \times m}$, the goal of interior-node clustering is to group the m nodes into k clusters, *i.e.,* to find a cluster indicator matrix $\mathbf{F} \in \mathbb{R}^{m \times k}$, whose entry indicates which cluster a node may belong to.

Intuitively, nodes with a higher correlation should have a similar cluster indicator. With this assumption, a graph regularization can be embedded to learn the cluster indicator matrix $\mathbf{F}$, which is formulated as the following minimization problem on the basis of the spectral analysis [22]:

$$\begin{aligned} &\min_{\mathbf{F}} \sum_{i,j=1}^{m} a_{ij} \left\| \frac{\mathbf{f}^i}{\sqrt{d_{ii}}} - \frac{\mathbf{f}^j}{\sqrt{d_{jj}}} \right\|_2^2 = \operatorname{Tr}\left(\mathbf{F}^{\mathrm{T}}\mathbf{L}'\mathbf{F}\right) \\ &\text{s.t.} \quad \mathbf{F}^{\mathrm{T}}\mathbf{F} = \mathbf{I}_k \end{aligned} \tag{3}$$

where $\mathbf{L}' = \mathbf{D}'^{-\frac{1}{2}}(\mathbf{D}' - \mathbf{A})\mathbf{D}'^{-\frac{1}{2}}$ is the symmetric normalized Laplacian matrix, and $\mathbf{D}'$ is a diagonal matrix with $d_{ii} = \sum_{j=1}^{m} a_{ij}$.

The above formulation provides a measure of the smoothness of $\mathbf{F}$ over the edges in G. Notice that when a node connects to the nodes in different clusters, it will lead to a relatively large value of $\mathrm{Tr}\left(\mathbf{F}^{\mathrm{T}}\mathbf{L}'\mathbf{F}\right)$ [21]. Therefore, it is expected to identify these boundary-spanning nodes to moderate this influence. In the following, we show how to model and leverage the topology of interior-node to achieve this goal.

From the definition of the topology, we know it is the mathematics of neighborhood relationships in space independent of metric distance. In the context of graph structures, such neighborhood relationships often correspond to the *connectivity* of nodes, *i.e.,* how nodes are connected to each other. In view of the involvement of graph, a naïve approach is that the value of $\mathbf{f}^i$ at every node v_i is the weighted average of $\mathbf{f}^i$ at neighbors of v_i, with the weights being proportional to the edge weights in adjacency matrix $\mathbf{A}$, which can be fitted as

$$\min_{\mathbf{F}} \left\|\mathbf{F} - \mathbf{D}'^{-1}\mathbf{A}\mathbf{F}\right\|_F^2 \tag{4}$$

Since there are some boundary-spanning nodes across clusters, and their neighbors naturally occur in different clusters, to exploit the formulation of (4) on interior-node clustering more effectively, it is crucial for the clustering indicator matrix $\mathbf{F}$ to have discriminative ability for such boundary-spanning nodes, *i.e.*, promoting row-wise sparsity to discriminate relevant boundary-spanning nodes, and thus achieving only characterizing interior nodes. Inspired by [8], we introduce the $\ell_{2,1}$-norm penalty to make it and thus we have the following optimization problem:

$$\begin{aligned} &\min_{\mathbf{F}} \ \mathrm{Tr}\left(\mathbf{F}^{\mathrm{T}}\mathbf{L}'\mathbf{F}\right) + \alpha\left\|\mathbf{F} - \mathbf{D}'^{-1}\mathbf{A}\mathbf{F}\right\|_{2,1} \\ &\text{s.t.} \quad \mathbf{F}^{\mathrm{T}}\mathbf{F} = \mathbf{I}_k \end{aligned} \tag{5}$$

where α is a parameter balancing two terms (*i.e.*, smoothness and sparsity). It can be seen the sparsity-inducing property of $\ell_{2,1}$ norm pushes the clustering indicator matrix $\mathbf{F}$ to be sparse in rows. More specifically, $\mathbf{f}^i$ shrinks to zero if the neighbors of node v_i belongs to different clusters. In particular, the more nodes having neighbors belonging to different clusters, the larger $\left\|\mathbf{f}_i - \mathbf{D}'^{-1}\mathbf{A}\mathbf{f}_i\right\|_2^2$ tends to be, so the value of $\mathbf{f}^i$ gets penalized more harshly. We can thus obtain a better clustering indicator $\mathbf{F}$ for interior nodes.

As we discussed earlier, the graphs clustered into the same group tend to have more similar topological structure, in each iteration of our framework, we hope to further improve the interior-node clustering of each graph by incorporating the interior-node clustering results of its neighbors, *i.e.*, the graphs clustered into the same group by the multi-graph clustering stage of the previous iteration. For two graphs in the same cluster, the closer they are, the more similar interior-node clustering they tend to have. Based on this assumption, for graph G_i, we consider only the graphs that are in the same cluster with G_i, and we aim to infer the weights of influence they should have on G_i.

Suppose we have a set of feature matrices $X = \{\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n\}$, where $\mathbf{X}_i$ can represent both the global and local structure of G_i, we aim to infer a weight matrix $\mathbf{W}$ by solving the following minimization problem:

$$\begin{aligned} \min_{\mathbf{W}} \ & \sum_i \left\| \mathbf{X}_i - \sum_j w_{ij} \mathbf{X}_j \right\|_F^2 \\ \text{s.t.} \ & \sum_j w_{ij} = 1 \end{aligned} \tag{6}$$

where w_{ij} denotes the weight of G_j for G_i, which will be used to control the extent that $\mathbf{F}_j$ will be used to influence $\mathbf{F}_i$ in the next iteration, and G_j can only be a graph from the cluster containing G_i. A larger w_{ij} implies a closer distance between G_i and G_j in the same cluster.

Now we can improve the interior-node clustering of each graph by adding a weighted influence from the neighbour graphs based on the multi-graph clustering. For a graph G_i, the interior-node clustering can be obtained by solving the following objective function extended from Eq. (5):

$$\begin{aligned} \min_{\mathbf{F}_i} \ & \mathrm{Tr}\left(\mathbf{F}_i^{\mathrm{T}} \mathbf{L}_i \mathbf{F}_i\right) + \alpha \left\| \mathbf{F}_i - \mathbf{D}_i^{-1} \mathbf{A}_i \mathbf{F}_i \right\|_{2,1} + \beta \left\| \mathbf{F}_i - \sum_j w_{ij} \mathbf{F}_j \right\|^2 \\ \text{s.t.} \ & \mathbf{F}_i^{\mathrm{T}} \mathbf{F}_i = \mathbf{I}_k \end{aligned} \tag{7}$$

where $\mathbf{A}_i$ is the weighted affinity matrix of G_i, $\mathbf{D}_i$ is the diagonal matrix, and $\mathbf{L}_i$ is the symmetric normalized Laplacian matrix.

With the two stages illustrated above, we can formulate the overall iterative process. We first obtain an initial multi-graph clustering indicator matrix $\mathbf{H}_0$ by Eq. (2), where $\mathbf{S}$ is computed by Eq. (1) with $\delta = 1$. Then we can infer the weight matrix $\mathbf{W}$ by solving (6), which will be used for optimizing the interior-node clustering of each graph in (7). With the resulted $\mathbf{F}_i$ for each graph G_i, a new similarity matrix can be created by Eq. (1), which leads to another iteration of multi-graph clustering by Eq. (2). The overall iterative algorithm with optimization solutions will be discussed in the following section.

3.2 Optimization

Since the minimization problem in Eq. (2) is a typical spectral clustering problem, we can directly solve it by computing the first c generalized eigenvectors of the eigenproblem as illustrated in [20].

To solve the minimization problem (7), we propose an iterative algorithm based on the half-quadratic minimization [16] and the following lemma [9].

Lemma 1. *Let $\phi(.)$ be a function satisfying the conditions: $x \to \phi(x)$ is convex on R; $x \to \phi(\sqrt{x})$ is convex on R_+; $\phi(x) = \phi(-x), \forall x \in R$; $\phi(x)$ is C^1 on R; $\phi'(0^+) \geq 0$, $\lim_{x \to \infty} \phi(x)/x^2 = 0$. Then for a fixed $\|\mathbf{u}^i\|_2$, there exists a dual potential function $\varphi(.)$, such that*

$$\phi(\|\mathbf{u}^i\|_2) = \inf_{p \in R}\{p\|\mathbf{u}^i\|_2^2 + \varphi(p)\} \tag{8}$$

where p is determined by the minimizer function $\varphi(.)$ with respect to $\phi(.)$.

Let $\mathbf{P}_i = \mathbf{F}_i - \mathbf{D}_i^{-1}\mathbf{A}_i\mathbf{F}_i$. According to the analysis for the $\ell_{2,1}$ norm in [9], if we define $\phi(x) = \sqrt{x^2 + \epsilon}$, we can replace $\|\mathbf{P}_i\|_{2,1}$ with $\sum_{j=1}^{n}\phi(\|\mathbf{p}_i^j\|_2)$. Thus, based on Lemma 1, we reformulate the objective function of Eq. (7) as follows:

$$\begin{aligned} &\min_{\mathbf{F}_i} \operatorname{Tr}\left(\mathbf{F}_i^{\mathrm{T}}\mathbf{L}_i\mathbf{F}_i\right) + \alpha\operatorname{Tr}\left(\mathbf{P}_i^{\mathrm{T}}\mathbf{Q}\mathbf{P}_i\right) + \beta\left\|\mathbf{F}_i - \sum_j w_{ij}\mathbf{F}_j\right\|^2 \\ &\text{s.t.} \quad \mathbf{F}_i^{\mathrm{T}}\mathbf{F}_i = \mathbf{I}_k \end{aligned} \tag{9}$$

where $\mathbf{Q} = Diag(\mathbf{q})$, and $\mathbf{q}$ is an auxiliary vector of the $\ell_{2,1}$ norm. The elements of $\mathbf{q}$ are computed by $q_j = \frac{1}{2\sqrt{\|\mathbf{p}_i^j\|_2^2 + \epsilon}}$, where ϵ is a smoothing term and is usually set to be a small constant value (we set $\epsilon = 10^{-4}$ in this paper).

The quadratic optimization problem with orthogonal constraint have been well studied, and can be solved by a lot of solvers [1,24]. Here we employ the solver Algorithm 2 in [24] to solve Eq. (9), which is a more efficient optimization algorithm with publicly available code.

Another optimization problem we need to solve is Eq. (6). In [19], such a minimization problem with respect to vectors is solved as a constrained least squares problem for locally linear embedding. Since the Frobenius norm for matrices is a straightforward generalization of the l_2 norm for vectors, we can directly obtain the following equation based on the analysis in [19].

$$\left\|\mathbf{X}_i - \sum_j w_{ij}\mathbf{X}_j\right\|_F^2 = \sum_{jr} w_{ij}w_{ir}\mathbf{C}_{jr} \tag{10}$$

where G_j and G_r denote two neighbors of G_i, *i.e.*, G_j and G_r are in the cluster containing G_i. $\mathbf{C}_{jr}$ is the local covariance matrix, which can be obtained by

$$\mathbf{C}_{jr} = \frac{1}{2}(M_j + M_r - m_{jr} - M_0) \tag{11}$$

where $m_{jr} = -s_{jr}$ denotes the squared distance between the jth and rth neighbors of G_i, thus can be obtained by Eq. (1), $M_j = \sum_z m_{jz}$, $M_r = \sum_z m_{rz}$ and $M_0 = \sum_{jr} m_{jr}$. Then the optimal weights can be obtained by:

$$w_{ij} = \frac{\sum_r \mathbf{C}_{jr}^{-1}}{\sum_{lz}\mathbf{C}_{lz}^{-1}} \tag{12}$$

For details about the derivation of the above solution, readers can refer to [19]. Based on the above analysis, we summarize the overall optimization algorithm of MGCT in Algorithm 1.

Algorithm 1. MGCT

Input: $D = \{G_1, G_2, \cdots, G_n\}$, c, k
Output: Assignments to c clusters

Initialize $\mathbf{H}_0$ s.t. $\mathbf{H}_0^{\mathrm{T}}\mathbf{H}_0 = \mathbf{I}_c$;
while not converge **do**
  Compute $\mathbf{W}$ according to Eq. (12);
  for $i = 1; i <= n; i++$ **do**
    Initialize $\mathbf{F}_{i0}$ s.t. $\mathbf{F}_{i0}^{\mathrm{T}}\mathbf{F}_{i0} = \mathbf{I}_k$,$t \leftarrow 0$;
    while not converge **do**
      Set $\mathbf{Q}_t \leftarrow Diag(\frac{1}{2\sqrt{\|\mathbf{p}_t^i\|_2^2+\epsilon}})$;
      Compute $\mathbf{F}_{it+1}$ by solving Eq. (9);
      $t \leftarrow t + 1$;
    end while
  end for
  Update $\mathbf{H}$ by solving Eq. (2);
  Cluster $\mathbf{H}$ by k-means;
end while

4 Experiments

In order to empirically evaluate the effectiveness of the proposed multi-graph clustering approach for brain network analysis, we test our model on real fMRI brain network data and compare with several state-of-the-art baselines.

4.1 Data Collection and Preprocessing

In this work, we use two real resting-state fMRI datasets as follows:

- *Human Immunodeficiency Virus Infection (HIV):* This dataset is collected from Chicago Early HIV Infection Study in Northwestern University [18]. The clinical cohort in this study includes 77 subjects, 56 of which are early HIV patients (positive) and the other 21 are seronegative controls (negative). The two groups did not differ in the demographic characteristics including age, gender, racial composition and education level.
- *Attention Deficit Hyperactivity Disorder (ADHD):* This dataset is collected from ADHD-200 global competition dataset[1], which contains the resting-state fMRI images of 768 subjects. Subjects are either ADHD patients or normal controls. In particular, the patient group in ADHD involves three stages of ADHD disease, which can be treated as three different groups, making the total number of groups be 4.

We use DPARSF toolbox[2] for fMRI data preprocessing. A time series of responds is extracted from each of the 116 anatomical volumes of interest (AVOI), which represents the 116 different brain regions. We perform the standard fMRI brain image processing steps, including functional images realignment, slice timing correction and normalization. Afterwards, spatial smoothing is performed on these images with an 8-mm Gaussian kernel for increasing signal-to-noise ratio, followed by the band-pass filtering (0.01–0.08 Hz) and the linear

[1] http://neurobureau.projects.nitrc.org/ADHD200/.
[2] http://rfmri.org/DPARSF.

trend removing of the time series. We also apply linear regression to reduce spurious variance coming from hardware reasons or subject factors such as thermal motion of electrons. After all these preprocessing steps, we compute the brain activity correlations among different brain regions based on the obtained time series for each of them, and then we use the positive correlations to form the links among the regions. Finally, we exclude the 26 cerebellar regions, and each brain is represented as a graph with 90 nodes, which correspond to the 90 cerebral regions.

4.2 Baselines and Metrics

We use four clustering methods as baselines.

- ***k*-means:** a classic clustering method [4]. We convert the adjacency matrix of each subject graph into vectors and then apply the k-means algorithm to cluster all the subject graphs. For the implementation of the k-means algorithm, we adopt the Litekmeans [5], which has been proven to be a fast MATLAB implementation of the k-means algorithm.
- **Spectral Clustering (SC)** [7]**:** a method for constructing graph partitions based on eigenvectors of the adjacency matrix of graph. In the experiment, we apply the normalized spectral clustering algorithm proposed in [20]. We first construct the similarity matrix for the multiple graphs only based on their adjacency matrices and then use the similarity matrix as the input for normalized spectral clustering of the multiple graphs.
- **Clustering Coefficient (CC):** the k-means clustering with clustering coefficient [17] as the feature representation of each graph.
- **Two-layer Spectral Clustering(TSC):** We adapt the typical spectral clustering into both of the two stages in our framework, where spectral clustering on the multi-graph is based on the spectral clustering on each graph. We call the model TSC.
- **MGCT:** our proposed multi-graph clustering method based on interior-node topology. To evaluate the discriminative ability of the sparsity-inducing term, *i.e.*, the $\ell_{2,1}$-norm penalty term in Eq. (7), we employ MGCT with and without the sparsity-inducing term and denote them as **MGCT** and $\mathbf{MGCT}_{nonST}$ respectively.

We adopt the following two measures for the evaluation.

- *Accuracy*. Let c_i represent the clustering label result of a multi-graph clustering algorithm and y_i represent the corresponding ground truth label of the graph G_i. Then *Accuracy* is defined as: $Accuracy = \frac{\sum_{i=1}^{n} \delta(y_i, map(c_i))}{n}$, where δ is the Kronecker delta function, and $map(c_i)$ is the best mapping function that permutes clustering labels to match the ground truth labels using the KuhnMunkres algorithm [13]. A larger *Accuracy* indicates better clustering performance.

– *Purity*. *Purity* is a measure used to evaluate the clustering method's ability to recover the groundtruth class labels, and it is defined as: $Purity = \frac{1}{n}\sum_{q=1}^{k} \max_{1\leq j\leq l} n_q^j$, where n is the total number of samples, and n_q^j is the number of samples in cluster q that belongs to original class j. Therefore, the purity is a real number in $[0, 1]$. The larger the *Purity*, the better the clustering performance.

The main parameters in our framework include the weight parameters α, β, and δ as well as the number of interior-node clusters k. Note that in the rest part of this paper, we use k specifically to denote the number of interior-node clusters in each graph although it might has been used for denoting other general variables in the equations above. For the convenience of evaluation, we directly use the number of distinct labels in each dataset as the number of clusters in multi-graph clustering. Since there are four possible labels of the samples in ADHD datasets, we set the number of clusters to be 4. For HIV dataset, we have two possible labels (positive, negative), so we set the cluster number to be 2. We apply the grid search to find the optimal values for α, β and δ. We do grid search for α in $\{10^{-2}, 10^{-1}, \cdots 10^{2}\}$, β in $\{10^{-4}, 10^{-3}, \cdots 10^{4}\}$, and δ in $\{0.1, 0.2, \cdots 0.9\}$. The optimal k is selected by the grid search from $\{2, 3, \cdots, 12\}$. For fair comparisons of all the methods, we employ Litekmeans [5] for all the k-means clustering step if it is needed in the implementation of the six methods listed above. We repeat clustering for 20 times with random initialization as k-means depends on initialization. For the evaluation, we repeat running the program of each methods for 50 times and report the average accuracy and purity as the results.

4.3 Performance Evaluations

As shown in Tables 1 and 2, our MGCT method performs the best on the two datasets in terms of both *accuracy* and *purity*. Among the six clustering methods, the first two methods (*i.e.*, k-means, Spectral Clustering) directly use the original matrix of each graph in the data set for calculating the distance or similarity between each pair of the graphs, which is utilized for the final multi-graph clustering. From Tables 1 and 2, we can see that the clustering accuracy and purity of these two methods are quite low. This is probably because that they do not consider the interior-node topology of these graphs when doing clustering. The CC achieves a slightly better result compared to k-means and Spectral Clustering. This is mainly due to the fact that CC does consider some local structure information while calculating the clustering coefficient. However, since it only assigns a single predefined measure to each node in the graph, which represents each brain region in the brain networks, the subtle topological structure of each brain network might not be fully characterized.

Comparatively, the last three methods (*i.e.*, TSC, MGCT, MGCT$_{nonST}$) all utilize the topological structure information but at different level. The TSC method first performs spectral clustering on each graph, and the resulted matrix containing the clustering indicator vectors are used in the multi-graph spectral clustering. This

process does include the topological structure, but it only has the one-way and one-time influence on the multi-graph clustering task. The result of multi-graph clustering does not have influence on the interior-node clustering. Different from TSC, the two methods we proposed namely the MGCT and MGCT$_{nonST}$ perform the task in an iterative way, and achieves the mutual reinforcement by leveraging the topology structure into multi-graph clustering and inferring a better topology structure for each graph from the multi-graph clustering result alternatively. According to Tables 1 and 2, we can also see that the proposed MGCT method outperforms the MGCT$_{nonST}$ in both accuracy and purity. This indicates the importance of the $\ell_{2,1}$ norm we add in Eq. (7), which has the sparsity-inducing property.

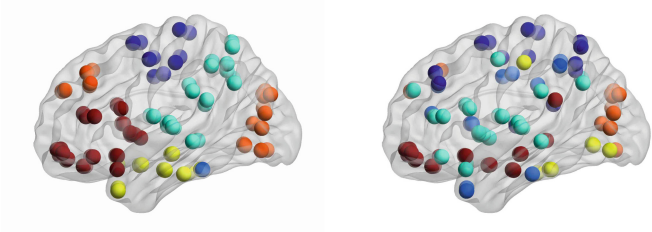
In order to evaluate the effectiveness of MGCT for interior-node topology extraction of brain networks, we investigate the resulted brain networks with interior-node clusters detected by MGCT and show the results of two brain networks in Fig. 2. We can find from the figure that the interior nodes of the normal brain network have been well grouped into several clusters, while the cluster boundaries in the patient's brain network are very blurred and the nodes widely spread out. Usually, the correlated regions of human brain will work together towards a task, and tend to present an approximately synchronized trend in their time series. Thus, the nodes representing these correlated regions would become more possible to be grouped into the same cluster. Therefore, the fuzzy cluster boundaries of the patient's interior nodes indicate that the collaboration activity

Table 1. Clustering *Accuracy*.

Methods	*Accuracy*	
	ADHD ($k = 6$)	HIV ($k = 9$)
k-means	52.0 %	60.3 %
Spectral Clustering	55.2 %	60.9 %
CC	56.8 %	63.7 %
TSC	57.6 %	62.5 %
MGCT$_{nonST}$	**59.3 %**	**64.9 %**
MGCT	**62.8 %**	**68.1 %**

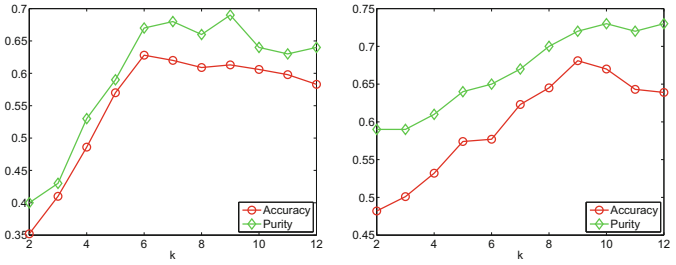
Table 2. Clustering *Purity*.

Methods	*Purity*	
	ADHD ($k = 6$)	HIV ($k = 9$)
k-means	0.55	0.63
Spectral Clustering	0.59	0.65
CC	0.57	0.66
TSC	0.57	0.64
MGCT$_{nonST}$	**0.62**	**0.69**
MGCT	**0.67**	**0.72**



(a) a typical normal control (b) a stage-2 ADHD patient

Fig. 2. Comparison of two brain networks with interior-node topology captured by MGCT from two subject graphs in ADHD dataset



(a) ADHD (b) HIV

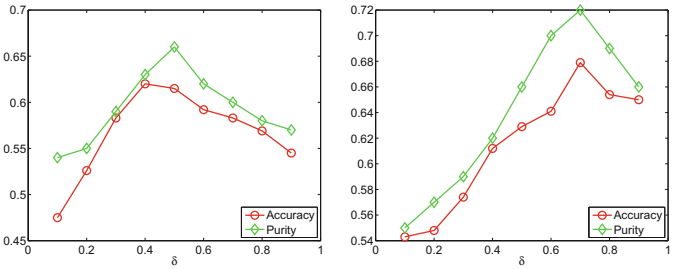
Fig. 3. Accuracy and purity with different k

of different regions might not be very organized. These observations imply that our proposed framework can be further used for distinguishing subjects with neurological disorders from healthy controls.

4.4 Parameter Sensitivity

In this section, we explore the sensitivity and effects of the four main parameters in our proposed method, including α, β, δ and k. We first evaluate the clustering performance of MGCT with different k values, ranging from 2 to 12. Figure 3 shows the clustering performance of MGCT in accuracy and purity with different k on both ADHD and HIV datasets. As we can see from the figure, the multi-graph clustering performance is very sensitive to the value of k, especially when the value for k keeps very small. For example, as shown in Fig. 3(a), the accuracy increases dramatically when the value of k goes from 2 to 6 before it reaches the peak value at 6. The main reason for such high sensitivity is that when k is set to be a small number, the interior-node clusters identified from each brain network tend to have large sizes, which could not capture the interior-node topological structure very well, resulting in a less discriminative measure for distinguishing subjects in different neurological states. A similar changing trend is shown for the purity, while noticeably the peak purity value shows up when $k = 9$ instead of $k = 6$. This can be traced back to the definition of *purity*. Since it counts the number of nodes in the dominated class for each cluster instead of counting the number of nodes only when they match the correct groundtruth labels. Thus, when the number of clusters increases, each cluster becomes easier to be dominated by one class, leading to a higher purity.

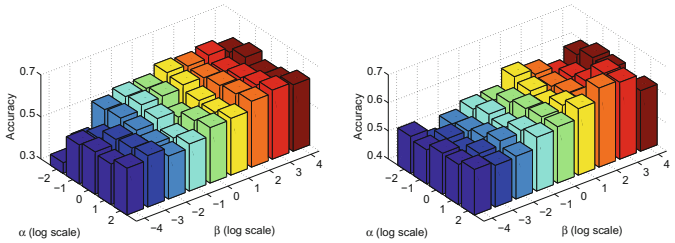
Now, we analyze the sensitivity of MGCT to δ, which balances the weights from the original affinity matrix and the interior-node clustering indicator matrix when creating the similarity matrix among multiple graphs. As shown in Fig. 4, MGCT achieves different level of accuracy and purity when the value of δ varies. For ADHD, the highest accuracy is achieved when $\delta = 0.4$, while for HIV, it achieves the highest accuracy when $\delta = 0.7$, and similar situations for the purity. These results indicate that both the global structure and the interior-node topo-



(a) ADHD (b) HIV

Fig. 4. Accuracy and purity with different δ

logical structure are important for the multi-graph clustering analysis, and their weights need to be determined for specific practical situations. Next, we evaluate the sensitivity of MGCT to α and β. We set k to be 6 and run the MGCT method with different values for α and β on ADHD and HIV data. The clustering accuracy of MGCT is plotted versus the values for α and β in Fig. 5. As shown in the figure, MGCT achieves the best performance when $\alpha = 10^2, \beta = 10^3$ on ADHD dataset, and $\alpha = 10^2, \beta = 10^2$ on HIV dataset. Parameter α controls the sparsity while parameter β controls the influence of iterative multi-graph clustering results on interior-node clustering. If the value for α is very small, then it will not really enforce the sparsity. Similarly, if the value for β is quite small, the iterative process would barely have influence on interior-node clustering optimizing. In these cases, the performance will decline. However, when the values for them are too large, they would enforce too much sparsity or influence, which might make the performance drop as well. Therefore, an optimal combination of the two parameters is crucial for improving the performance of MGCT.



(a) ADHD (b) HIV

Fig. 5. Accuracy with different α, β

5 Related Work

Our work relates to several bodies of studies, including the multi-graph clustering, node clustering in graphs, and brain network analysis.

In the context of multi-graph clustering, there are a few of strategies that have been proposed and widely used [3], for example the structural summary method discussed in [2], and the hierarchical algorithm with graph structure proposed in [15]. However, these methods only focus on finding a summary from the global structure of the graphs without looking into the topological structure, thus would lose very important local structural information, leading to a less effective clustering of multiple graphs.

For node clustering in graphs, there has also been a vast literature of works. One classic category of these methods are the spectral clustering algorithms [22], which use the eigenvalues of the Laplacian matrix to perform dimension reduction and then cluster the data in fewer dimensions. Recently, new methods of node clustering have been proposed for various applications, such as the works for social network analysis [25,26], which utilize the heterogeneous information in aligned networks for node clustering. Although these work use information from multiple graphs, they focus on the mutual relationship of graphs at the node level instead of the graph-graph neighbourhood relationship as we consider.

Brain network analysis has become a hot research topic of medical data mining these years. A major task in brain network analysis is to identify the difference of a healthy subject and a neurological demented subject in brain network structure. In the past decade, quite a few of works have been done to solve this problem. In [11], a discriminative subgraph mining method is proposed for classifying brain networks. In [14], they find a unified cut and a contrast cut of multiple graphs for studying brain networks of multiple subjects. This work is the most related one of ours. However, they study the brain networks when the labels of subjects (healthy or demented) are given, while we cluster the unlabeled subjects into groups with their brain network features.

6 Conclusions

In this paper, we propose an iterative framework MGCT for multi-graph clustering based on interior-node topology of graphs. To capture the local topological structure of the graphs, we perform the sparsity-inducing interior-node clustering on each graph. In this framework, the interior-node clustering and the multi-graph clustering are performed alternatively, where the results of interior-node clustering are exerted on multi-graph clustering and the multi-graph clustering in turn improves the interior-node clustering of each graph. After this iterative mutual reinforcement process, we can obtain a refined multi-graph clustering result, which can be used for further analysis of the graphs. Experiments on two real brain network datasets demonstrate the superior performance of the proposed model in multi-graph clustering for brain network analysis.

Acknowledgments. This work is supported in part by NSF through grants III-1526499, NSFC through grants 61503253 and 61472089, and NIH through grant $R01 - MH080636$.

References

1. Absil, P.A., Mahony, R., Sepulchre, R.: Optimization Algorithms on Matrix Manifolds. Princeton University Press, Princeton (2009)
2. Aggarwal, C.C., Ta, N., Wang, J., Feng, J., Zaki, M.: XProj: a framework for projected structural clustering of XML documents. In: KDD, pp. 46–55. ACM (2007)
3. Aggarwal, C.C., Wang, H.: A survey of clustering algorithms for graph data. In: Aggarwal, C.C., Wang, H. (eds.) Managing and Mining Graph Data, pp. 275–301. Springer, New York (2010)
4. Berkhin, P.: A survey of clustering data mining techniques. In: Kogan, J., Nicholas, C., Teboulle, M. (eds.) Grouping Multidimensional Data, pp. 25–71. Springer, Heidelberg (2006)
5. Cai, D.: Litekmeans: the fastest matlab implementation of kmeans. Software (2011). http://www.zjucadcg.cn/dengcai/Data/Clustering.html
6. Cao, B., Kong, X., Zhang, J., Philip, S.Y., Ragin, A.B.: Mining brain networks using multiple side views for neurological disorder identification. In: ICDM, pp. 709–714. IEEE (2015)
7. Donath, W.E., Hoffman, A.J.: Lower bounds for the partitioning of graphs. IBM J. Res. Dev. **17**(5), 420–425 (1973)
8. He, L., Lu, C.T., Ma, J., Cao, J., Shen, L., Philip, S.Y.: Joint community and structural hole spanner detection via harmonic modularity. In: KDD. ACM (2016)
9. He, R., Tan, T., Wang, L., Zheng, W.S.: $\ell_{2,1}$ regularized correntropy for robust feature selection. In: CVPR, pp. 2504–2511 (2012)
10. Jie, B., Zhang, D., Gao, W., Wang, Q., Wee, C., Shen, D.: Integration of network topological and connectivity properties for neuroimaging classification. Biomed. Eng. **61**(2), 576 (2014)
11. Kong, X., Ragin, A.B., Wang, X., Yu, P.S.: Discriminative feature selection for uncertain graph classification. In: SDM, pp. 82–93. SIAM (2013)
12. Kong, X., Yu, P.S.: Brain network analysis: a data mining perspective. ACM SIGKDD Explor. Newsl. **15**(2), 30–38 (2014)
13. Kuhn, H.W.: The hungarian method for the assignment problem. Naval Res. Logistics Q. **2**(1–2), 83–97 (1955)
14. Kuo, C.T., Wang, X., Walker, P., Carmichael, O., Ye, J., Davidson, I.: Unified and contrasting cuts in multiple graphs: application to medical imaging segmentation. In: KDD, pp. 617–626. ACM (2015)
15. Lian, W., Mamoulis, N., Yiu, S.M., et al.: An efficient and scalable algorithm for clustering XML documents by structure. IEEE Trans. Knowl. Data Eng. **16**(1), 82–96 (2004)
16. Nikolova, M., Ng, M.K.: Analysis of half-quadratic minimization methods for signal and image recovery. SIAM J. Sci. Comput. **27**(3), 937–966 (2005)
17. Onnela, J.P., Saramäki, J., Kertész, J., Kaski, K.: Intensity and coherence of motifs in weighted complex networks. Phys. Rev. E **71**(6), 065103 (2005)
18. Ragin, A.B., Du, H., Ochs, R., Wu, Y., Sammet, C.L., Shoukry, A., Epstein, L.G.: Structural brain alterations can be detected early in HIV infection. Neurology **79**(24), 2328–2334 (2012)
19. Saul, L.K., Roweis, S.T.: An introduction to locally linear embedding (2000). http://www.cs.toronto.edu/~roweis/lle/publications.html
20. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **22**(8), 888–905 (2000)

21. Spielman, D.A.: Algorithms, graph theory, and linear equations in Laplacian matrices. In: ICM, vol. 4, pp. 2698–2722 (2010)
22. Von Luxburg, U.: A tutorial on spectral clustering. Stat. Comput. **17**(4), 395–416 (2007)
23. Wee, C.Y., Yap, P.T., Zhang, D., Denny, K., Browndyke, J.N., Potter, G.G., Welsh-Bohmer, K.A., Wang, L., Shen, D.: Identification of MCI individuals using structural and functional connectivity networks. Neuroimage **59**(3), 2045–2056 (2012)
24. Wen, Z., Yin, W.: A feasible method for optimization with orthogonality constraints. Math. Program. **142**(1–2), 397–434 (2013)
25. Zhang, J., Yu, P.S.: Community detection for emerging networks. In: SDM. SIAM (2015)
26. Zhang, J., Yu, P.S.: Mutual clustering across multiple heterogeneous networks. In: IEEE BigData Congress (2015)