Mining Brain Networks using Multiple Side Views for Neurological Disorder Identification

Bokai Cao*, Xiangnan Kong†, Jingyuan Zhang*, Philip S. Yu*‡ and Ann B. Ragin§
*Department of Computer Science, University of Illinois at Chicago, IL, USA; {caobokai, jzhan8, psyu}@uic.edu
†Department of Computer Science, Worcester Polytechnic Institute, MA, USA; xkong@wpi.edu
‡Institute for Data Science, Tsinghua University, Beijing, China
§Department of Radiology, Northwestern University, IL, USA; ann-ragin@northwestern.edu

***Abstract*—Mining discriminative subgraph patterns from graph data has attracted great interest in recent years. It has a wide variety of applications in disease diagnosis, neuroimaging, *etc.* Most research on subgraph mining focuses on the graph representation alone. However, in many real-world applications, the side information is available along with the graph data. For example, for neurological disorder identification, in addition to the brain networks derived from neuroimaging data, hundreds of clinical, immunologic, serologic and cognitive measures may also be documented for each subject. These measures compose multiple side views encoding a tremendous amount of supplemental information for diagnostic purposes, yet are often ignored. In this paper, we study the problem of discriminative subgraph selection using multiple side views and propose a novel solution to find an optimal set of subgraph features for graph classification by exploring a plurality of side views. We derive a feature evaluation criterion, named gSide, to estimate the usefulness of subgraph patterns based upon side views. Then we develop a branch-and-bound algorithm, called gMSV, to efficiently search for optimal subgraph features by integrating the subgraph mining process and the procedure of discriminative feature selection. Empirical studies on graph classification tasks for neurological disorders using brain networks demonstrate that subgraph patterns selected by the multi-side-view guided subgraph selection approach can effectively boost graph classification performances and are relevant to disease diagnosis.**

***Index Terms*—subgraph pattern, graph mining, side information, brain network.**

I. Introduction

Recent years have witnessed an increasing amount of data in the form of graph representations, which involve complex structures, *e.g.*, brain networks, social networks. These data are inherently represented as a set of nodes and links, instead of feature vectors as traditional data. For example, brain networks are composed of brain regions as the nodes, *e.g.*, *insula*, *hippocampus*, *thalamus*, and functional/structural connectivities between the brain regions as the links. The linkage structure in these brain networks can encode tremendous information concerning the integrated activity of the human brain. For example, in brain networks derived from functional magnetic resonance imaging (fMRI), connections/links can encode correlations between brain regions in functional activity. While structural links in diffusion tensor imaging (DTI) can capture white matter fiber pathways connecting different brain regions. The complex structures and the lack of vector representations within these graph data raise a challenge for data mining. An effective model for mining the graph data should be able to extract a set of subgraph patterns for further analysis. Motivated by such challenges, graph mining research problems, in particular graph classification, have received considerable attention in the last decade.

The graph classification problem has been studied extensively. Conventional approaches focus on mining discriminative subgraphs from graph view alone. This is usually feasible for applications like molecular graph analysis, where a large set of graph instances with labels are available. For brain network analysis, however, usually we only have a small number of graph instances, ranging from 30 to 100 brain networks [12]. In these applications, the information from the graph view alone may not be sufficient for mining important subgraphs. Commonly, however, in neurological studies, hundreds of clinical, serologic and cognitive measures are available for each subject in addition to brain networks derived from the neuroimaging data [2], [3]. These measures comprise multiple side views. This supplemental information, which is generally ignored, may contain a plurality of side views to guide the process of subgraph mining in brain networks.

Despite its value and significance, the feature selection problem for graph data using auxiliary views has not been studied in this context so far. There are two major difficulties in learning from multiple side views for graph classification, as follows:

The primary view in graph representation: Graph data naturally composes the primary view for graph mining problems, from which we want to select discriminative subgraph patterns for graph classification. However, it raises a challenge for data mining with the complex structures and the lack of vector representations. Conventional feature selection approaches in vector spaces usually assume that a set of features are given before conducting feature selection. In the context of graph data, however, subgraph features are embedded within the graph structures and usually it is not feasible to enumerate the full set of subgraph features for a graph dataset before feature selection. Actually, the number of subgraph features grows exponentially with the size of graphs.

The side views in vector representations: In many applications, side information is available along with the graph data and usually exists in the form of vector representations. That is to say, an instance is represented by a graph and additional

vector-based features at the same time. It introduces us to the problem of how to leverage the relationship between the primary graph view and a plurality of side views, and how to facilitate the subgraph mining procedure by exploring the vector-based auxiliary views. For example, in brain networks, discriminative subgraph patterns for neurological disorders indicate brain injuries associated with particular regions. Such changes can potentially express in other medical tests of the subject, *e.g.*, clinical, immunologic, serologic and cognitive measures. Thus, it would be desirable to select subgraph features that are consistent with these side views.

In conventional graph classification approaches, the valuable information embedded in side views is not fully leveraged in process of selecting subgraph patterns. Most subgraph mining approaches focus on the drug discovery problem which have access to a great amount of graph data for chemical compounds. For neurological disorder identification, however, there are usually limited subjects with a small sample size of brain networks available. Therefore, it is critical to learn knowledge from other possible sources. We notice that transfer learning can borrow supervision knowledge from the source domain to help the learning on the target domain, *e.g.*, finding a good feature representation [7], mapping relational knowledge [15], [16], and learning across graph database [18]. However, to our knowledge, they do not consider transferring complementary information from vector-based side views to graph database whose instances are complex structural graphs.

To solve the above problems, in this paper, we introduce a novel framework for discriminative subgraph selection using multiple side views. In contrast to existing subgraph mining approaches that focus on a single view of the graph representation, our method can explore multiple vector-based side views to find an optimal set of subgraph features for graph classification. We first verify side information consistency via statistical hypothesis testing. Based on auxiliary views and the available label information, we design an evaluation criterion for subgraph features, named gSide. By deriving a lower bound, we develop a branch-and-bound algorithm, called gMSV, to efficiently search for optimal subgraph features with pruning, thereby avoiding exhaustive enumeration of all subgraph features. In order to evaluate our proposed model, we conduct experiments on graph classification tasks for neurological disorders, using fMRI and DTI brain networks. The experiments demonstrate that our subgraph selection approach using multiple side views can effectively boost graph classification performances. Moreover, we show that gMSV is more efficient by pruning the subgraph search space via gSide. A full version of this paper is available in [4].

II. Problem Formulation

A motivation for this work is the premise that side information could be strongly correlated with neurological status. Before presenting the subgraph feature selection model, we first introduce the notations that will be used throughout this paper. Let $\mathcal{D} = \{G_1, \cdots, G_n\}$ denote the graph dataset, which consists of n graph objects. The graphs within $\mathcal{D}$ are labeled by $[y_1, \cdots, y_n]^\top$, where $y_i \in \{-1, +1\}$ denotes the binary class label of G_i.

Definition *1 (Side view):* A side view is a set of vector-based features $\mathbf{z}_i = [z_1, \cdots, z_d]^\top$ associated with each graph object G_i, where d is the dimensionality of this view. A side view is denoted as $\mathcal{Z} = \{\mathbf{z}_1, \cdots, \mathbf{z}_n\}$.

We assume that there are multiple side views $\{\mathcal{Z}^{(1)}, \cdots, \mathcal{Z}^{(v)}\}$ available for the graph dataset $\mathcal{D}$, where v is the number of side views. We employ kernels $\kappa^{(p)}$ on $\mathcal{Z}^{(p)}$, such that $\kappa_{ij}^{(p)}$ represents the similarity between G_i and G_j from the perspective of the p-th view. The RBF kernel is used as the default kernel in this paper, unless otherwise specified.

Definition *2 (Graph):* A graph is represented as $G = (V, E)$, where $V = \{v_1, \cdots, v_{n_v}\}$ is the set of vertices, $E \subseteq V \times V$ is the set of edges.

Definition *3 (Subgraph):* Let $G' = (V', E')$ and $G = (V, E)$ be two graphs. G' is a subgraph of G (denoted as $G' \subseteq G$) iff $V' \subseteq V$ and $E' \subseteq E$. If G' is a subgraph of G, then G is supergraph of G'.

In this paper, we adopt the idea of subgraph-based graph classification approaches, which assume that each graph object G_j is represented as a binary vector $\mathbf{x}_j = [x_{1j}, \cdots, x_{mj}]^\top$ associated with the full set of subgraph patterns $\{g_1, \cdots, g_m\}$ for the graph dataset $\{G_1, \cdots, G_n\}$. Here $x_{ij} \in \{0, 1\}$ is the binary feature of G_j corresponding to the subgraph pattern g_i, and $x_{ij} = 1$ iff g_i is a subgraph of G_j ($g_i \subseteq G_j$), otherwise $x_{ij} = 0$. Let $X = [x_{ij}]^{m \times n}$ denote the matrix consisting of binary feature vectors using $\mathcal{S}$ to represent the graph dataset $\mathcal{D}$. $X = [\mathbf{x}_1, \cdots, \mathbf{x}_n] = [\mathbf{f}_1, \cdots, \mathbf{f}_m]^\top \in \{0, 1\}^{m \times n}$. The full set $\mathcal{S}$ is usually too large to be enumerated. There is usually only a subset of subgraph patterns $\mathcal{T} \subseteq \mathcal{S}$ relevant to the task of graph classification.

III. Data Analysis

A motivation for this work is that the side information could be strongly correlated with the health state of a subject. Before proceeding, we first introduce real-world data used in this work and investigate whether the available information from side views has any potential impact on neurological disorder identification.

A. Data Collections

In this paper, we study the real-world datasets collected from the Chicago Early HIV Infection Study at Northwestern University [17]. The clinical cohort includes 56 HIV (positive) and 21 seronegative controls (negative). The datasets contain functional magnetic resonance imaging (fMRI) and diffusion tensor imaging (DTI) for each subject, from which brain networks can be constructed, respectively. A detailed description about network construction is available in [4].

In addition, for each subject, hundreds of clinical, imaging, immunologic, serologic and cognitive measures were documented. Seven groups of measurements were investigated in our datasets, including *neuropsychological tests*, *flow cytometry*, *plasma luminex*, *freesurfer*, *overall brain microstructure*,

localized brain microstructure, *brain volumetry*. Each group can be regarded as a distinct view that partially reflects subject status, and measurements from different medical examinations can provide complementary information. Moreover, we preprocessed the features by min-max normalization before employing the RBF kernel on each view.

B. Verifying Side Information Consistency

We study the potential impact of side information on selecting subgraph patterns via statistical hypothesis testing. Side information consistency suggests that the similarity of side view features between instances with the same label should have higher probability to be larger than that with different labels. We use hypothesis testing to validate whether this statement holds in the fMRI and DTI datasets.

For each side view, we first construct two vectors $\mathbf{a}_s^{(p)}$ and $\mathbf{a}_d^{(p)}$ with an equal number of elements, sampled from the sets $\mathcal{A}_s^{(p)}$ and $\mathcal{A}_d^{(p)}$, respectively:

$$\mathcal{A}_s^{(p)} = \{\kappa_{ij}^{(p)} | y_i y_j = 1\} \tag{1}$$

$$\mathcal{A}_d^{(p)} = \{\kappa_{ij}^{(p)} | y_i y_j = -1\} \tag{2}$$

Then, we form a two-sample one-tail t-test to validate the existence of side information consistency. We test whether there is sufficient evidence to support the hypothesis that the similarity score in $\mathbf{a}_s^{(p)}$ is larger than that in $\mathbf{a}_d^{(p)}$. The null hypothesis is $H_0 : \mu_s^{(p)} - \mu_d^{(p)} \leq 0$, and the alternative hypothesis is $H_1 : \mu_s^{(p)} - \mu_d^{(p)} > 0$, where $\mu_s^{(p)}$ and $\mu_d^{(p)}$ represent the sample means of similarity scores in the two groups, respectively.

The t-test results show that there is strong evidence, with significance level $\alpha = 0.05$, to reject the null hypothesis on the two datasets. In other words, we validate the existence of side information consistency in neurological disorder identification, thereby paving the way for our next study of leveraging multiple side views for discriminative subgraph selection.

IV. Multi-Side-View Discriminative Subgraph Selection

In this section, we address the first problem discussed in Section II by formulating the discriminative subgraph selection problem as a general optimization framework as follows:

$$\mathcal{T}^* = \operatorname*{argmin}_{\mathcal{T} \subseteq \mathcal{S}} \mathcal{F}(\mathcal{T}) \quad \text{s.t. } |\mathcal{T}| \leq k \tag{3}$$

where $|\cdot|$ denotes the cardinality and k is the maximum number of feature selected. $\mathcal{F}(\mathcal{T})$ is the evaluation criterion to estimate the score (can be the lower the better in this paper) of a subset of subgraph patterns $\mathcal{T}$. $\mathcal{T}^*$ denotes the optimal set of subgraph patterns $\mathcal{T}^* \subseteq \mathcal{S}$.

A. Exploring Multiple Side Views: gSide

Following the observations in Section III-B that the side view information is clearly correlated with the prespecified label information, we assume that the set of optimal subgraph patterns should have the following properties. The similarity/distance between instances in the space of subgraph features should be consistent with that in the space of a side view. That is to say, if two instances are similar in the space of the p-th view (*i.e.*, a high $\kappa_{ij}^{(p)}$ value), they should also be close to each other in the space of subgraph features (*i.e.*, a small distance between subgraph feature vectors). On the other hand, if two instances are dissimilar in the space of the p-th view (*i.e.*, a low $\kappa_{ij}^{(p)}$ value), they should be far away from each other in the space of subgraph features (*i.e.*, a large distance between subgraph feature vectors). Therefore, our objective function could be to minimize the distance between subgraph features of each pair of similar instances in each side view, and maximize the distance between dissimilar instances. This idea is formulated as follows:

$$\operatorname*{argmin}_{\mathcal{T} \subseteq \mathcal{S}} \frac{1}{2} \sum_{p=1}^{v} \lambda^{(p)} \sum_{i,j=1}^{n} \|\mathcal{I}_{\mathcal{T}} \mathbf{x}_i - \mathcal{I}_{\mathcal{T}} \mathbf{x}_j\|_2^2 \Theta_{ij}^{(p)} \tag{4}$$

where $\mathcal{I}_{\mathcal{T}}$ is a diagonal matrix indicating which subgraph features are selected into $\mathcal{T}$ from $\mathcal{S}$, $(\mathcal{I}_{\mathcal{T}})_{ii} = 1$ iff $g_i \in \mathcal{T}$, otherwise $(\mathcal{I}_{\mathcal{T}})_{ii} = 0$. The parameters $\lambda^{(p)} \geq 0$ are employed to control the contributions from each view.

$$\Theta_{ij}^{(p)} = \begin{cases} \frac{1}{|\mathcal{H}^{(p)}|} & (i,j) \in \mathcal{H}^{(p)} \\ -\frac{1}{|\mathcal{L}^{(p)}|} & (i,j) \in \mathcal{L}^{(p)} \end{cases} \tag{5}$$

where $\mathcal{H}^{(p)} = \{(i,j) | \kappa_{ij}^{(p)} \geq \mu^{(p)}\}$, $\mathcal{L}^{(p)} = \{(i,j) | \kappa_{ij}^{(p)} < \mu^{(p)}\}$, and $\mu^{(p)}$ is the mean value of $\kappa_{ij}^{(p)}$, *i.e.*, $\frac{1}{n^2} \sum_{i,j=1}^{n} \kappa_{ij}^{(p)}$. This normalization is to balance the effect of similar instances and dissimilar instances.

Intuitively, Eq. (4) will minimize the distance between subgraph features of similar instance-pairs with $\kappa_{ij}^{(p)} \geq \mu^{(p)}$, while maximize the distance between dissimilar instance-pairs with $\kappa_{ij}^{(p)} < \mu^{(p)}$ in each view. In this way, the side view information is effectively used to guide the process of discriminative subgraph selection. The fact verified in Section III-B that the side view information is clearly correlated with the prespecified label information can be very useful, especially in the semi-supervised setting.

With prespecified information for labeled graphs, we further consider that the optimal set of subgraph patterns should satisfy the following constraints: labeled graphs in the same class should be close to each other; labeled graphs in different classes should be far away from each other. Intuitively, these constraints tend to select the most discriminative subgraph patterns based on the graph labels. Such an idea has been well explored in the context of dimensionality reduction and feature selection [1], [19].

The constraints above can be mathematically formulated as minimizing the loss function:

$$\operatorname*{argmin}_{\mathcal{T} \subseteq \mathcal{S}} \frac{1}{2} \sum_{i,j=1}^{n} \|\mathcal{I}_{\mathcal{T}} \mathbf{x}_i - \mathcal{I}_{\mathcal{T}} \mathbf{x}_j\|_2^2 \Omega_{ij} \tag{6}$$

where

$$\Omega_{ij} = \begin{cases} \frac{1}{|\mathcal{M}|} & (i,j) \in \mathcal{M} \\ -\frac{1}{|\mathcal{C}|} & (i,j) \in \mathcal{C} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

and $\mathcal{M} = \{(i,j)|y_i y_j = 1\}$ denotes the set of pairwise constraints between graphs with the same label, and $\mathcal{C} = \{(i,j)|y_i y_j = -1\}$ denotes the set of pairwise constraints between graphs with different labels.

By defining matrix $\Phi \in \mathbb{R}^{n\times n}$ as

$$\Phi_{ij} = \Omega_{ij} + \sum_{p=1}^{v} \lambda^{(p)} \Theta_{ij}^{(p)} \tag{8}$$

we can combine and rewrite the function in Eq. (4) and Eq. (6) as

$$\begin{aligned} \mathcal{F}(\mathcal{T}) &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \|\mathcal{I}_{\mathcal{T}}\mathbf{x}_i - \mathcal{I}_{\mathcal{T}}\mathbf{x}_j\|_2^2 \Phi_{ij} \\ &= \text{tr}(\mathcal{I}_{\mathcal{T}}^{\top} X(D-\Phi)X^{\top}\mathcal{I}_{\mathcal{T}}) \\ &= \text{tr}(\mathcal{I}_{\mathcal{T}}^{\top} X L X^{\top}\mathcal{I}_{\mathcal{T}}) \\ &= \sum_{g_i \in \mathcal{T}} \mathbf{f}_i^{\top} L \mathbf{f}_i \end{aligned} \tag{9}$$

where $\text{tr}(\cdot)$ is the trace of a matrix.

DEFINITION *4 (gSide):* Let $\mathcal{D} = \{G_1, \cdots, G_n\}$ denote a graph dataset with multiple side views. Suppose Φ is a matrix defined as Eq. (8), and L is a Laplacian matrix defined as $L = D - \Phi$, where D is a diagonal matrix, $D_{ii} = \sum_j \Phi_{ij}$. We define an evaluation criterion q, called gSide, for a subgraph pattern g_i as

$$q(g_i) = \mathbf{f}_i^{\top} L \mathbf{f}_i \tag{10}$$

where $\mathbf{f}_i = [f_{i1}, \cdots, f_{in}]^{\top} \in \{0,1\}^n$ is the indicator vector for subgraph pattern g_i, $f_{ij} = 1$ iff $g_i \subseteq G_j$, otherwise $f_{ij} = 0$. Since the Laplacian matrix L is positive semi-definite, for any subgraph pattern g_i, $q(g_i) \geq 0$.

Based on gSide as defined above, the optimization problem in Eq. (3) can be written as

$$\mathcal{T}^* = \underset{\mathcal{T}\subseteq\mathcal{S}}{\text{argmin}} \sum_{g_i \in \mathcal{T}} q(g_i) \quad \text{s.t. } |\mathcal{T}| \leq k \tag{11}$$

B. Searching with A Lower Bound: gMSV

Now we address the second problem discussed in Section II, and propose an efficient method to find the optimal set of subgraph patterns from a graph dataset with multiple side views.

A straightforward solution to the goal of finding an optimal feature set is the exhaustive enumeration, *i.e.*, we could first enumerate all subgraph patterns from a graph dataset, and then calculate the gSide values for all subgraph patterns. In the context of graph data, however, it is usually not feasible to enumerate the full set of subgraph patterns before feature selection. Actually, the number of subgraph patterns grows exponentially with the size of graphs. Inspired by recent advances in graph classification approaches [5], [13], [14], [20], which nest their evaluation criteria into the subgraph mining process and develop constraints to prune the search space, we adopt a similar approach by deriving a different constraint based upon gSide.

By adopting the gSpan algorithm proposed by Yan and Han [21], we can enumerate all the subgraph patterns for a graph dataset in a canonical search space. In order to prune the subgraph search space, we now derive a lower bound of the gSide value:

THEOREM *1:* Given any two subgraph patterns $g_i, g_j \in \mathcal{S}$, g_j is a supergraph of g_i, *i.e.*, $g_i \subseteq g_j$. The gSide value of g_j is bounded by $\hat{q}(g_i)$, *i.e.*, $q(g_j) \geq \hat{q}(g_i)$. $\hat{q}(g_i)$ is defined as

$$\hat{q}(g_i) \triangleq \mathbf{f}_i^{\top} \hat{L} \mathbf{f}_i \tag{12}$$

where the matrix $\hat{L}$ is defined as $\hat{L}_{pq} \triangleq \min(0, L_{pq})$.

PROOF. The detailed proof is presented in [4].

We can now nest the lower bound into the subgraph mining steps in gSpan to efficiently prune the DFS code tree. During the depth-first search through the DFS code tree, we always maintain the currently top-k best subgraph patterns according to gSide and the temporally suboptimal gSide value (denoted by θ) among all the gSide values calculated before. If $\hat{q}(g_i) \geq \theta$, the gSide value of any supergraph g_j of g_i should be no less than $\hat{q}(g_i)$ according to Theorem 1, *i.e.*, $q(g_j) \geq \hat{q}(g_i) \geq \theta$. Thus, we can safely prune the subtree rooted from g_i in the search space. If $\hat{q}(g_i) < \theta$, we cannot prune this subtree since there might exist a supergraph g_j of g_i such that $q(g_j) < \theta$. As long as a subgraph g_i can improve the gSide values of any subgraphs in $\mathcal{T}$, it is added into $\mathcal{T}$ and the least best subgraph is removed from $\mathcal{T}$. Then we recursively search for the next subgraph in the DFS code tree.

V. EXPERIMENTS

In order to evaluate the performance of the proposed solution to the problem of feature selection for graph classification using multiple side views, we tested our algorithm on brain network datasets derived from neuroimaging, as introduced in Section III-A.

A. Experimental Setup

To the best of our knowledge, this paper is the first work on leveraging side information in feature selection problem for graph classification. In order to evaluate the performance of the proposed method, we compare our method with other methods using different statistical measures and discriminative score functions. For all the compared methods, gSpan [21] is used as the underlying searching strategy. Note that although alternative algorithms are available [20], [10], [11], the search step efficiency is not the focus of this paper. The compared methods are summarized as follows:

- gMSV: The proposed discriminative subgraph selection method using multiple side views. Following the observation in Section III-B that side information consistency is verified to be significant in all the side views, the parameters in gMSV are simply set to $\lambda^{(1)} = \cdots = \lambda^{(v)} = 1$ for experimental purposes. In the case where some side views are suspect to be redundant, we can adopt the alternative optimization strategy to iteratively select discriminative subgraph patterns and update view weights.
- gSSC: A semi-supervised feature selection method for graph classification based upon both labeled and unlabeled graphs. The parameters in gSSC are set to $\alpha = \beta = 1$ unless otherwise specified [14].

- Discriminative Subgraphs (Conf, Ratio, Gtest, HSIC): Supervised feature selection methods for graph classification based upon confidence [8], frequency ratio [9], [11], [10], G-test score [20] and HSIC [13], respectively. The top-k discriminative subgraph features are selected in terms of different discrimination criteria.
- Frequent Subgraphs (Freq): In this approach, the evaluation criterion for subgraph feature selection is based upon frequency. The top-k frequent subgraph features are selected.

We append the side view data to the subgraph-based graph representations computed by the above algorithms before feeding the concatenated feature vectors to the classifier. Another baseline that only uses side view data is denoted as MSV.

For a fair comparison, we used LibSVM [6] with linear kernel as the base classifier for all the compared methods. In the experiments, 3-fold cross validations were performed on balanced datasets. To get the binary links, we performed simple thresholding over the weights of the links. The *threshold* for fMRI and DTI datasets was 0.9 and 0.3, respectively.

B. Performance on Graph Classification

The experimental results on the fMRI dataset are shown in Figure 1. Results on the DTI dataset are available in [4]. The average performances with different number of features of each method are reported. Classification accuracy is used as the evaluation metric.

In Figure 1, our method gMSV can achieve the classification accuracy as high as 97.16%, which is significantly better than the union of other subgraph-based features and side view features. The black solid line denotes the method MSV, the simplest baseline that uses only side view data. Conf and Ratio can do slightly better than MSV. Freq adopts an unsupervised process for selecting subgraph patterns, resulting in a comparable performance with MSV, indicating that there is no additional information from the selected subgraphs. Other methods that use different discrimination scores without leveraging the guidance from side views perform even worse than MSV in graph classification, because they evaluate the usefulness of subgraph patterns solely based on the limited label information from a small sample size of brain networks. The selected subgraph patterns can potentially be redundant or irrelevant, thereby compromising the effects of side view data. Importantly, gMSV outperforms the semi-supervised approach gSSC which explores the unlabeled graphs based on the separability property. This indicates that rather than simply considering that unlabeled graphs should be separated from each other, it would be better to regularize such separability/closeness to be consistent with the available side views.

Overall, these results support our premise that exploring a plurality of side views can boost the performance of graph classification, and the gSide evaluation criterion in gMSV can find more informative subgraph patterns for graph classification than subgraphs based on frequency or other discrimination scores. Figure 2 displays the most discriminative subgraph patterns selected by gMSV from the fMRI dataset.

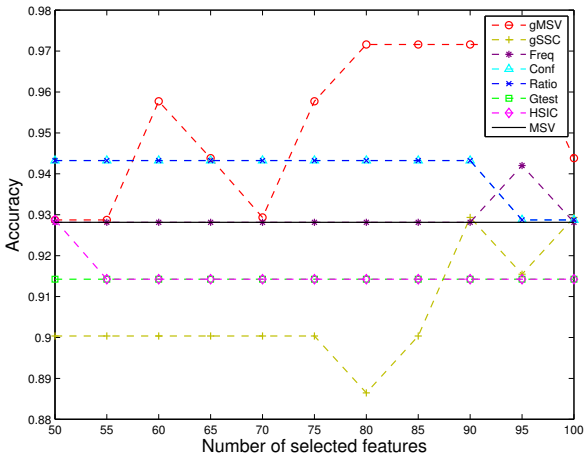
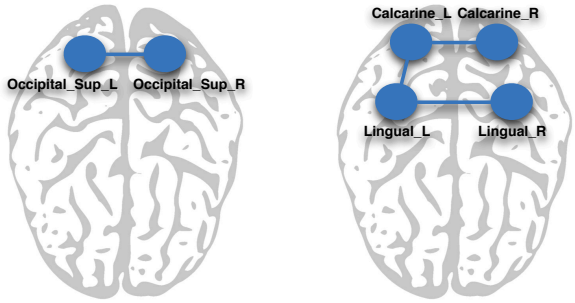


Fig. 1. Classification performance on the fMRI dataset.



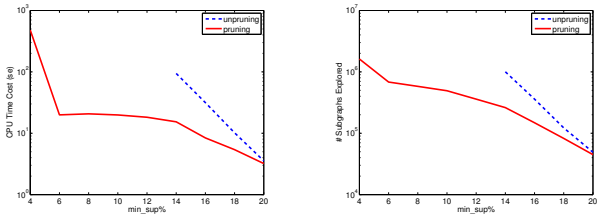
(a) (b)

Fig. 2. Discriminative subgraph patterns that are associated with HIV, selected from the fMRI dataset.

C. Time and Space Complexity

Next, we evaluate the effectiveness of pruning the subgraph search space by adopting the lower bound of gSide in gMSV. In this section, we compare the runtime performance of two implementation versions of gMSV: the pruning gMSV uses the lower bound of gSide to prune the search space of subgraph enumerations; the unpruning gMSV denotes the method without pruning in the subgraph mining process. We test both approaches and recorded the average CPU time used and the average number of subgraph patterns explored during the procedure of subgraph mining and feature selection.

The comparisons with respect to the time complexity and the space complexity on the fMRI dataset are shown in Figure 3. We observe that the unpruning gMSV needs to explore exponentially larger subgraph search space as we decrease the *min_sup* value in the subgraph mining process. When the *min_sup* value is too low, the subgraph enumeration step in the unpruning gMSV can run out of the memory. However, the pruning gMSV is still effective and efficient when the *min_sup* value goes to very low, because its running time and space requirement do not increase as much as the unpruning gMSV by reducing the subgraph search space via



(a) Average CPU time.

(b) Average number of subgraph patterns explored in the mining process.

Fig. 3. Complexity analysis on the fMRI dataset.

TABLE I
AVERAGE CLASSIFICATION PERFORMANCES OF GMSV ON THE FMRI DATASET WITH DIFFERENT SINGLE SIDE VIEWS.

Side views	Acc.	Prec.	Rec.	F1
neuropsychological tests	0.743	0.851	0.679	0.734
flow cytometry	0.887	0.919	0.872	0.892
plasma luminex	0.715	0.769	0.682	0.710
freesurfer	0.786	0.851	0.737	0.785
overall brain microstructure	0.672	0.824	0.500	0.618
localized brain microstructure	0.628	0.686	0.605	0.637
brain volumetry	0.701	0.739	0.737	0.731
All side views	0.972	1.000	0.949	0.973

the lower bound of gSide.

D. Effects of Side Views

In this section, we first investigate contributions from different side views. Table I shows performance of gMSV on the fMRI dataset by considering only one side view each time. In general, the best performance is achieved by simultaneously exploring all side views. Specifically, we observe that the side view *flow cytometry* can independently provide the most informative side information for selecting discriminative subgraph patterns on the fMRI brain networks. This is plausible as it implies that HIV brain alterations in terms of functional connectivity are most likely to express from this side view (*i.e.*, in measures of immune function, the HIV hallmark). It is consistent with our finding in Section III-B that the side view *flow cytometry* is the most significantly correlated with the prespecified label information. Similar results on the DTI dataset are available in [4].

VI. CONCLUSION

We presented an approach for selecting discriminative subgraph features using multiple side views. By leveraging available information from multiple side views together with graph data, the proposed method gMSV can achieve very good performance on the problem of feature selection for graph classification, and the selected subgraph patterns are relevant to disease diagnosis. This approach has broad applicability for yielding new insights into brain network alterations in neurological disorders and for early diagnosis.

VII. ACKNOWLEDGEMENTS

This work is supported in part by NSF through grants III-1526499, CNS-1115234, and OISE-1129076, Google Research Award, the Pinnacle Lab at Singapore Management University, and NIH through grant R01-MH080636.

REFERENCES

[1] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall. Learning a mahalanobis metric from equivalence constraints. *Journal of Machine Learning Research*, 6(6):937–965, 2005.

[2] B. Cao, L. He, X. Kong, P. S. Yu, Z. Hao, and A. B. Ragin. Tensor-based multi-view feature selection with applications to brain diseases. In *ICDM*, pages 40–49. IEEE, 2014.

[3] B. Cao, X. Kong, C. Kettering, P. S. Yu, and A. B. Ragin. Determinants of HIV-induced brain changes in three different periods of the early clinical course: A data mining analysis. *NeuroImage: Clinical*, 2015.

[4] B. Cao, X. Kong, J. Zhang, P. S. Yu, and A. B. Ragin. Mining brain networks using multiple side views for neurological disorder identification. *arXiv*, 2015.

[5] B. Cao, L. Zhan, X. Kong, P. S. Yu, N. Vizueta, L. L. Altshuler, and A. D. Leow. Identification of discriminative subgraph patterns in fMRI brain networks in bipolar affective disorder. In *Brain Informatics and Health*. Springer, 2015.

[6] C. Chang and C. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[7] W. Dai, G. Xue, Q. Yang, and Y. Yu. Co-clustering based classification for out-of-domain documents. In *KDD*, pages 210–219. ACM, 2007.

[8] C. Gao and J. Wang. Direct mining of discriminative patterns for classifying uncertain data. In *KDD*, pages 861–870. ACM, 2010.

[9] N. Jin and W. Wang. LTS: Discriminative subgraph mining by learning from search history. In *ICDE*, pages 207–218. IEEE, 2011.

[10] N. Jin, C. Young, and W. Wang. Graph classification based on pattern co-occurrence. In *CIKM*, pages 573–582. ACM, 2009.

[11] N. Jin, C. Young, and W. Wang. GAIA: graph classification using evolutionary computation. In *SIGMOD*, pages 879–890. ACM, 2010.

[12] X. Kong, A. B. Ragin, X. Wang, and P. S. Yu. Discriminative feature selection for uncertain graph classification. In *SDM*, pages 82–93. SIAM, 2013.

[13] X. Kong and P. S. Yu. Multi-label feature selection for graph classification. In *ICDM*, pages 274–283. IEEE, 2010.

[14] X. Kong and P. S. Yu. Semi-supervised feature selection for graph classification. In *KDD*, pages 793–802. ACM, 2010.

[15] L. Mihalkova, T. Huynh, and R. J. Mooney. Mapping and revising markov logic networks for transfer learning. In *AAAI*, volume 7, pages 608–614, 2007.

[16] L. Mihalkova and R. J. Mooney. Transfer learning from minimal target data by mapping across relational domains. In *IJCAI*, volume 9, pages 1163–1168, 2009.

[17] A. B. Ragin, H. Du, R. Ochs, Y. Wu, C. L. Sammet, A. Shoukry, and L. G. Epstein. Structural brain alterations can be detected early in HIV infection. *Neurology*, 79(24):2328–2334, 2012.

[18] X. Shi, X. Kong, and P. S. Yu. Transfer significant subgraphs across graph databases. In *SDM*, pages 552–563. SIAM, 2012.

[19] W. Tang and S. Zhong. Pairwise constraints-guided dimensionality reduction. In *SDM Workshop on Feature Selection for Data Mining*, 2006.

[20] X. Yan, H. Cheng, J. Han, and P. S. Yu. Mining significant graph patterns by leap search. In *SIGMOD*, pages 433–444. ACM, 2008.

[21] X. Yan and J. Han. gspan: Graph-based substructure pattern mining. In *ICDM*, pages 721–724. IEEE, 2002.