Brain Imaging and Behavior (2019) 13:879–892
https://doi.org/10.1007/s11682-018-9899-8

ORIGINAL RESEARCH

Identifying disease-related subnetwork connectome biomarkers by sparse hypergraph learning

Chen Zu[1,2] · Yue Gao[3] · Brent Munsell[4] · Minjeong Kim[5] · Ziwen Peng[6] · Jessica R. Cohen[7] · Daoqiang Zhang[1] · Guorong Wu[2]

Published online: 14 June 2018

Abstract

The functional brain network has gained increased attention in the neuroscience community because of its ability to reveal the underlying architecture of human brain. In general, majority work of functional network connectivity is built based on the correlations between discrete-time-series signals that link only two different brain regions. However, these simple region-to-region connectivity models do not capture complex connectivity patterns between three or more brain regions that form a connectivity subnetwork, or *subnetwork* for short. To overcome this current limitation, a hypergraph learning-based method is proposed to identify subnetwork differences between two different cohorts. To achieve our goal, a hypergraph is constructed, where each vertex represents a subject and also a hyperedge encodes a subnetwork with similar functional connectivity patterns between different subjects. Unlike previous learning-based methods, our approach is designed to jointly optimize the weights for all hyperedges such that the learned representation is in consensus with the distribution of phenotype data, i.e. clinical labels. In order to suppress the spurious subnetwork biomarkers, we further enforce a sparsity constraint on the hyperedge weights, where a larger hyperedge weight indicates the subnetwork with the capability of identifying the disorder condition. We apply our hypergraph learning-based method to identify subnetwork biomarkers in Autism Spectrum Disorder (ASD) and Attention Deficit Hyperactivity Disorder (ADHD). A comprehensive quantitative and qualitative analysis is performed, and the results show that our approach can correctly classify ASD and ADHD subjects from normal controls with 87.65 and 65.08% accuracies, respectively.

Keywords Hypergraph learning · Brain network · Biomarker · Autism spectrum disorder · Attention deficit hyperactivity disorder

Introduction

Human brain can be partitioned into different regions according to various functions (Van Den Heuvel and Pol 2010). An integrated network is composed of different brain regions and the information flows that are continuously processed between those functionally linked brain regions. In order to understand the pathological underpinnings of a neurological disorder, many functional neuroimaging studies have been developed to investigate abnormal alterations among brain con-

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11682-018-9899-8) contains supplementary material, which is available to authorized users.

✉ Daoqiang Zhang
dqzhang@nuaa.edu.cn

✉ Guorong Wu
grwu@med.unc.edu

[1] Department of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China

[2] Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[3] School of Software, Tsinghua University, Beijing, China

[4] Department of Computer Science, College of Charleston, Charleston, SC, USA

[5] Department of Computer Science, University of North Carolina, Greensboro, NC, USA

[6] Centre for Studies of Psychological Application, School of Psychology, South China Normal University, Guangzhou, China

[7] Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

nections. Recently, diagnosis of brain disease at individual level with functional connectivity patterns has gained attention in computer-assisted diagnosis (Zeng et al. 2012).

It is well known that the complex and oscillatory activities behind cognition are essentially the large-scale collaborative work among millions of neurons through multiple brain regions. However, the bivariate region-to-region interaction (link-wise connectivity) is not sufficient to capture the characteristics of functional connectivity involving three or more brain regions that form a connectivity subnetwork, or *subnetwork* for short. Recently, there is overwhelming evidence that brain networks display hierarchical modularity, making the investigation of biomarkers beyond link-wise connectivity more attractive to neuroscience and clinical practice than ever before.

In this paper, we propose a novel learning-based method to discover the subnetwork biomarkers that are able to distinguish two clinical cohorts and also diagnose brain disorders at individual level. Without doubt, there are thousands of subnetworks varying as the number and combination of involved brain regions. Considering the computational cost, we only investigate a 3-node subnetwork that is the simplest types of subnetwork. Intuitively, two criteria are used to select a subnetwork:

1. Discriminative power across clinical groups. The entire functional connectivity inside the subnetwork, instead of the particular predominant connection link, shows significant difference between two clinical cohorts.
2. Consistency within each clinical group. The characteristic of functional connectivity inside the subnetwork should be similar for two subjects that fall in the same clinical group.

The naive solution is to calculate both the discriminative power and consistency measurement for each subnetwork via independent statistical *t*-test. Since the subnetworks are highly correlated (e.g., existing a large amount of overlap of edges among brain regions), independent statistical test can be hardly effective in identifying the most influential subnetworks that may be related to the brain disorder. In light of this, we propose a novel learning-based method to jointly find a set of subnetwork biomarkers where the combination of selected biomarkers meets the above criteria with low redundancy.

To achieve this goal, we resort to hypergraph technique for identifying the complex subject-to-subject relationships based on the subnetwork connectivity inside all possible subnetwork combinations. We treat each subject as a vertex in the hypergraph. The *star-expansion* strategy is involved in constructing hyperedges. Specifically, for each subnetwork type, a vertex chosen as the central node and its *k*-nearest neighbors compose a hyperedge. Thus, the hypergraph eventually encodes a wide spectrum of subject-to-subject relationships in the population. The next step is to identify the subnetworks related to the brain disorder. Since each subject is assigned a clinical label, it now becomes an optimization problem that learns a set of weights (one for each hyperedge) to create partitions in the hypergraph that maximally agrees with the observed clinical labels. Hence, the learned weights reflect the significance in distinguishing two clinical cohorts. Since we construct hyperedges for all possible subnetworks in an exhaustive way, we further apply sparse constraint to suppress spurious subnetworks.

Our proposed learning-based method is close to the network motif study (Sporns and Kötter 2004; Milo et al. 2002), which seeks for a recurrent subnetwork that occurs more often than in a pre-defined and randomly organized benchmark network with the same degree distribution. However, our work has the following advantages over the network motifs.

1. Our proposed method is a learning-based approach, where the objective function is aimed to estimate the influence of each subnetwork in separating two clinical groups. On the contrary, the current work on the motif highly relies on the pre-defined benchmark network and only counts for the occurrence of each motif, which cannot be used as biomarker for calssification.
2. Our method considers the dependency among subnetworks. Thus, our method jointly selects a small number of subnetwork biomarkers such that the combination of these selected subnetwork biomarkers has the highest sensitivity and lowest redundancy. To the best of our knowledge, the current motif methods select each motif independently.
3. Our method is fully data-driven and free of presumptions. Specifically, we use the phenotype information (clinical labels) to guide the selection of subnetworks such that it is straightforward to interpret the biological meanings of selected subnetworks. However, the current motif selection methods have the issue of selecting motifs that are not consistent with real network since random rewiring in constructing benchmark network removes topological information in real brain network (Sporns and Kötter 2004).

Our learning-based method is applied to discover subnetwork biomarkers that can identify childhood autism spectrum disorder (ASD) and attention deficit hyperactivity disorder (ADHD) subjects, respectively. Promising classification results have been achieved, which demonstrate the power of the learned subnetwork patterns.

The rest of this paper is organized as follows. We first introduce the background of hypergraph learning in second section. Then, we present our learning-based method to find the significant subnetwork biomarkers in third section. After that, we apply the method to the two public imaging databases (ABIDE and ADHD-200), and present the comparison results to validate the advantages of our method in forth section. Finally, we conclude our method in fifth section.

Related work

In recent years, hypergraph learning has been utilized in many applications for its merit of exploring complex sample relationships. For example, Zhou et al. (Zhou et al. 2006) introduced a general hypergraph learning framework for clustering, classification, and embedding. A hypergraph Laplacian is proposed to completely present the complex relationships among subjects. Hypergraph learning also has many successful applications in computer vision and medical imaging area. To name a few, Gao et al. have applied hypergraph learning technique for 3-D object retrieval and recognition in (Gao et al. 2012). Yu et al. (2012) proposed an adaptive hypergraph learning method for transductive image classification by adjusting the weights of hyperedges. Furthermore, Huang et al. (Huang et al. 2010) proposed a transductive learning framework for image retrieval. In their method, each image is taken as a vertex in a weighted hypergraph and the task of image retrieval is treated as a problem of hyperedge ranking. In medical imaging area, hypergraph learning was used for classifying gene express in (Tian et al. 2009) and identifying MCI (mild cognitive impairment) and AD (Alzheimer's disease) patients from normal controls using neuroimaging data in (Gao et al. 2015a; b). Some studies (Davison et al. 2015; Jie et al. 2016) give the connections between the neuroscience and the hypergraph. For example, (Davison et al. 2015) finds the groups of brain functional interactions that fluctuate coherently in strength over time instead of dyadic (region-to-region) relationships. In (Jie et al. 2016), a hyper-connectivity network of brain functions is constructed by hypergraph technique. By extracting features from the hyper-connectivity brain network, the better classification performance can be achieved on MCI and ADHD dataset. Other interesting applications of hypergraph can be found in (Zass and Shashua 2008; Bu et al. 2010; Sun et al. 2008; Yu et al. 2014; Huang et al. 2011; Agarwal et al. 2005; Zhang et al. 2014).

In general, most works on hypergraph learning are built on the hypergraph Laplacian proposed in (Zhou et al. 2006). Similarly, our method considers each subject as a vertex in the hypergraph and encodes complex subject-to-subject relationships via a set of hyperedges. Then, we leverage the optimized hyperedge weights to reflect the importance of subnetwork architectures. To be clear, we first briefly explain the principle of hypergraph leaning in "Introduction of hypergraph learning" section and then present our learning-based method for identification of subnetwork biomarkers in "Our method" section.

Introduction of hypergraph learning

In the conventional graph technique, one edge only can connect two related vertices at a time. Graph can be undirected or directed which depends on whether the pairwise relationships among vertices is symmetric (Zhou et al. 2006). Since graph is an efficient tool to represent data distribution, it has been widely used in manifold learning (Zhang et al. 2012) and transductive learning (Chapelle et al. 2009). However, in many real-world problems, relationships among the objects are much more complex than pairwise. Simply projecting the complex relationships into pairwise links will inevitably lead to the loss of information which could be valuable in many learning tasks (Zhou et al. 2006). On the contrary, hypergraph is a generalization of traditional graph in which each hyperedge can connect any number of vertices. It is worth noting that a simple graph is just a special case of hypergraph, where each edge allows linking only two vertices. For convenience, Table 1 summarizes several important notations and their definitions for hypergraph.

A hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{w})$ is formed by a vertex set $\mathcal{V}$, a hyperedge set $\mathcal{E}$, and a hyperedge weight $\boldsymbol{w}$. Each hypergraph e_θ is assigned a weight $w(\theta)$ ($\theta = 1, \ldots, |\mathcal{E}|$). The hypergraph $\mathcal{G}$ can be represented by a $|\mathcal{V}| \times |\mathcal{E}|$ incidence matrix $\mathbf{H}$ with the elements defined by

$$H(v,\theta) = \begin{cases} 1, \text{if } v \in e_\theta \\ 0, \text{if } v \notin e_\theta. \end{cases}$$

For a vertex $v \in \mathcal{V}$, its degree is defined based on $\mathbf{H}$ by

$$d(v) = \sum_{\theta=1}^{|\mathcal{E}|} w(\theta) H(v,\theta).$$

Similarly, the edge degree of hyperedge $e \in \mathcal{E}$ is

$$\delta(\theta) = \sum_{v \in V} H(v,\theta)$$

$\mathbf{D}_v$ and $\mathbf{D}_e$ denote the diagonal matrices of the vertex degrees and the hyperedge degrees, respectively. Let $\mathbf{D}_w$ denote the $|\mathcal{E}| \times |\mathcal{E}|$ diagonal matrix of the hyperedge weights.

Hypergraph learning has been applied to various applications such as image classification, clustering, and embedding. Here, we use the binary classification problem as an example to explain the principle of hypergraph learning. Suppose $y(v)$ returns the label of a given vertex v, where $y(v) = +1$ and $y(v) = -1$ denote the positive and negative training subjects, respectively. Since the labels on the testing data are not known yet, we usually set the values of their respective label as 0. On the other hand, f is a classification function, which returns a label confidence $f(v)$ ($-1 \le f(v) \le 1$) on each vertex $v \in \mathcal{V}$, where the underlying vertex v is assigned to positive sample if $0 < f(v) \le 1$. Otherwise, vertex v is considered as negative subject. After we stack $f(v)$ and $y(v)$ into the column vectors $\vec{\mathbf{f}}$ and $\vec{\mathbf{y}}$, the objective function can be defined as:

$$\arg\min_{\vec{\mathbf{f}}} \Omega_{\mathbf{f}} + \lambda R_{emp}\left(\vec{\mathbf{f}}\right). \quad (1)$$

Table 1 Notations and definitions

Notation	Definition
$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{w})$	$\mathcal{G}$ denotes a hypergraph, and $\mathcal{V}$ and $\mathcal{E}$ indicate the vertex set and hyperedge set, respectively; and $\boldsymbol{w}$ represents the weights of the hyperedges.
$\mathcal{V}$	The set of vertices of the hypergraph $\mathcal{G}$, containing N elements.
$\mathcal{E}$	The set of edges of the hypergraph $\mathcal{G}$, which contains Θ items.
N	The number of subjects in the data, i.e., $\lvert\mathcal{V}\rvert$.
Θ	The number of hyperedges in the hypergraph, i.e., $\lvert\mathcal{E}\rvert$.
$\mathbf{H}$	The incidence matrix of the hypergraph.
$d(v)$	The degree of the vertex v.
$\delta(e)$	The degree of the hyperedge e.
$\mathbf{D}_v$	The diagonal matrix of the vertex degrees.
$\mathbf{D}_e$	The diagonal matrix of the edge degrees.
$\mathbf{D}_w$	The diagonal matrix of the hyperedge weights, with its (θ, θ)-th entry as $w(\theta)$.
$\vec{\mathbf{y}} = [y_1, \ldots, y_N]^T$	The $N \times 1$ label vector for hypergraph learning.
$\vec{\mathbf{f}} = [f_1, \ldots, f_N]^T$	The $N \times 1$ to-be-learned label.
L	The Laplacian matrix of the hypergraph.

where $\Omega_{\mathbf{f}}$ is a regularizer on the constructed hypergraph (the hypergraph construction approach will be detailed in "Method overview" section), and $R_{emp}\left(\vec{\mathbf{f}}\right) = \left\|\vec{\mathbf{y}} - \vec{\mathbf{f}}\right\|_2^2$ is an empirical loss. $\lambda > 0$ is a weighting parameter used to balance the two terms in Eq. (1). Specifically, the hypergraph regularization term $\Omega_{\mathbf{f}}$ is defined as:

$$\Omega_{\mathbf{f}} = \frac{1}{2}\sum_{\theta=1}^{|\mathcal{E}|}\sum_{u,v\in\mathcal{V}}\frac{w(\theta)H(u,\theta)H(v,\theta)}{\delta(\theta)}\left(\frac{f(u)}{\sqrt{d(u)}} - \frac{f(v)}{\sqrt{d(v)}}\right)^2. \quad (2)$$

The intuition behind Eq. (2) is to encourage the normalized label probabilities of any two vertices u and v to be similar as long as they fall into the same hyperedge. By letting $\boldsymbol{\Lambda} = \mathbf{D}_v^{-\frac{1}{2}}\mathbf{H}\mathbf{D}_w\mathbf{D}_e^{-1}\mathbf{H}^T\mathbf{D}_v^{-\frac{1}{2}}$, and $\mathscr{L} = \mathbf{I} - \boldsymbol{\Lambda}$, the normalized regularization term can be rewritten as:

$$\Omega_{\mathrm{f}} = \vec{\mathbf{f}}^T \mathscr{L} \vec{\mathbf{f}}. \quad (3)$$

where $\mathscr{L}$ is a positive semidefinite matrix, called *hypergraph Laplacian*. The objective function in Eq. (1) has a closed-form solution:

$$\hat{\vec{\mathbf{f}}} = \left(\mathbf{I} + \frac{1}{\lambda}\mathscr{L}\right)^{-1}\vec{\mathbf{y}}. \quad (4)$$

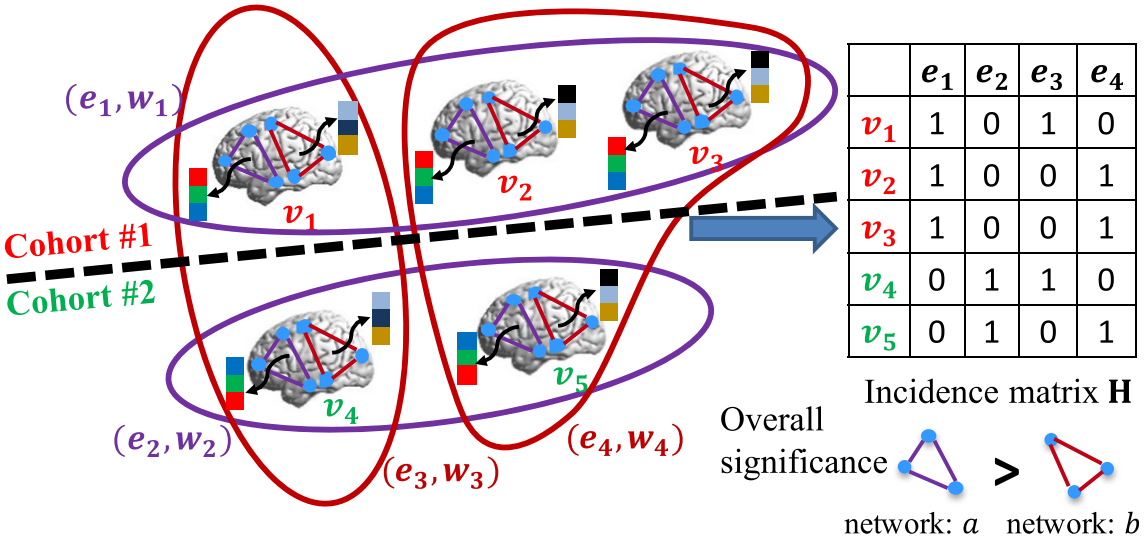
It is worth noting that the hypergraph regularization term $\Omega_{\mathbf{f}}$ is the function of hyperedge weights $\{w(\theta)\}$, although we assume that the hyperedge weights are identical in the above example. As we will explain in "Encoding complex subject-wise relationship in hypergraph" section, each hyperedge is associated with one particular subnetwork. Thus, it is reasonable to optimize both the label probability $\vec{\mathbf{f}}$ and hyperedge weights $\{w(\theta)\}$ in the objective function of the standard hypergraph learning. Hence, the optimized hyperedge weights $\{w(\theta)\}$ respect *not only* the distribution of phenotype data $\vec{\mathbf{y}}$, *but also* the local embedding of functional behaviors encoded in the hypergraph $\mathcal{G}$.

Our method

Method overview

Figure 1 illustrates the intuition behind our proposed learning-based method. For clarity, we assume three subjects in one cohort (top-left in Fig. 1) and two subjects in another cohort (bottom-left in Fig. 1). Only two possible subnetwork patterns (purple and red triangles) are under investigation in this example. So, the goal is to find out which subnetwork is able to separate subjects from two cohorts more accurately than others, based on the functional connectivity flow inside the subnetwork pattern. Eventually, the selected subnetworks are considered as biomarkers to identify other individual subjects.

Hypergraph is employed to measure the complex subject-wise relationships based on the functional connectivity flow running inside each subnetwork pattern. Specifically, subjects are considered as vertices (v_i, $i = 1, 2, \ldots, 5$ in Fig. 1) in the hypergraph. In general, a set of subjects fall into the same hyperedge only if the difference of the overall functional connectivity inside the same subnetwork is small. Thus, hyperedge can accommodate the complex relationship beyond two subjects. For example, subject v_2 and v_3 stay in the same hyperedge e_1 with v_1 since their functional connectivity flows (designated by the black arrows) are very similar inside



	e_1	e_2	e_3	e_4
v_1	1	0	1	0
v_2	1	0	0	1
v_3	1	0	0	1
v_4	0	1	1	0
v_5	0	1	0	1

Fig. 1 The overview of our learning-based method in discovering complex subject relation patterns by hypergraph

the purple subnetwork. The standard way to construct hyperedges is to exhaustively visit each subject per each subnetwork. As shown in Fig. 1, we can obtain four hyperedges (e_1-e_4) as indicated by curves. Note that the identical hyperedges are discarded and the color on each hyperedge indicates the associated subnetwork.

It is obvious that the hypergraph has comprehensively encoded a wide spectrum of subject-wise relationships for various subnetworks. A hypergraph learning technique is used to jointly quantify the significance of each subnetwork based on the ground-truth clinical label on each subject. Intuitively, the more label discrepancies occur within the hyperedges related to the underlying subnetwork pattern, the lower the significance of that particular subnetwork becomes. Finally, the subnetwork patterns with high overall significance value across related hyperedges are regarded as the biomarkers from rs-fMRI image. As shown in the left panel of Fig. 1, the labels of vertices in e_1 and e_2 (purple curves) are highly consistent, suggesting that the functional connectivity flow running on the purple subnetwork pattern is a good biomarker to separate subjects from two different cohorts. In contrast, the functional connectivity flow inside the red subnetwork pattern fails to be the biomarker since the hyperedges built on the red subnetwork pattern have subjects with different clinical labels, e.g., v_1 and v_4 belonging to different categories are both included in the hyperedge e_3.

Encoding complex subject-wise relationship in hypergraph

Given a training set of N subjects, where each subject has already been partitioned to R anatomical regions. Without loss of generality, we use '+1' and '-1' to distinguish the label for two clinical cohorts, and thus form a column vector $\vec{\mathbf{y}} = [y_1, y_2, \ldots, y_N]^T$. Considering the computational cost and efficiency, we first construct the pool of all possible subnetworks $\Delta = \{\Delta_j | j = 1, \ldots, C\}$, where each subnetwork pattern Δ_j consists of three brain regions randomly picked up from the total R regions. Therefore, there are $C = \binom{R}{3}$ subnetworks in total. Given a subject v_n and a particular subnetwork Δ_j, we can obtain a three-element vector of functional connectivity flow $\boldsymbol{\alpha}_{j,n} = \left[\alpha^1_{j,n}, \alpha^2_{j,n}, \alpha^3_{j,n}\right]$, where each element in $\boldsymbol{\alpha}_{j,n}$ is the Pearson's correlation coefficient of the mean rs-fMRI signals, from subject v_n, between any two brain regions within the subnetwork Δ_j.

Next, we construct hypergraph, as denoted by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the hypergraph vertex set $\mathcal{V} = \{v_n | n = 1, \ldots, N\}$ includes all subjects in the population with the known clinical labels. We use *star-expansion* strategy to build a set of hyperedges by exhaustively visiting each vertex v_n for each particular subnetwork Δ_j, thus forming the hyperedge set $\mathcal{E} = \{e_{j,n} | j = 1, \ldots, C, n = 1, \ldots, N\}$. Specifically, a subject v_n is treated as a centroid vertex. For subnetwork Δ_j, we examine the distance between functional connectivity flow $\boldsymbol{\alpha}_{j,n}$ at current vertex v_n and $\boldsymbol{\alpha}_{j,n'}$ ($n' = 1, \ldots, N, n' \neq n$) at all other vertices. If $\boldsymbol{\alpha}_{j,n'}$ is within the k-nearest neighborhood of the centroid vertex, the subject $v_{n'}$ is included in the hyperedge $e_{j,n}$ (i.e., $v_{n'} \in e_{j,n}$). Thus, each hyperedge consists of $(k+1)$ vertices.

Since each hyperedge $e_{j,n}$ is related with both vertex v_n and subnetwork Δ_j, we use the index θ to delegate the bivariate index (j, n), i.e., $\theta \leftrightarrow (j, n)$, where θ ranges from 1 to $\Theta = C \times N$. Thus, $\mathbf{H}$ is a $N \times \Theta$ matrix. For each entry $H(n, \theta)$, we set $H(n, \theta) = 1$, if the vertex v_n is contained in hyperedge $e(j, n)$. Otherwise, $H(n, \theta) = 0$. The example of the incidence matrix is shown in the right panel of Fig. 1. Apparently, the incidence matrix conveys more information than the affinity matrix used in the conventional approaches based on simple graphs. Furthermore, we construct a $C \times N$ matrix $\mathbf{W}$ where each

element $w_{j,n}$ denotes the weight of hyperedge $e_{j,n}$. The weight of the j-th subnetwork is defined as $\sum_{n=1}^{N} w_{j,n}$ Since each column of incidence matrix $\mathbf{H}$ encodes the connected vertices inside a particular hyperedge, each $w_{j,n}$ has one-to-one correspondence to the θ-th column vector in $\mathbf{H}$.

Discovering subnetwork biomarkers by sparse hypergraph learning

Our learning-based method aims to find out the subnetwork by inspecting the performance of each hyperedge in separating subjects from two cohorts. To this end, we first assume that the label on each subject is not known yet. Thus, we use hypergraph learning technique to estimate the likelihood f_n for each subject v_n, which is driven by (a) the minimization of discrepancies between the ground-truth label vector $\vec{\mathbf{y}}$ and the estimated likelihood vector $\vec{\mathbf{f}} = [f_1, f_2, \ldots, f_N]^T$, and (b) the consistency of clinical labels within each hyperedge. Similar to Eq. (2), the consistency requirement can be defined as:

$$\Omega_{\mathbf{f}}(\mathbf{W}) = \frac{1}{2}\sum_{j=1}^{C}\sum_{n=1}^{N}\sum_{n'=1}^{N}\frac{w_{j,n}H(n,\theta)H(n',\theta)}{\delta(\theta)}\left(\frac{f_n}{\sqrt{d(n)}}-\frac{f_{n'}}{\sqrt{d(n')}}\right)^2. \tag{5}$$

The regulation term $\Omega_{\mathbf{f}}(\mathbf{W})$ penalizes the label discrepancy by encouraging the difference between the normalized likelihoods $f_n/\sqrt{d(n)}$ and $f_{n'}/\sqrt{d(n')}$ to be as small as possible if v_n and $v_{n'}$ are in the same hyperedge e_θ. It is clear that the regularization term $\Omega_{\mathbf{f}}(\mathbf{W})$ is a function of both $\mathbf{W}$ and $\vec{\mathbf{f}}$, which eventually makes the optimization of $\mathbf{W}$ reflecting the quality of each hyperedge as a biomarker.

Of note, we construct N hyperedges in total for each possible subnetwork, where each subject in turn acts as a central vertex. In order to make more sense on the selected subnetwork biomarkers, we go one step further to enforce sparsity on the subnetwork, i.e., most rows in $\mathbf{W}$ are null vectors (with all elements in the row vector as zeroes). We encourage the selected subnetwork working consistently across subjects, in order to suppress the predominant weights on a few subjects misleading the learning of subnetwork biomarkers. It is obvious that the integration of the above two criteria falls into the typical $\ell_{2,1}$ regularization (Argyriou et al. 2008). Thus, we further derive our objective function of learning subnetwork biomarkers as:

$$\arg\min_{\mathbf{W},\vec{\mathbf{f}}} \Omega_{\mathbf{f}}(\mathbf{W}) + \lambda\left\|\vec{\mathbf{y}}-\vec{\mathbf{f}}\right\|_2^2 + \mu\sum_{j=1}^{C}\sqrt{\sum_{n=1}^{N} w_{j,n}^2}. \tag{6}$$

where λ and μ are two positive scalars controlling the strength of data fitting term and the $\ell_{2,1}$-norm on the weighting matrix $\mathbf{W}$, respectively. Term $\left\|\vec{\mathbf{y}}-\vec{\mathbf{f}}\right\|_2^2$ guarantees that the new generated results $\vec{\mathbf{f}}$ are not far away from the ground-truth label information. We use Fig. 2 to demonstrate the effectiveness of $\ell_{2,1}$-norm. Figure 2 visualizes the matrix $\mathbf{W}$, in which the elements with deep blue colors have large values. The $\ell_{2,1}$-norm accentuates the individual weight learning across multiple types of hyperedge. Through the joint sparsity-inducing norm, many weights of the hyperedges constructed via important subnetworks will learn large values and a small number of weights are in the hyperedges based on irrelevant subnetworks.

Optimization

Since the objective function in Eq. (6) is convex only *with respect to* (*w.r.t.*) variables $\mathbf{W}$ and $\vec{\mathbf{f}}$ separately, we propose the following solution to optimize $\mathbf{W}$ and $\vec{\mathbf{f}}$, alternatively.

First, we fix $\mathbf{W}$ and optimize $\vec{\mathbf{f}}$, and then the objective function becomes:

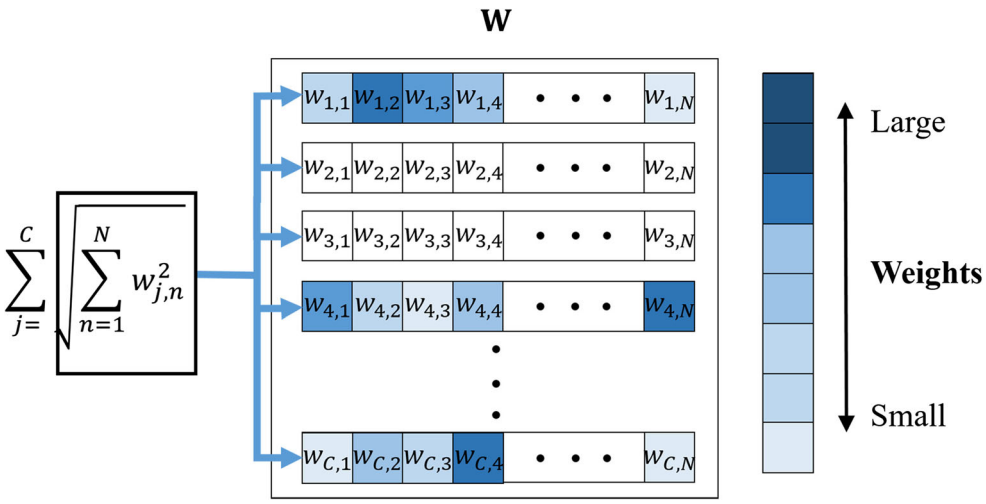


Fig. 2 Illustration of the hyperedge weight matrix $\mathbf{W}$

$$\arg\min_{\vec{\mathbf{f}}}\phi\left(\vec{\mathbf{f}}\right)=\vec{\mathbf{f}}^{T}\,\mathscr{L}\,\vec{\mathbf{f}}+\lambda\left\|\vec{\mathbf{y}}-\vec{\mathbf{f}}\right\|_2^2. \tag{7}$$

The conventional hypergraph inference method can be used to estimate $\vec{\mathbf{f}}$ by letting $\frac{\partial\phi}{\partial\mathbf{f}}=0$, which leads to the deterministic solution: $\hat{\vec{\mathbf{f}}}=\left(\mathbf{I}+\frac{1}{\lambda}L\right)^{-1}\vec{\mathbf{y}}$.

After obtaining $\vec{\mathbf{f}}$, we can optimize $\mathbf{W}$ with $\vec{\mathbf{f}}$ fixed. After discarding the unrelated terms w.r.t. $\mathbf{W}$ in Eq. (6), we derive the objective function for hypergraph weight as:

$$\arg\min_{\mathbf{W}}\varphi(\mathbf{W})=\vec{\mathbf{f}}^{T}L\,\vec{\mathbf{f}}+\mu\sum_{j=1}^{C}\sqrt{\sum_{n=1}^{N}w_{j,n}^2}. \tag{8}$$

Taking the derivative of the objective w.r.t. $w_{j,\,n}$ and also setting it to zero, we obtain:

$$-\delta(\theta)^{-1}\Gamma_\theta^2+\mu b_j w_{j,n}=0. \tag{9}$$

where Γ_θ denotes the θ-th column of the $\Theta\times\Theta$ matrix $\mathbf{\Gamma}=\mathbf{H}^T\mathbf{D}_v^{-\frac{1}{2}}\mathbf{f}$ and $b_j=\left(\sum_j w_{j,n}^2\right)^{-\frac{1}{2}}$. Thus, we have:

$$w_{j,n}=\frac{\delta(\theta)^{-1}\Gamma_\theta^2}{\mu b_j}. \tag{10}$$

Note that b_j is a latent variable which depends on the estimation of other entries in $\mathbf{W}$. Thus, our solution turns to an iterative manner, as summarized in **Algorithm 1**.

Features for classification

A specific subnetwork $\Delta_j$with discriminative ability is chosen after hypergraph learning. Then given a subject v_n, we can obtain a three-element vector of functional connectivity flow $\boldsymbol{\alpha}_{j,n}=\left[\alpha_{j,n}^1,\alpha_{j,n}^2,\alpha_{j,n}^3\right]$, where each element in $\boldsymbol{\alpha}_{j,\,n}$ is the Pearson's correlation coefficient of the mean rs-fMRI signals, from subject v_n, between any two brain regions within the subnetwork Δ_j. In this work, traditional Support Vector Machine (SVM) (Cortes and Vapnik 1995) and the Support Tensor Machine (STM) (Tao et al. 2005) are adopted for classification. Suppose we have m subnetworks. For classification with SVM, the functional connectivity flows are concatenated as a long feature vector $\boldsymbol{\alpha}_n=[\boldsymbol{\alpha}_{1,\,n},\boldsymbol{\alpha}_{2,\,n},\ldots,\boldsymbol{\alpha}_{m,\,n}]$ for subject v_n and the total number of the elements in $\boldsymbol{\alpha}_n$ is $3\times m$. For classification with STM the functional connectivity flows are stacked as a feature matrix $\boldsymbol{\alpha}_n=[\boldsymbol{\alpha}_{1,\,n};\boldsymbol{\alpha}_{2,\,n};\ldots;\boldsymbol{\alpha}_{m,\,n}]\in R^{3\times m}$ for subject v_n. Finally, for all subjects these feature vectors and feature matrices are combined as a feature matrix and a feature tensor for SVM and STM, respectively.

Algorithm 1 An iterative algorithm to solve the optimization problem in Eq. (6)

Input: Vertex set $\mathcal{V}=\{v_n|n=1,\ldots,N\}$, subnetwork $\Delta=\{\Delta_j|j=1,\ldots,C\}$ and ground-truth clinical label $\vec{\mathbf{y}}=[y_1,\ldots,y_N]^T$.

1. Construct hypergraph via the *star-expansion* strategy. Initialize $\mathbf{W}$ and obtain $\mathbf{H},\mathbf{D}_v,\mathbf{D}_e,\mathbf{D}_w$ and $\boldsymbol{\mathcal{L}}$.

While not converge **do**

2. Calculate $\vec{\mathbf{f}}^{(t+1)}=\left(\mathbf{I}+\frac{1}{\lambda}\boldsymbol{\mathcal{L}}^{(t)}\right)^{-1}\vec{\mathbf{y}}$.
3. Calculate $w_{j,n}^{(t+1)}=\frac{\delta(\theta)^{-1}\Gamma_\theta^2}{\mu b_j}$ based on the latest estimation $\mathbf{W}^{(t)}$, and update $\mathbf{D}_v$ with $\mathbf{W}^{(t)}$.
4. $t=t+1$.

End while

Output: $\mathbf{W}$ and $\vec{\mathbf{f}}$.

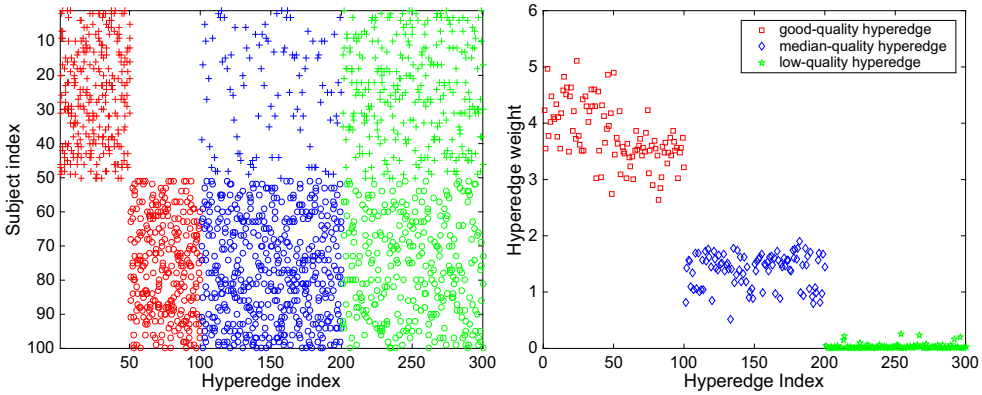


Fig. 3 A toy example demonstrating the principle of our hypergraph learning where the learned hyperedge weights reflect the quality of each subnetwork pattern being as the biomarker

Experiments

Validation of the simulated dataset

In this simulated dataset, we suppose having two groups of subjects (50 patients and 50 normal controls). We simulate three types of hyperedges: (1) good-quality hyperedges consisting of 6 subjects randomly selected from the same group (hyperedge index from 1 to 100); (2) median-quality hyperedges consisting of 5 subjects randomly selected from the same group and 1 subject from another group (hyperedge index from 101 to 200), and (3) low-quality hyperedges consisting of 3 subjects randomly selected from one group and the remaining 3 subjects randomly selected from another group (hyperedge index from 201 to 300). In the left of Fig. 3, we show the incidence matrix $\mathbf{H}$. Each column denotes one hyperedge, and the colors indicate the type of simulated hyperedges. Red, blue, and green colors represent good-quality hyperedges, median-quality hyperedges, and low-quality hyperedges, respectively. Symbols '+' and 'o' indicate the vertexes belonging to two categories. The optimized weights obtained by our sparse hypergraph learning approach are shown in the right of Fig. 3. It is clear that the values of optimized hyperedge weights are reasonable to reflect the quality of hyperedges. The hyperedges contain more subjects from the same class, then have the higher hyperedge weights. After this validation, we apply our learning-based approach to identify subnetwork biomarkers in ASD and ADHD studies, respectively, as detailed below.

Discovering subnetwork biomarkers for autism Spectrum disorder

Critical subnetworks learned by hypergraph inference

In this section, we apply our learning-based method to identify the most influential subnetworks based on 45 ASD and 47 normal control (NC) subjects from the NYU site of Autism

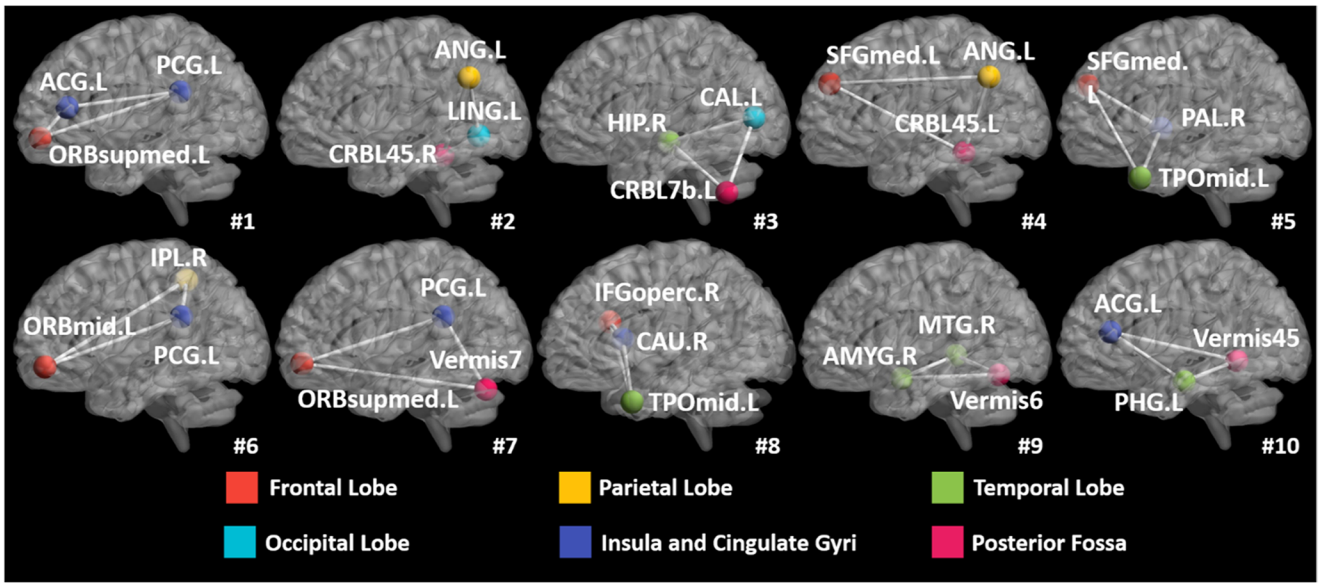


Fig. 4 The top 10 selected subnetworks (white triangle cliques), where the functional connectivity flow running inside has significant difference between ASD and NC cohorts

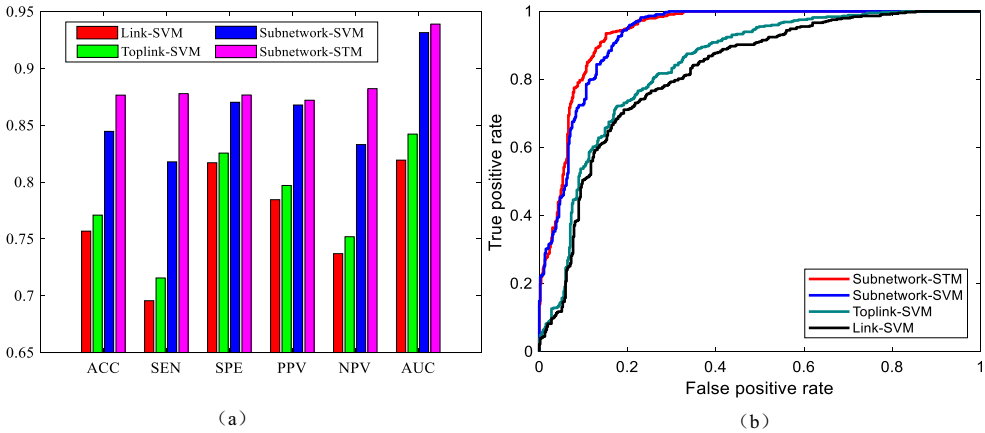


Fig. 5 Classification performance of four different classification methods on ASD dataset of NYU

Brain Imaging Data Exchange (ABIDE) database (Di Martino et al. 2014). The first 10 obtained rs-fMRI images of each subject are removed to ensure magnetization equilibrium. After slice timing and head motion correction, all images are normalized into MNI space and then segmented into 116 regions-of-interest (ROIs) according to Automated Anatomical Labeling (AAL) template (Tzourio-Mazoyer et al. 2002). Following this, the images undergo signal detrending and bandpass filtering (0.01–0.08HZ). For each subject, the mean time series of each ROI is obtained by averaging the resting state fMRI time series over all voxels in that particular ROI. Note, the total number of possible subnetworks is $\binom{116}{3} = 253{,}460$. We jointly find the best parameters for λ and μ in Eq. (6) using the line grid search strategy with the parameter values selected from the range of [10^{-3}, 10^{-2}, 10^{-1}, 1,10,10^{2}, 10^{3}] via 10-fold cross validation. Specifically, we randomly split the data into 10 subgroups. At each time, 9 subgroups are treated as training data and the left-out one is used the testing data. For selecting the parameter values, another 10-fold cross validation inside the training data is also conducted. The training data is separated into 10 parts, one part is called validation set, and the other parts are used for training the classification model. This procedure is repeated 10 times to avoid introducing the bias.

Since we randomly run the experiment 10 times to get the average results, and, for each time, the obtained top 10 subnetworks are not always the same for different parameter values. Thus, we present the top 10 most frequently appearing subnetworks. We rank these subnetworks according to their weights and Fig. 4 shows the top 10 most critical subnetworks (white triangle cliques) out of 253,460 candidates between ASD and NC cohorts. The color on each vertex differentiates the functions in human brain. It is clear that (a) most of the brain regions involved in the selected top 10 critical subnetworks locate at the key areas related with ASD, such as amygdala, middle temporal gyrus, superior frontal gyrus; and (b) most of the selected subnetworks travel across subcortical and cortical regions, which is in consensus with the recent discovery of autism pathology in neuroscience community (Minshew and Williams 2007).

Identification of ASD subjects with the learned subnetwork

In the following experiments, we use functional connectivity flows on the selected critical subnetwork patterns as feature representation (where each feature vector of subnetwork pattern is a 3-dimensional vector) to classify ASD and NC subjects. Then, the traditional Support Vector Machine (SVM) (Cortes and Vapnik 1995) is adopted to train the classifier directly based on the concatenated feature vector, denoted as *Subnetwork-SVM*. Since the functional connectivity flow comes from each subnetwork pattern, it is straightforward to organize them to a tensor representation and then use the advanced Support Tensor Machine (STM) (Tao et al. 2005) to take advantage of the structured feature representation, denoted as *Subnetwork-STM* in the following experiments. Details of SVM and STM are provided in Appendix. In order to demonstrate the advantage of subnetwork over the conventional

Table 2 Classification performance on ASD dataset of NYU. The best results are denoted in bold

Method	ACC	SEN	SPE	PPV	NPV	AUC
Link-SVM	0.7568 ± 0.1323	0.6956 ± 0.0021	0.8170 ± 0.0016	0.7845 ± 0.0016	0.7370 ± 0.0013	0.8193 ± 0.1472
Toplink-SVM	0.7709 ± 0.1339	0.7156 ± 0.0022	0.8255 ± 0.0016	0.7970 ± 0.0016	0.7519 ± 0.0014	0.8422 ± 0.1327
Subnetwork-SVM	0.8446 ± 0.1179	0.8178 ± 0.0017	0.8702 ± 0.0015	0.8678 ± 0.0015	0.8330 ± 0.0013	0.9315 ± 0.0972
Subnetwork-STM	**0.8765 ± 0.1059**	**0.8778 ± 0.0015**	**0.8766 ± 0.0015**	**0.8720 ± 0.0014**	**0.8822 ± 0.0012**	**0.9390 ± 0.0886**

Table 3 Classification results of Hypergraph learning on ASD dataset of NYU

	ACC	SEN	SPE	PPV	NPV	AUC
Hypergraph	0.7646 ± 0.1339	0.7067 ± 0.0016	0.8213 ± 0.0016	0.7910 ± 0.0016	0.7452 ± 0.0014	0.8299 ± 0.1423

region-to-region connection in brain network, we compare with two counterpart methods *Link-SVM* (using the Pearson's correlations on each link as the feature) and *Toplink-SVM* (selecting significant links by *t*-test ($p = 0.05$) and using the Pearson's correlation on the selected links to form the feature vector).

Evaluation on discrimination power In this experiment, we use 10-fold cross validation strategy to evaluate the classification accuracy (ACC), sensitivity (SEN), specificity (SPE), positive predictive value (PPV), and negative predictive value (NPV) on 45 ASD and 47 NC subjects from NYU site in ABIDE database. As shown by both the classification performance plots and the ROC curves in Fig. 5, the classifiers trained on connectome features from our learned subnetworks have achieved much higher classification performance than those trained by the same classification tools but based on the connectome features from the conventional region-to-region connection links. Also, the substantial classification improvements by *Subnetwork-STM* over *Subnetwork-SVM* indicate the benefit of using structured data presentation in classification, where such complex relation information is clearly delivered in the learned subnetworks.

Table 2 shows the classification performance on ASD dataset of NYU. The conventional method, such as Link-SVM, obtains only an accuracy of 0.7568, while our proposed methods achieve accuracies of 0.8446 and 0.8765 with SVM and STM, respectively. At the same time, other classification performance metrics also demonstrate the superiority of our proposed method. That means, the subnetwork patterns (features) selected by our method contain more discriminative information, which validates that the identified subnetworks are the more suitable ASD biomarkers.

The $\vec{\mathbf{f}}$ learned in the objective function (6) gives the predicted label of testing subjects which can be directly used to evaluate the classification power of hypergraph learning. Table 3 presents the classification performance of hypergraph learning method.

As we can see from Table 3, the Hypergraph method achieves an accuracy of 0.7646, which is superior to the accuracy obtained by link-SVM, but inferior to the accuracy got by Toplink-SVM. That is probably because the Hypergraph method can adequately utilize the importance of hyperedges. During the learning process, the hyperedges consisting of subnetworks with less discriminability will get small weights, which can enhance the performance of the classification. While the classification performance of Hypergraph method is inferior to that of Toplink-SVM, this is probably because the Toplink-SVM not only selects useful features, but also finds the optimal hyperplane to separate the two categories subjects. The similar phenomenon can be observed from other

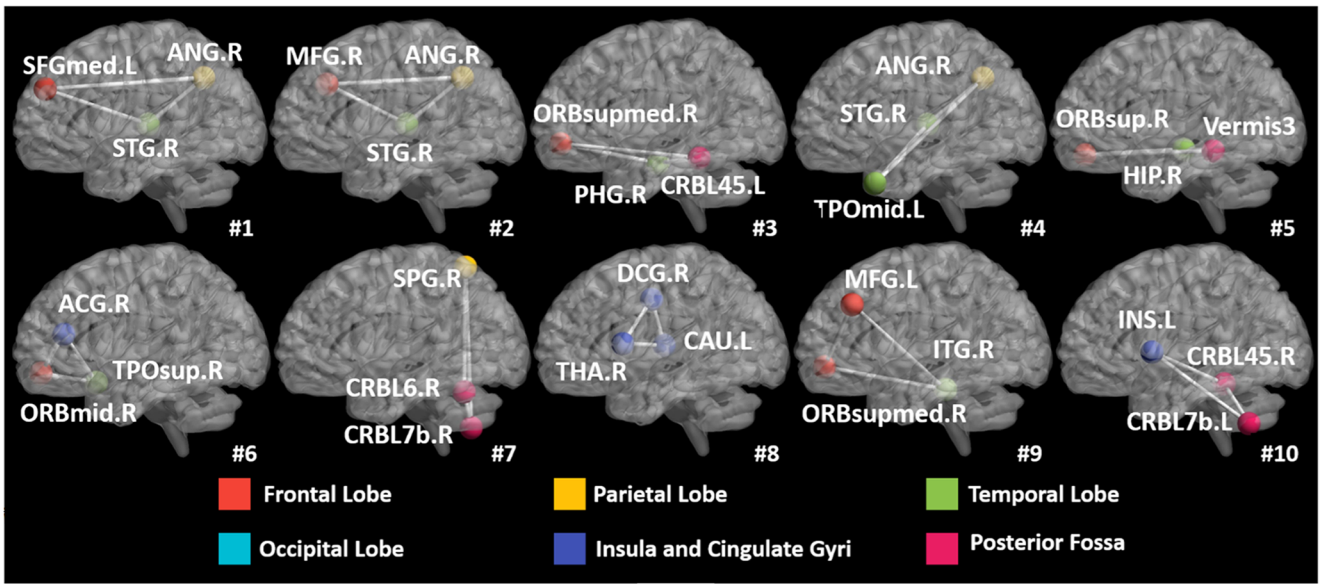


Fig. 6 The top 10 selected subnetworks (white triangle cliques), where the functional connectivity flow running inside has significant difference between ADHD and NC cohorts

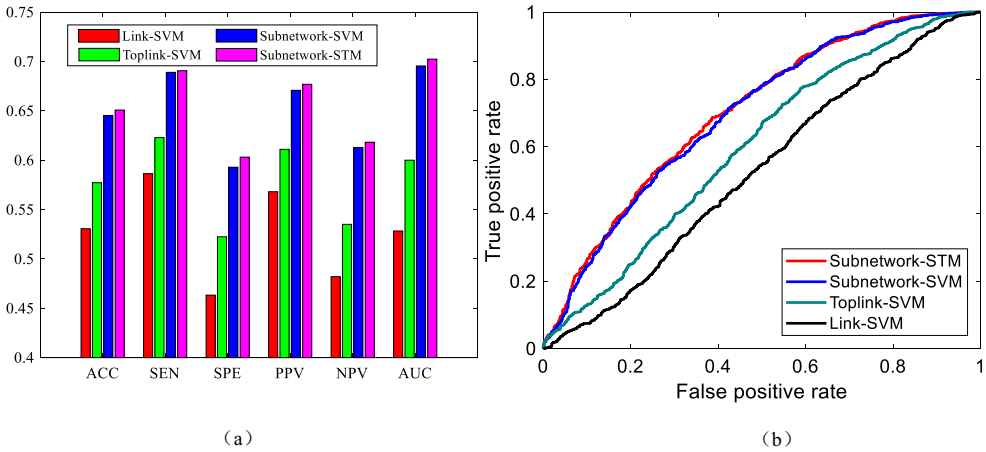


Fig. 7 Classification performances of four different classification methods on ADHD

measures such as sensitivity and specificity. In order to consider the false positives found in hypergraph learning, false positive rate (FPR) is listed in this experiment. The false positive rate obtained by the Hypergraph method is FPR = 1−SPE = 0.1787.

Evaluation on generality To verify the generality of learned subnetworks, we also directly apply the subnetworks learned on ASD dataset of NYU to the classification of 44 ASD and 53 TC subjects from the UM (University of Michigan) site in ABIDE database. The accuracies obtained by Link-SVM and Toplink-SVM are 0.6086 and 0.6253, respectively, which are comparable to those in (Nielsen et al. 2013). Our *Subnetwork-SVM* and *Subnetwork-STM* can improve the accuracies up to 0.6668 and 0.6968, respectively. Again, the classification methods using the features extracted from the learned top subnetworks achieve much higher classification accuracies than the counterpart *Link-SVM* and *Toplink-SVM* methods.

Discovering subnetwork biomarkers for attention deficit hyperactivity disorder (ADHD)

Here, we deploy our proposed method to identify subnetwork biomarkers of attention deficit hyperactivity disorder (ADHD). As we know, ADHD is one of the most common diseases in school-aged children. The data used here are from the New York University Child Study Center (NYU), as a part of the eight databases of ADHD-200 Global Competition. The ADHD dataset of NYU consists of 216 subjects, including 118 ADHD with the ages ranging from 7.24 to 17.61 (mean age 11.26) and 98 healthy controls with the ages ranging from 7.17 to 17.96 (mean age 12.17). For processing, the resting state fMRI data are first performed with a series of preprocessing steps, including removal of a central spike caused by MR signal offset, slice timing, realign, image registration, normalization and spatial smoothing (Matthews and Jezzard 2004). AAL template is also applied to measure the functional connectivity among brain regions. Note that the parameter setting for the ADHD experiments is the same as for the ASD experiments.

Similar to the experiments on ASD data, we also present the top 10 subnetwork biomarkers learned by our proposed method in ADHD in Fig. 6. The current models of ADHD emphasize the dysfunction of frontal, parietal, and cerebellar brain regions in ADHD, as well as disrupted connectivity between these regions in Fig. 6, we can see that these regions predominate in the subnetworks that differentiate children with ADHD from typically developing children. These results demonstrate that our novel method identifies subnetworks that may be meaningful biomarkers for separating children with ADHD from typically developing controls.

To validate the selected subnetwork biomarkers, we also treat the subnetworks as features and then use SVM and STM classifiers for classification. Figure 7 presents the classification performances of different methods on ADHD data. Table 4 gives the details of the classification results. As we can see from Table 4, the classifiers (SVM and STM) using

Table 4 Classification performance on ADHD dataset. The best results are denoted in bold

Method	ACC	SEN	SPE	PPV	NPV	AUC
Link-SVM	0.5306 ± 0.1066	0.5864 ± 0.0016	0.4633 ± 0.0013	0.5681 ± 0.0010	0.4820 ± 0.0013	0.5283 ± 0.1206
Toplink-SVM	0.5772 ± 0.1024	0.6229 ± 0.0015	0.5224 ± 0.0014	0.6110 ± 0.0009	0.5350 ± 0.0013	0.6000 ± 0.1166
Subnetwork-SVM	0.6453 ± 0.0937	0.6890 ± 0.0014	0.5929 ± 0.0013	0.6708 ± 0.0009	0.6129 ± 0.0012	0.6955 ± 0.1043
Subnetwork-STM	**0.6508 ± 0.0967**	**0.6907 ± 0.0014**	**0.6031 ± 0.0014**	**0.6769 ± 0.0009**	**0.6182 ± 0.0012**	**0.7024 ± 0.1044**

Table 5 Classification results of Hypergraph learning on ADHD dataset

	ACC	SEN	SPE	PPV	NPV	AUC
Hypergraph	0.5629 ± 0.1066	0.6110 ± 0.0015	0.5051 ± 0.0014	0.5978 ± 0.0010	0.5189 ± 0.0013	0.5794 ± 0.1191

the features selected by our proposed method achieve better accuracies than those of other traditional methods. More specifically, the best accuracy is 0.6508 obtained by subnetwork-STM, while the accuracy of the conventional method using all the features is just 0.5306, much lower than those of our methods. It is worth nothing that the classification results on NYU dataset are the worst among ADHD-200 datasets at this point, and the best accuracy of 0.6508 achieved by our method is better than the average best imaging-based diagnostic performance of 0.6154 achieved in the ADHD-200 global competition.

Table 5 presents the classification performance of hypergraph learning method. Similar to the results given by Table 3, the Hypergraph method achieves higher classification accuracy than Link-SVM, but lowers than Toplink-SVM. The last column of Table 5 also reports the false positive rate achieved by Hypergraph method FPR =1−SPE = 0.4949.

Conclusion

In this paper, we propose a novel learning method to discover complex connectivity biomarkers that are beyond the widely-used region-to-region connections in the conventional brain network analysis. Specifically, a hypergraph learning technique is introduced to encode the complex subject-wise relationships in terms of various subnetworks, and then quantify the significance of each subnetwork based on the discrimination power across clinical groups as well as consistency within each group. We apply our learning-based method to finding the subnetwork biomarkers for ASD and ADHD, respectively. The learned top subnetworks are *not only* in consensus with the recent clinical findings, *but also* able to significantly improve accuracy in identifying ASD (or ADHD) subjects from normal controls, strongly supporting their potential use and impact in neuroscience study and clinic practice.

One limitation of our proposed method is the increasing computational capacity for subnetworks with more than 3 nodes. Considering the 3-node subnetworks in brain with 116 regions-of-interest, the total number of possible subnetworks is $\binom{116}{3} = 253{,}460$. If increasing 3-node to 4-node subnetworks, the number of the whole subnetworks is $\binom{116}{4} = 7{,}160{,}245$. For the huge 4-node subnetworks, it is impossible to conduct the experiment with the limited memory and CPU resources. In our future work, pre-selection of 116 ROIs will be performed, which will remove some brain regions unrelated to the brain diseases. We hope it can release the demand on computer memory and raise the efficiency. Another limitation is the interpretation of the findings of our method. Most of the brain regions involved in the selected subnetworks locate at the key areas, which are in consensus with some discovery in neuroscience, however the functions of these critical subnetworks still need to be studied. In our following work, we expect to collaborate with neuroscientists to investigate the foundations of these subnetworks in neuroscience for better understanding the brain diseases.

Compliance with ethical standards

Conflict of interest All authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

Appendix

Support Vector Machine:

Given a data set $D = \{\boldsymbol{x}_i, y_i\}_{i=1}^n$ of labeled samples, where $y_i \in \{-1, +1\}$. SVM wants to find the optimal hyperplane which can separate the data and minimize the generalization error at the same time. The optimization problem of SVM can be defined as follows:

$$\begin{aligned} &\min_{\boldsymbol{w},b,\boldsymbol{\xi}} \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_i^n \xi_i \\ &\text{s.t.} y_i\left(\boldsymbol{w}^T\boldsymbol{x}_i + b\right) \geq 1-\xi_i \\ &\xi_i \geq 0, i = 1, 2, \ldots, n \end{aligned} \tag{11}$$

where $\boldsymbol{w}$ is a vector orthogonal to the hyperplane. Equation (11) is a constrained optimization problem and can be solved by using quadratic programming tsechniques.

When a new testing data point coming, a label is assigned to the new sample via the following decision function:

$$g(\boldsymbol{x}) = \text{sign}\left(\boldsymbol{w}^T\boldsymbol{x} + b\right) \tag{12}$$

Support Tensor Machine.

Given a set of training samples $\{\boldsymbol{X}_i, y_i\}$, $i = 1, 2, \ldots, n$, where $\boldsymbol{X}_i$ is the data point in order-2 tensor space, $\boldsymbol{X}_i \in \mathbb{R}^{d_1} \otimes$

$\mathbb{R}^{d_2}$ and $y_i \in \{-1, +1\}$ is the label of $\boldsymbol{X}_i$. The goal of STM is to find a tensor classifier $f(\boldsymbol{X}) = \boldsymbol{u}^T \boldsymbol{X} \boldsymbol{p} + b$ such that the two classes can be separated with maximum margin. Thus, the optimization problem of STM is as follows:

$$\begin{aligned} &\min_{u,p,b,\xi} \frac{1}{2} \left\| \boldsymbol{u}\boldsymbol{p}^T \right\|^2 + c \sum_i^n \xi i \\ &\text{s.t.} y_i \left(\boldsymbol{u}^T \boldsymbol{X}_i \boldsymbol{p} + b \right) \geq 1 - \xi i \\ &\xi i \geq 0, i = 1, 2, \ldots, n \end{aligned} \tag{13}$$

As we can see from Eq. (13), STM is a tensor generation of SVM. The algorithm to solve STM is stated below:

1. Initialization: Let $\boldsymbol{u} = (1, \ldots, 1)^T$.
2. Calculating p: Let $\boldsymbol{x}_i = \boldsymbol{X}_i^T \boldsymbol{u}$ and $\beta_1 = \|\boldsymbol{u}\|^2$, p can be computed by solving the following problem:

$$\begin{aligned} &\min_{p,b,\xi} \frac{1}{2} \beta_1 \boldsymbol{p}^T \boldsymbol{p} + \boldsymbol{c} \sum_i^n \xi_i \\ &\text{s.t.} y_i \left(p^T x_i + b \right) \geq 1 - \xi i \\ &\xi i \geq 0, 1 = 1, 2, \ldots, n \end{aligned} \tag{14}$$

It is worth noting that problem (14) is the same as objective function (11) of SVM. Thus, the standard optimization approach for SVM can be adopted for Eq. (14).

3. Calculating u:When $\boldsymbol{p}$ is solved, let $\widetilde{\boldsymbol{x}_i} = \boldsymbol{X}_i \boldsymbol{p}$ and $\beta_2 = \|\boldsymbol{p}\|^2$. u can be computed by solving the following problem:

$$\begin{aligned} &\min_{u,b,\xi} \frac{1}{2} \beta_2 \boldsymbol{u}^T u + c \sum_i^n \xi_i \\ &s.t. y_i \left(\boldsymbol{u}^T x_i + b \right) \geq 1 - \xi_i \\ &\xi_i \geq 0, 1 = 1, 2, \ldots, n \end{aligned} \tag{15}$$

As above, the standard optimization method for SVM can also be used to solve Eq. (15).

4. Iteratively computing $\boldsymbol{u}$ and $\boldsymbol{p}$: $\boldsymbol{u}$ and $\boldsymbol{p}$ can be iteratively calculated by step 2 and 3 until a convergence attained.

References

Agarwal, S., Lim, J., Zelnik-Manor, L., Perona, P., Kriegman, D., & Belongie, S. (2005). Beyond pairwise clustering. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (vol. 2, pp. 838–845). IEEE.

Argyriou, A., Evgeniou, T., & Pontil, M. (2008). Convex multi-task feature learning. *Machine Learning, 73*(3), 243–272.

Bu, J. et al. (2010) Music recommendation by unified hypergraph: combining social media information and music content. In *Proceedings of the 18th ACM international conference on Multimedia* (pp. 391–400). ACM.

Chapelle, O., Scholkopf, B., & Zien, A. (2009). Semi-supervised learning (Chapelle, O. et al., Eds.; 2006)[Book reviews]. *IEEE Transactions on Neural Networks, 20*(3), 542–542.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.

Davison, E. N., Schlesinger, K. J., Bassett, D. S., Lynall, M. E., Miller, M. B., Grafton, S. T., & Carlson, J. M. (2015). Brain network adaptability across task states. *PLoS Computational Biology, 11*(1), e1004029.

Di Martino, A., et al. (2014). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular Psychiatry, 19*(6), 659–667.

Gao, Y., Wang, M., Tao, D., Ji, R., & Dai, Q. (2012). 3-D object retrieval and recognition with hypergraph analysis. *IEEE Transactions on Image Processing, 21*(9), 4290–4303.

Gao, Y., Adeli-M, E., Kim, M., Giannakopoulos, P., Haller, S., and Shen, D. (2015a) Medical image retrieval using multi-graph learning for MCI diagnostic assistance. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 86–93). Springer.

Gao, Y. et al. (2015b) MCI identification by joint learning on multiple MRI data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 78–85) Springer.

Huang, Y., Liu, Q., Zhang, S., and Metaxas, D. N. (2010) Image retrieval via probabilistic hypergraph ranking. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pp. 3376–3383: IEEE.

Huang, Y., Liu, Q., Lv, F., Gong, Y., & Metaxas, D. N. (2011). Unsupervised image categorization by hypergraph partition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 33*(6), 1266–1273.

Jie, B., Wee, C.-Y., Shen, D., & Zhang, D. (2016). Hyper-connectivity of functional networks for brain disease diagnosis. *Medical Image Analysis, 32*, 84–100.

Matthews, P., & Jezzard, P. (2004). Functional magnetic resonance imaging. *Journal of Neurology, Neurosurgery & Psychiatry, 75*(1), 6–12.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science, 298*(5594), 824–827.

Minshew, N. J., & Williams, D. L. (2007). The new neurobiology of autism: cortex, connectivity, and neuronal organization. *Archives of Neurology, 64*(7), 945–950.

Nielsen, J. A. et al. (2013) Multisite functional connectivity MRI classification of autism: ABIDE results.

Sporns, O., & Kötter, R. (2004). Motifs in brain networks. *PLoS Biology, 2*(11), e369.

Sun, L., Ji, S., & Ye, J. (2008). Hypergraph spectral learning for multi-label classification. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 668–676). ACM.

Tao, D., Li, X., Hu, W., Maybank, S., and Wu, X. (2005) Supervised tensor learning. In *Fifth IEEE International Conference on Data Mining (ICDM'05)* (pp. 8). IEEE.

Tian, Z., Hwang, T., & Kuang, R. (2009). A hypergraph-based learning algorithm for classifying gene expression and arrayCGH data with prior knowledge. *Bioinformatics, 25*(21), 2831–2838.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage, 15*(1), 273–289.

Van Den Heuvel, M. P., & Pol, H. E. H. (2010). Exploring the brain network: a review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology, 20*(8), 519–534.

Yu, J., Tao, D., & Wang, M. (2012). Adaptive hypergraph learning and its application in image classification. *IEEE Transactions on Image Processing, 21*(7), 3262–3272.

Yu, J., Rui, Y., Tang, Y. Y., & Tao, D. (2014). High-order distance-based multiview stochastic learning in image classification. *IEEE Transactions on Cybernetics, 44*(12), 2431–2442.

Zass, R. and Shashua, A. (2008) Probabilistic graph and hypergraph matching. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (pp. 1–8) IEEE.

Zeng, L.-L., Shen, H., Liu, L., Wang, L., Li, B., Fang, P., Zhou, Z., Li, Y., & Hu, D. (2012). Identifying major depression using whole-brain functional connectivity: a multivariate pattern analysis. *Brain, 135*(5), 1498–1507.

Zhang, Z., Wang, J., & Zha, H. (2012). Adaptive manifold learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 34*(2), 253–265.

Zhang, L., Gao, Y., Hong, C., Feng, Y., Zhu, J., & Cai, D. (2014). Feature correlation hypergraph: exploiting high-order potentials for multi-modal recognition. *IEEE Transactions on Cybernetics, 44*(8), 1408–1419.

Zhou, D., Huang, J., and Schölkopf, B. (2006). Learning with hypergraphs: Clustering, classification, and embedding. In *Advances in neural information processing systems* (pp. 1601–1608).