Joint Estimation and Regularized Aggregation in Brain Network Analysis

by

Jongik Chung

(Under the Direction of Cheolwoo Park)

Abstract

In the Gaussian graphical model framework, precision matrices reveal conditional dependence structure among random variables. In functional magnetic resonance imaging (fMRI) data, estimating such precision matrices of multi-subjects and aggregating them to a group-level is an essential step for constructing a group brain network. In this dissertation, I consider joint estimation of multiple precision matrices with regularized aggregation under both Gaussian and Non-Gaussian assumptions. In the estimation of individual precision matrices, I take a regularization approach to induce sparsity, which makes brain network estimation more realistic. Also, in the construction of a group precision matrix, the simple average of individual precision matrices may be affected by outliers and provide inconsistent outcomes between subject-level and group-level networks. In contrast, the proposed methods yield a robust group graph which can identify and ease the effect of outliers. I demonstrate the effectiveness of the proposed methods through simulated examples and analyses on saccade tasks fMRI data.

Index words: Aggregation, FMRI data, Graphical models, Precision matrix, Regularization

Joint Estimation and Regularized Aggregation in Brain Network Analysis

by

Jongik Chung

B.B.A., Korea University, Republic of Korea, 2011

M.B.A., Korea University, Republic of Korea, 2014

M.S., University of Georgia, 2016

A Dissertation Submitted to the Graduate Faculty of the

University of Georgia in Partial Fulfillment of the Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2020

Joint Estimation and Regularized Aggregation in Brain Network Analysis

by

Jongik Chung

Major Professor: Cheolwoo Park

Committee: Nicole Lazar
Jennifer E. McDowell
Lynne Seymour
Qian Xiao

Electronic Version Approved:

Ron Walcott
Interim Dean of the Graduate School
The University of Georgia
August 2020

Acknowledgments

First, I would like to express the deepest appreciation to my advisor professor Dr. Cheolwoo Park for the continuous support of my Ph.D study and related research, for his patience, motivation, and immense knowledge. He continually and convincingly conveyed a spirit of adventure in regard to research and scholarship. This dissertation would not have been possible without his guidance and persistent help.

My appreciation also extends to my committee professors, Dr. Nicole Lazar, Dr. Jennifer McDowell, Dr. Lynne Seymour, and Dr. Qian Xiao, for their insightful suggestions and encouragements, but also for the challenging issues which inspired me to widen my research from various perspectives. In addition, I would like to give special thanks to the members of the Neuroimaging Data Analysis Group at the University of Georgia, for their practical suggestions and helpful advices. I thank to all of my colleagues and friends for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last few years.

Last but not least, I would like to thank my wife Wooree for her personal support and great patience at all times. My parents, my wife's family, and my brother's family have given me their unequivocal support throughout, as always, for which my mere expression of thanks likewise does not suffice.

Financial support was partly provided by National Science Foundation (NSF IIS-1607919).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In graph theory, a graphical model illustrates probability distributions using graphical description. It aims to represent the joint distribution and graphically reveal the associations between random variables based on their conditional dependence structure. A graph comprises nodes (vertices) connected by edges (links). A node represents a random variable or a group of random variables, and an edge describes a probabilistic relationship between nodes. Two types of graphs are used to model the relationships: directed graphs, also called Bayesian networks, and undirected graphs, which are known as Markov networks. While directed graphs use arrows to show directionality between associations and are useful to explain causal relationships, undirected graphs demonstrate connections by simple lines and do not require any specific paths from one to other variables (Bishop, 2006; Koller et al., 2009). Undirected graphical models are known to be appropriate for spatial or relational data (Jordan et al., 2004; Wainwright et al., 2008; Murphy, 2012) because conditional independence can be determined by simple graph separation. Hence, we focus on undirected graphical models in this dissertation to describe brain networks for functional magnetic resonance imaging (fMRI) data, which is our main application.

The basic measurement in fMRI data is blood oxygen level dependent (BOLD) signal, which measures inhomogeneities in the magnetic field through changes in the level of oxygen

in the blood (Ogawa et al., 1990). Standard fMRI data scanned on a single subject typically contain 4,096 voxels (64×64 grid, $3 \times 3 \times 3$ mm^3) in a single slice of the brain (Huettel et al., 2004; Lazar, 2008). Generally, multiple slices are measured to illustrate 3-dimensional brain volume. At each voxel of the brain, BOLD signals are measured along the time of the experiment, for example, collected every 2 second over 5 - 7 minute protocol. Hence, the BOLD signals are inherently correlated in both space and time. Not only do we consider a single subject, but commonly we also analyze groups of subjects in order to draw a general conclusion about differences between groups. Our main question is how to estimate relationships between brain regions, and simultaneously produce subject-wise individual graphs and a representative group graph that can show common characteristics that are shared by the subjects in the group.

The application of graphical models to fMRI data is an active field of research. Cribben et al. (2012) applied Dynamic Connectivity Regression to identify brain connectivities in anxiety-inducing experimental fMRI data. They constructed dynamic undirected graphs based on functional connectivity changes along with experimental conditions. Ng et al. (2013) proposed Sparse Group Gaussian Graphical Model for joint estimation of intra-subject and group-level functional connectivities, and they showed its high sensitivity in activation detection for fMRI data. Li and Solea (2018) suggested Functional Additive Precision Operator estimator without the Gaussian assumption. Their estimator showed superior performance when a nonlinear dependence exists in the data, and it inferred that the brain networks of healthy children and children with Attention Deficit Hyperactivity Disorder (ADHD) were different. Another study of a graphical model analyzing fMRI data was introduced by Zhu and Li (2018). The authors argued that their tensor-based method could estimate multiple graphs from matrix-valued data and showed its statistical properties based on resting state fMRI data with a number of subjects with Autism Spectrum Disorder and ADHD.

In the framework of Gaussian graphical models, the most essential step is estimation of a precision matrix, which is an inverse covariance matrix. Because the precision matrix from the Gaussian distribution represents conditional dependence structure, one can build an undirected graph from the estimated precision matrix. However, high-dimensional data structures that we frequently observe these days, such as fMRI data, make the estimation more challenging. In the early models, many researchers took a penalized maximum likelihood approach to construct a sparse precision matrix (Meinshausen et al., 2006; Yuan and Lin, 2007; Banerjee et al., 2008; Dahl et al., 2008; Friedman et al., 2008; Fan et al., 2009). For example, Friedman et al. (2008) suggested the estimation of a sparse undirected graph through the l_1 penalty (Lasso).

More recently, several other techniques have been utilized to solve the estimation problem. In order to reflect no relationship between some nodes or to deal with high dimensional data structures, many approaches take a penalized maximum likelihood approach to construct a sparse precision matrix. They include likelihood approaches using a group penalty (Friedman et al., 2010a; Ng et al., 2013; Danaher et al., 2014), a dynamic model with changing points (Cribben et al., 2012), and joint estimation of multiple precision matrices (Guo et al., 2011; Danaher et al., 2014; Lee and Liu, 2015; Cai et al., 2016; Zhu and Li, 2018). A state-of-the-art design is the latter approach, which jointly estimates multiple precision matrices at the same time based on a constrained optimization procedure. Since our main interest is group fMRI data analysis, it is natural to consider the joint estimation of group and subject-wise precision matrices. Cai et al. (2011) proposed a method utilizing a constrained l_1 minimization problem and producing a sparse precision matrix without the Gaussian distribution assumption. Lee and Liu (2015) established their method as an advancement of Cai et al.'s method. With a similar sparse constrained optimization, Lee and Liu (2015) offered joint estimation of multiple precision matrices by decomposing those matrices into common and unique structures. By estimating group and subject-wise precision matrices

simultaneously, one can obtain more interpretable information than what can be obtained from individual modeling. Danaher et al. (2014) also suggested the joint estimation for inverse covariance using graphical lasso that requires the Gaussian distribution assumption. They employed the generalized fused lasso or the group lasso penalty in their method, which is based on maximizing a penalized log-likelihood. Our work is in line with Lee and Liu (2015) and Danaher et al. (2014), and it further employs an optimal aggregation of subject-wise precision matrices in order to produce a robust and representative group brain network.

A common issue in group fMRI data analysis is aggregation of subject-wise results. After estimating a precision matrix from each subject, one needs to determine how to combine those individual precision matrices into a group. Aggregation is an important step in group analysis since it can lead scientists to generalized results about the whole group and provides more interpretable outcomes. Aggregation methods in neuroimaging data can be categorized into two approaches, which are combining hypothesis tests and combining estimates of treatment effects (Lazar et al., 2002). The latter approach directly relates to the problem in which we are interested, and random effects models are frequently used in this approach (Council et al., 1992). However, random effects models assume the Gaussian distribution and are known to yield conservative results that can lead to loss of information in outcomes (Lazar, 2008; Wang et al., 2019). Also, outliers can have a big impact on aggregation results (McNamee and Lazar, 2004), so it is desirable to consider a robust approach that is not affected by normality or homogeneity assumptions. Wang et al. (2019) proposed a regularized aggregation approach that reduces the influence of outliers in constructing group statistical parametric maps.

In this dissertation, we develop a graphical model for group fMRI data that incorporates joint estimation of multiple precision matrices, with or without the Gaussian assumption, and regularized aggregation simultaneously. Since the group precision matrix estimation can be affected by individual precision matrices with different patterns (i.e., outliers), the integration of robust aggregation to the estimation can provide more accurate and robust results than

the simple average as in Lee and Liu (2015) and Danaher et al. (2014). We achieve this through a regularization technique to assign weights to individual precision matrices, which can provide the degree of outlyingness of each subject. In other words, subjects that share a similar covariance structure have similar high weights, whereas subjects that have a different covariance structure from the majority have similar but lower weights. One of the approaches that we suggest requires the Gaussian distribution assumption, while the other proposed approach does not assume a distribution, either in constructing individual precision matrices or in aggregating them. We compare the performance between the Non-Gaussian-based and the Gaussian-based estimation. In order to evaluate our methods, we conduct both simulation and real data analysis and show that our methods can detect outliers and provide robust group graphs. In the real data analysis, we consider a wavelet transform as a pre-whitening procedure to mitigate temporal correlation in the fMRI data.

The rest of the dissertation is organized as follows. In Chapter 2, we review the properties of fMRI data analysis, graphical models, and the methods proposed by Lee and Liu (2015), Danaher et al. (2014), and Wang et al. (2019). Then we describe the proposed joint estimation of multiple precision matrices with regularized aggregation and the joint graphical lasso with regularized aggregation in Chapter 3. In Chapter 4, we report the behavior of the proposed methods in various simulation settings, which resemble real fMRI data. Then, we describe the probability saccade tasks fMRI data collected by Pierce and McDowell (2017), apply the two proposed methods to the fMRI data, and evaluate the performances in Chapter 5. We also compare the performance of the proposed methods with that of Lee and Liu's method and Danaher et al.'s method in Chapters 4 and 5. Finally, we conclude with discussion and future research directions in Chapter 6.

Chapter 2

Literature Review

In this chapter, we review several aspects of fMRI data analysis in Section 2.1. We also review general properties of graphical models introduced by Bishop (2006) and Koller et al. (2009) in Section 2.2. Then the methods which can jointly estimate multiple precision matrices based on the Non-Gaussian and Gaussian assumption, suggested by Lee and Liu (2015) and Danaher et al. (2014), and the regularized aggregation technique by Wang et al. (2019) are respectively reviewed in Sections 2.3, 2.4 and 2.5.

2.1 FMRI Data Analysis

2.1.1 Properties

The basic measurement in fMRI is Blood Oxygen Level Dependent (BOLD) signal. BOLD signal measures inhomogeneities in the magnetic field through changes in the level of oxygen in the blood. It originates from the difference in the magnetic properties of de-oxygenated and oxygenated blood. De-oxygenated hemoglobin is more paramagnetic than oxygenated hemoglobin, which is diamagnetic (Pauling and Coryell, 1936). This difference can be captured by magnetic resonance (MR) signal, indeed, BOLD effect. The balance between oxygen

consumption and oxygen supply effects on the total amount of de-oxygenated hemoglobin in a brain region, and it reveals as BOLD contrast (Ogawa et al., 1990). In fact, when neuronal activity increases, the vascular system supplies more oxygenated hemoglobin than necessary, and it leads to over-compensatory increase of regional cerebral blood flow. This decreases the relative amount of de-oxygenated homoglobin, which results in a higher T2* signal strength and a brighter MR image (Huettel et al., 2004).

The changes in the ratio of oxygenated to de-oxygenated hemoglobin in the MR signal from neuronal activity reflects the hemodynamic response. Before any change in the MR signal, there exists a small decrease which is called the initial dip. This can be explained by the time demands for the vascular system to respond to the brain's need. Based on the stimuli, response increases and peaks after 4 to 6 seconds, and returns to baseline after approximately 20 seconds. The decrease in BOLD signal goes below the baseline, which is called the undershoot, is due to the combination of reduced blood flow and increased blood volume (Huettel et al., 2004). This regional BOLD response to a brief stimulus is known as the hemodynamic response function (HRF), and it can vary among regions and subjects.

In order to evaluate BOLD signal explicitly, we need to determine spatial and temporal resolutions. Spatial resolution indicates the ability to discriminate differences between nearby spatial locations, which can be measured by the size of voxels. Voxels are three-dimensional rectangular prisms, consist of field of view, matrix size, and slice thickness. The field of view refers to the area of a slice, and the matrix size represents the grid imposed on the slice. Naturally, larger voxels contain more neurons on average and include more blood flow, so have more signal than smaller voxels. In addition, acquisition time increases as the size of voxels smaller. Therefore, it is not always good to use the smallest possible voxel size in experiments, though increased spatial resolution improves the ability to distinguish between neighboring areas. The large-vessel effect is caused by the drainage system in large veins. It should be minimized for good spatial resolution, since the signal from large vessels are not related with

the neuronal activity. We can manage this by using spin-echo or diffusion-weighted sequences (Huettel et al., 2004). In short, the appropriate spatial resolution depends on the objective of the research.

Temporal resolution is the smallest time period which can accurately separate the timing of brain activity based on the BOLD signal. It can be measured by TR, which is the basic time resolution parameter. We can have many samples by reducing TR, which results in improved hemodynamic response estimates. However, using very short TRs introduces less measurement of MR signal and reduced spatial coverage with a given pulse sequence. Therefore, similar with spatial resolution, preferred temporal resolution depends on a given experimental condition and research objective.

Generally, an increase in the positive BOLD signal can be interpreted as a net increase in brain activation, since it relies on the level of de-oxygenated to oxygenated hemoglobin in the blood. BOLD signals also reflect the sum of the effects of oxygen consumption and blood flow increase (Hall et al., 2016). However, we note that the amplitude of the BOLD signal does not firmly reflect behavioral performance. As the subjects get used to performing the task, the amplitude may decreases while the efficiency remains the same (Huettel et al., 2004).

The nature of the BOLD signal requires specific types of paradigm designs. Typically, two designs are possible; block and event-related designs. In block designs, periods of task alternate with periods of rest or different task (Lazar, 2008) with periods lasting on the order of 20 - 40 seconds. They are powerful to detect which voxels are activated. In the task blocks, increase coincident with constant stimulation of task-related neurons. Recovering to the baseline only happens during the control blocks. Hence, the activated voxels can be determined as the ones display increased levels of signal during the task blocks, since the deactivated voxels do not show response in both task and control blocks. Moreover, block designs are relatively easy to implement and straightforward by presenting one stimulus type

in each block. On the other hand, block designs also have some limitations. First, if we consider two or more stimuli, not all tasks can be blocked. Specifically, first task may affect second task if we compose two task blocks consecutively. Second, subjects may anticipate the order and duration of conditions so that it can affect the result of brain activities. Last but not least, block designs do not allow separation of BOLD response to individual trials (for instance correct versus error trials), since it averages over hemodynamic responses.

In contrast, event-related designs do not block any experimental conditions. Instead, stimuli occur individually, separated by an inter-stimulus interval (ISI). There are fewer restrictions on types of stimuli, so different type of stimuli can be shown irregularly at different timings. This allows more flexible experiment rather than block designs and removes anticipation from subjects. Furthermore, event-related designs provide good estimate of the hemodynamic response function, since the BOLD response following a stimulus will return to baseline if the ISI is long enough. Event-related designs also can separate response with respect to the tasks, for example, trials in which the subject remembered versus trials in which the subject forgotten stimuli. Even though event-related designs introduce the greater flexibility than block designs, they also have some disadvantages. Since they can handle any types of stimuli, their statistical analysis is generally more complicated than that of block designs. Also they have lower efficiency than block designs, take more time for the same power to detect effect because of a lower signal-to-noise (SNR) ratio. Moreover, functional contrast from event-related designs tends to decrease rather than block designs, with shortened constant ISI (Bandettini and Cox, 2000). In general, the major disadvantage of event-related designs is that they have less detection power than block designs.

2.1.2 Pros and Cons of fMRI

MRI uses strong magnetic fields to create brain images. It provides high spatial resolution images so that it is able to detect and represent different spatial locations. FMRI produces

improved images which represent increased oxygen flow in response to energy needs of certain brain regions based on BOLD signals. There are some advantages and disadvantages of MRI as a brain imaging technology, compared with other brain imaging techniques such as Positron Emission Tomography (PET), Electroencephalogram (EEG), and Magnetoencephalogram (MEG).

FMRI has several advantages over PET, EEG, and MEG. FMRI does not involve radioactive substances as does PET, so it is possible to noninvasively set up extended or detailed longitudinal studies (Munoz-Cespedes et al., 2005). Moreover, it has higher spatial resolution than PET, EEG, and MEG which enables to accurately depicts structural data. The PET can measure the blood flow changes in approximately 5 - 10 cubic millimeters, while fMRI can measure in approximately 3 cubic millimeters (Afrasiabi and Noroozian, 2015). FMRI can record signal from all regions of the brain, unlike EEG or MEG which does not accurately measure neural activity that occurs below the surface of the cortex. In addition, fMRI has a reasonable temporal resolution which is higher than that of PET, which allows to expand the type of experimental designs for detecting activation of brain regions.

Although it has many benefits than other brain imaging techniques, fMRI also has some limitations. First, BOLD signal, which is the measurement of fMRI, is only an indirect measure of neural activity. Hence it is susceptible to non-neural changes in the body (Turner, 2016). Although PET measures the flow of blood in the brain which based on a similar technique with fMRI, EEG and MEG directly measure brain activation by electric signals of neurons. In this point of view, EEG or MEG may more accurately detect brain activations than fMRI and PET. Second, there is a possibility that the BOLD signal is captured in a non-activity area, since BOLD signals are strongly related with the input to a given area (Turner, 2016). Third, fMRI has lower temporal resolution than EEG or MEG. Since the BOLD signal peaks approximately 5 seconds after neuronal activity, it may difficult to distinguish BOLD response with other factors within a short amount of time. There also exist some

minor disadvantages. Unlike PET which is not motion sensitive, fMRI can be obscured by any physical movement (Afrasiabi and Noroozian, 2015). fMRI also does not have portable machines, whereas EEG has portable devices.

2.1.3 Applications and Statistical Issues

As described in Section 2.1.2, fMRI has several advantages as a brain imaging technique. FMRI has an acceptable temporal resolution with high spatial resolution, and it allows multiple measurements without any risk of radioactive materials. We can use brain imaging techniques such as fMRI to diagnose and treat mental illnesses, since it has been known that there are relationships between brain functions and psychiatric disorders (Biological Sciences Curriculum Study, 2007). For example, investigating the disorders such as depression, Attention Deficit Hyperactivity Disorder (ADHD), Schizophrenia and so on by using fMRI is an active field of study. The benefits of fMRI can explain the reason why it is often used to study psychiatric populations.

For the sake of enabling the longitudinal study, fMRI is able to indirectly detect brain activation for the assessment of psychiatric disease treatment effects. It can identify changes in brain functions according to the specific treatment, so researchers can verify such treatment is whether effective or not. For example, Mitterschiffthaler et al. (2006) introduced three fMRI studies which found normalization of attenuated amygdala activation in response to negative facial expressions prior to treatment after antidepressant treatment. The authors also provided two more fMRI studies of effective antidepressant treatment on patients who displayed greater relative anterior cingulate activation and subgenual anterior cingulate. Correspondingly, we can compare BOLD signals from baseline and treatment to identify mediums of treatment outcome. Recent studies show that fMRI can be a reliable tool for predicting treatment response by employing longitudinal approaches.

We can also compare healthy subjects with those who have mental illnesses through fMRI. Specifically, different neural responses from two groups based on the same emotional or cognitive task indicate distinguishable localized brain activation, which implies different information processing in two groups' brain regions. By studying differences in neural response between healthy subjects and psychiatric patients with respect to activated brain regions, researchers have the opportunity to evaluate detailed dissimilarities. We can also find that how different the brains of patients and normal subjects are, in terms of receiving information and performing tasks (Filippi, 2016).

Moreover, some studies have shown that the fMRI signal can be regarded as an intermediate phenotype (observable characteristics) in genetics studies of neuropsychiatric disorders (Mitterschiffthaler et al., 2006). In this case, BOLD response is represented as the phenotype under investigation. The authors in the studies argue that they can identify neural correlates of a behavior of interest by experimentally isolating task and control in block designs. If there exists behavioral deviation between healthy controls and psychiatric patients through a certain behavioral or cognitive task, we may claim important differences at a neurobiological level.

In a statistical point of view, there are several issues regarding fMRI data analysis. The first problem is complexity and massive amounts of the data themselves (Lindquist et al., 2008). The data consist of a sequence of MRIs along time, each has a large number of voxels, partitioning whole brain regions. In reaction to neuronal activity, changes in BOLD signal for each voxel are recorded, and these values across time can be used to infer the timing and location of brain activation. Further, the experiment may be repeated several times for the same subject, in a group level. In this data environment, there exist potentially high correlations between observations (Lazar et al., 2001). Since voxels in nearby regions can be expected to have similar characteristics, spatial correlations can be presented. Also, repeated measurements over time introduce temporal correlations. It is not easy to deal with this type

of data in a traditional statistical methods, since the assumptions in those methods, such as independently and identically distributed observations, may not be satisfied.

Second, the low signal to noise ratio in fMRI experiments is one of the common problems. The noise induced by machine, such as unstable magnetic field, unusual vibration, and temperature fluctuation should be also handled before analyzing. Another source of noise is the movements from subjects, especially head movements. This type of noise can be dealt with the software in preprocessing steps, but it is possible to remain and be present in the data. The noise may blur detecting brain activation regions, so it is important to control it effectively.

The third problem commonly induced in fMRI data analysis is subject-wise variation. In order to focus on group data to perform population inference, a normalization procedure is required since brain size and shape of each subject are different. In this normalization, varying brain features are standardized and warped onto a typical brain template. However, this procedure introduce spatial imprecision, which includes incorrect localization or even absence of small activation regions (Lindquist et al., 2008). Not only brain images, but also BOLD signals may have subject-wise variation problem. Even though sample mean is a powerful statistic which can represent a group, there may exist subjects which can be considered as outliers. Such outliers have crucial effect on simple average and obscure researchers to get accurate population inference. Therefore, we need to deal with outliers in our modelling procedure.

2.2 Graphical Models

A graphical model is a probabilistic model for which a graph expresses the conditional dependence structure between random variables, using diagrammatic representations of probability distributions (Bishop, 2006). We can simply observe a structure and properties of a probabilistic model when we investigate the graph. Such graphical manipulations can also

express complex computations in advanced models. In graph theory, there are two different categories in the graph: directed graphical models and undirected graphical models. As their names indicate, directed graphical models use arrows to point in directions, whereas undirected graphical models use simple lines to connect the nodes. We first review conditional independence property in Section 2.2.1.

2.2.1 Conditional Independence

Conditional independence is an essential property of probability distributions in the construction of graphs. Suppose we have three random variables x, y, and z. x and y are conditionally independent given z if

$$p(x, y|z) = p(x|y, z)p(y|z) = p(x|z)p(y|z).$$

Note that if $x \perp\!\!\!\perp y|z$ where the notation "$\perp\!\!\!\perp$" indicates the independence between two variables,

$$\begin{aligned} p(x, y, z) &= p(x, y|z)p(z) = p(x|z)p(y|z)p(z) = p(x, z)p(y, z)/p(z) \\ &= h(x, z)k(x, z), \end{aligned}$$

which means that the joint distribution factorizes into the product of marginal distributions. Figure 2.1(a) describes an example for conditional independence in a directed graph. By applying the chain rule, we can write the joint distribution as:

$$p(x, y, z) = p(x|z)p(y|z)p(z).$$

Conditioning on the variable z, as represented by the shaded node in Figure 2.1(a), the conditional distribution of x and y given z has the form

$$p(x, y|z) = \frac{p(x, y, z)}{p(z)} = \frac{p(x|z)p(y|z)p(z)}{p(z)} = p(x|z)p(y|z),$$

so we obtain the conditional independence property

$$x \perp\!\!\!\perp y|z.$$

Here the node z is known as tail-to-tail with respect to the path from node x to y via z, since the node is connected to the tails of the two arrows. As the conditioned node z blocks the path from x to y, the conditional independence between x and y is induced.

For an undirected graphical model that will be covered in Section 2.2.3, the conditional independence property becomes more simple. Figure 2.1(b) shows the conditional independence between x and y, given z for an undirected graph. The conditional independence of x and y, given z is satisfied in this case since all possible paths that connect node x and y are through the node z. In general, suppose we have three sets of nodes T_1, T_2, and T_3. Then the property must hold when there is no path that connects any node in set T_1 to any node T_2, when all nodes in set T_3 with connected edges are eliminated.

2.2.2 Directed Graphical Models

In directed graphical models, the links of the graphs have a certain directionality, and arrows represent them. We can add directed arrows to the graph from the nodes corresponding to the variables on which the distribution is conditioned. To describe, assume that we have three random variables x, y, and z with an arbitrary joint distribution of $p(x, y, z)$. Note that a graph can represent any type of variables, either discrete or continuous, and we do not need to specify any class of distributions for the graph. The joint distribution can be

decomposed by applying the chain rule:

$$p(x, y, z) = p(x|y, z)p(y, z) = p(x|y, z)p(y|z)p(z). \tag{2.1}$$

Figure 2.2(a) illustrates a simple directed graph which shows the probabilistic relationships in the right-hand side of equation (2.1). Each of the random variable x, y, and z represents each node. Then we add directed arrows to the graph from the nodes paring with the variables on which the distribution is conditioned for each conditional distribution in equation (2.1). For example, $p(x|y, z)$ can be illustrated by two arrows from nodes y and z to node x. If there exists a link from a node x to a node y, node x is called the parent of node y, and node y is called the child of node x. Note that the joint distribution in equation (2.1) can be decomposed in different ordering, and such different ordering results in different graph representation. In general for m variables, the joint distribution can be written by a product of conditional distributions:

$$p(x_1, \cdots, x_m) = p(x_n|x_1, \cdots, x_{m-1}) \cdots p(x_2|x_1)p(x_1). \tag{2.2}$$

Again, the joint distribution in equation (2.2) can be represented as a directed graph having m nodes, with each node having arrows from all conditioned nodes. Since there exists an arrow between every pair of nodes, we say the graph is fully connected.

Another example of a directed graph is provided in Figure 2.2(b). The graph is not fully connected since there are no arrows, for example, from x_1 to x_2, or from x_2 to x_5. Then the graph in Figure 2.2(b) can be written in terms of the product of conditional distributions:

$$p(\boldsymbol{x}) = p(x_1)p(x_2)p(x_3|x_1, x_2)p(x_4|x_1, x_2)p(x_5|x_3, x_4), \tag{2.3}$$

where $\boldsymbol{x} = \{x_1, x_2, \cdots, x_5\}$. Each of the conditional distribution is conditioned only on the parents of the corresponding node. This graph is an example of a directed acyclic graph (DAG), which does not include any directed cycles. There are no closed paths within the graph such that the starting node and the ending node is the same when we move along the direction of the arrows. Generally, the joint distribution for a graph with m nodes can be presented by:

$$p(\boldsymbol{x}) = \prod_{i=1}^{m} p(x_i|par_i), \tag{2.4}$$

where $\boldsymbol{x} = \{x_1, \cdots, x_n\}$ and par_i is the set of parents of x_i. This equation (2.4) explains the factorization properties of the joint distribution for a directed graphical model. Notice that the right-hand side of equation (2.4) is appropriately normalized subject to the normalized individual conditional distributions.

2.2.3 Undirected Graphical Models

An undirected graphical model is another class of models that provides a graph-based representation of conditional independence, by removing the directionality from the links of the graph. In order to introduce a factorization property for undirected graphs, we first need to define a clique. A clique is defined as a subset of the nodes in a graph when all the pairs of nodes in the subset are connected. Additionally, a maximal clique is a clique that cannot be extended by including any other nodes from the graph in the set. Figure 2.3 illustrates two undirected graphs. The graph in Figure 2.3(a) has three cliques given by $\{a, b\}$, $\{b, c, d\}$, and $\{c, e\}$, while the graph in Figure 2.3(b) has two cliques that are $\{a, d\}$ and $\{a, b, c\}$. Note that all of these cliques are also maximal cliques. The set of cliques decides the graph, and therefore we can define the factors to be functions of the variables in the cliques (maximal cliques) in the decomposition of the joint distribution.

Let C be a clique and $\boldsymbol{x}_C$ be the set of variables in C. Then the joint distribution can be decomposed by a product of potential functions $\phi_C(x_C)$ over the maximal cliques of the graph:

$$p(\boldsymbol{x}) = \frac{1}{Z} \prod_C \phi_C(\boldsymbol{x}_C),$$

where Z is a normalizing constant, which is

$$Z = \sum_{\boldsymbol{x}} \prod_C \phi_C(\boldsymbol{x}_C).$$

A choice of potential function can be different based on model configurations.

Another important property of undirected graphical models is a relationship with the Gaussian distribution assumption. Suppose we have a p-dimensional multivariate Gaussian random vector $\boldsymbol{x}$ with the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$, that is, $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let us partition the vector into two parts, r-dimensional random vector of $\boldsymbol{x}_1$ and s-dimensional random vector of $\boldsymbol{x}_2$, where $r + s = p$. Then we have

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix}, \ \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \ \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \text{ and } \boldsymbol{\Omega} \stackrel{def}{=} \boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix}.$$

Using the property of the Gaussian distribution, we know that $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are independent if $\boldsymbol{\Sigma}_{12} = \mathbf{0}$. Then the precision matrix $\boldsymbol{\Omega}$ can be computed by

$$\boldsymbol{\Omega} = \begin{pmatrix} \left(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right)^{-1} & -\boldsymbol{\Sigma}_{11\cdot 2}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22\cdot 1}^{-1} \\ (-\boldsymbol{\Sigma}_{11\cdot 2}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22\cdot 1}^{-1})^T & \boldsymbol{\Sigma}_{22\cdot 1}^{-1}, \end{pmatrix}$$

where

$$\boldsymbol{\Sigma}_{11\cdot 2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}, \text{ and } \boldsymbol{\Sigma}_{22\cdot 1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}.$$

The conditional distribution of $\boldsymbol{x}_1$ given $\boldsymbol{x}_2$ is given by

$$\boldsymbol{x}_1|\boldsymbol{x}_2 \sim N_r\left(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2),\ \boldsymbol{\Sigma}_{11\cdot 2}\right),$$

Notice that $\boldsymbol{\Sigma}_{11\cdot 2}^{-1} = \boldsymbol{\Omega}_{11}$. Without loss of generality, let us assume that $r = 2$. Since

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} |x_3, \cdots, x_p \sim N_2\left(\boldsymbol{\mu}_{1,2|3,\cdots,p}, \boldsymbol{\Sigma}_{1,2|3,\cdots,p}\right),$$

if the $(1,2)$th element of $\boldsymbol{\Sigma}_{1,2|3,\cdots,p} = 0$, then $x_1 \perp\!\!\!\perp x_2|x_3, \cdots, x_p$. Now, noting that $\boldsymbol{\Sigma}_{1,2|3,\cdots,p} = \boldsymbol{\Omega}_{11}^{-1}$ since $\boldsymbol{\Omega}_{11} = \boldsymbol{\Sigma}_{1,2|3,\cdots,p}^{-1}$, if $\boldsymbol{\Omega}_{11}$ is diagonal, then $\boldsymbol{\Sigma}_{1,2|3,\cdots,p}$ is also diagonal. Thus, if the $(1,2)$th element of $\boldsymbol{\Omega}$ is zero, then $x_1 \perp\!\!\!\perp x_2|x_3, \cdots, x_p$. Extending this fact to the general r, it is clear that if the (i,j)th component of $\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Omega}$ is zero, then variables i and j are conditionally independent, given the other variables.

As an illustrative example, consider a 5-dimensional multivariate Gaussian random vector $\boldsymbol{x} = \{x_1, x_2, \cdots, x_5\}$ and we have an inverse covariance matrix of

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Omega} = \begin{pmatrix} \omega_{11} & \omega_{12} & \omega_{13} & \omega_{14} & \omega_{15} \\ & \omega_{22} & \omega_{23} & 0 & 0 \\ & & \omega_{33} & 0 & 0 \\ & & & \omega_{44} & \omega_{45} \\ & & & & \omega_{55} \end{pmatrix}. \tag{2.5}$$

Then using the Gaussian property, we can construct an undirected graph based on $\boldsymbol{\Omega}$ as Figure 2.4. The graph in Figure 2.4 has two maximal cliques given by $\{x_1, x_2, x_3\}$ and $\{x_1, x_4, x_5\}$, and we can also show that the joint distribution can be factorized by

$$p(\boldsymbol{x}) = g(x_1, x_2, x_3)h(x_1, x_4, x_5).$$

This property is extensively used for producing sparse undirected graphs with estimated precision matrices. In this dissertation, we also focus on imposing penalty for the estimation of $\boldsymbol{\Omega}$ to increase its sparsity. To highlight some of such estimation procedures in detail, we will review Lee and Liu (2015) and Danaher et al. (2014) in Sections 2.3 and 2.4, respectively.

2.3 Joint Estimation of Multiple Precision Matrices

We briefly introduce some notation that will be used throughout this dissertation. Suppose we have a data set with S subjects. The sth subject, $s = 1, \ldots, S$, is composed of $\{\boldsymbol{x}_1^{(s)}, \cdots, \boldsymbol{x}_{n_s}^{(s)}\}$, an independent and identically distributed random sample of size n_s, which forms the covariance matrix $\boldsymbol{\Sigma}^{(s)}$ and the corresponding precision matrix $\boldsymbol{\Omega}^{(s)} := (\boldsymbol{\Sigma}^{(s)})^{(-1)}$. Here, $\boldsymbol{x}_k^{(s)} = (x_{k1}^{(s)}, \cdots, x_{kp}^{(s)})^T$, $k = 1, \cdots, n_s$, is a p-dimensional random vector. For the rest of this article, we define some notations similar to those in Lee and Liu (2015) as follows. Let $\boldsymbol{I}$ denote a $p \times p$ identity matrix, $\hat{\boldsymbol{\Sigma}}^{(s)}$ be the sample covariance matrix, and $\boldsymbol{\Omega}^{(s)} = (\omega_{ij}^{(s)}); i, j = 1, \cdots, p; s = 1, \cdots, S$ be the precision matrix of the sth subject. For a matrix $\boldsymbol{X} = (x_{ij}) \in \mathbb{R}^{p \times q}$, $\|\boldsymbol{X}\|_1 = \sum_{i=1}^{p} \sum_{j=1}^{q} |x_{ij}|$ indicates the element-wise l_1 norm, $|\boldsymbol{X}|_\infty = \max_{1 \le i \le p, 1 \le j \le q} |x_{ij}|$ the element-wise l_∞ norm, $\|\boldsymbol{X}\|_{L_1} = \max_{1 \le j \le q} \sum_{i=1}^{p} |x_{ij}|$ the matrix l_1 norm, and $\|\boldsymbol{X}\|_F = \sqrt{\sum_i \sum_j x_{ij}^2}$ the Frobenius norm. Similarly, for a vector $\boldsymbol{x} = (x_1, \cdots, x_p)^T \in \mathbb{R}^p$, the vector l_1 and l_∞ norms are expressed by $\|\boldsymbol{x}\|_1$ and $|\boldsymbol{x}|_\infty$, respectively.

The Joint Estimation of Multiple Precision matrices (JEMP) method proposed by Lee and Liu (2015) is inspired by the Constrained L_1 minimization for Inverse Matrix Estimation (CLIME), introduced in Cai et al. (2011). The CLIME estimator of $\boldsymbol{\Omega}^{(s)}$ can be obtained by solving the following minimization problem:

$$\min \left\| \boldsymbol{\Omega}^{(s)} \right\|_1 \quad \text{s.t.} \quad \left| \hat{\boldsymbol{\Sigma}}^{(s)} - \boldsymbol{I} \right|_\infty \le \lambda_s, \tag{2.6}$$

where λ_s is a tuning parameter. We note that the CLIME estimator does not assume the Gaussian distribution. If the optimal solution in (2.6) is not symmetric, the method contains a symmetrization process in the last step of its algorithm. To estimate the precision matrices $\boldsymbol{\Omega}^{(1)}, \cdots, \boldsymbol{\Omega}^{(S)}$ using CLIME, one needs to conduct the optimization process in (2.6) for S times. However, Lee and Liu (2015) pointed out that such separate estimation may lose some information if individual precision matrices share a common structure. To circumvent this issue, Lee and Liu (2015) suggested JEMP, which takes a common structure into account and is also free of the Gaussian assumption. They define the common structure $\boldsymbol{C}$ of the S precision matrices and the unique structure $\boldsymbol{U}^{(s)}$ for the sth subject as

$$\boldsymbol{C} := \frac{1}{S}\sum_{s=1}^{S}\boldsymbol{\Omega}^{(s)}, \;\; \boldsymbol{U}^{(s)} := \boldsymbol{\Omega}^{(s)} - \boldsymbol{C}; \;\; s = 1, \cdots, S. \tag{2.7}$$

Then, $\sum_{s=1}^{S}\boldsymbol{U}^{(s)} = 0$ by Equation (2.7). By decomposing $\boldsymbol{\Omega}^{(s)}$ into common and unique features, the authors argued that JEMP produces a more accurate estimation of precision matrices than CLIME. Assuming $\boldsymbol{C}$ is sparse, they proposed the following constrained l_1 minimization criterion:

$$\begin{aligned} &\min\left\{\|\boldsymbol{C}\|_1 + \nu\sum_{s=1}^{S}\|\boldsymbol{U}^{(s)}\|_1\right\} \\ \text{s.t.} \;\; &\left|\frac{1}{S}\sum_{s=1}^{S}\{\hat{\boldsymbol{\Sigma}}^{(s)}(\boldsymbol{C}+\boldsymbol{U}^{(s)}) - \boldsymbol{I}\}\right|_{\infty} \le \lambda_1, \; \left|\hat{\boldsymbol{\Sigma}}^{(s)}(\boldsymbol{C}+\boldsymbol{U}^{(s)}) - \boldsymbol{I}\right|_{\infty} \le \lambda_2, \; \sum_{s=1}^{S}\boldsymbol{U}^{(s)} = 0, \end{aligned} \tag{2.8}$$

where ν is a predetermined weight, and λ_1 and λ_2 are tuning parameters. The first and second inequalities in (2.8) control how similar the final estimates are to the inverse covariance matrices on average and at the individual level, respectively. In the case of $\lambda_1 > \lambda_2$, because the second inequality implies the first inequality, the authors considered only $\lambda_1 \le \lambda_2$. Similar to CLIME, the final solution of JEMP also can be obtained after the following symmetrization

step. Define $\{\hat{\boldsymbol{C}}, \hat{\boldsymbol{U}}^{(1)}, \cdots, \hat{\boldsymbol{U}}^{(S)}\}$ as the solution to (2.8), and $\hat{\boldsymbol{\Omega}}^{(s)} = (\hat{\omega}_{ij}^{(s)}) := \hat{\boldsymbol{C}} + \hat{\boldsymbol{U}}^{(s)}$. Then, the symmetry can be achieved by choosing the element that has a smaller magnitude, i.e.,

$$\hat{\omega}_{ij}^{(s)*} = \hat{\omega}_{ij}^{(s)} I\left\{\sum_{s=1}^{S} |\hat{\omega}_{ij}^{(s)}| \leq \sum_{s=1}^{S} |\hat{\omega}_{ji}^{(s)}|\right\} + \hat{\omega}_{ji}^{(s)} I\left\{\sum_{s=1}^{S} |\hat{\omega}_{ij}^{(s)}| > \sum_{s=1}^{S} |\hat{\omega}_{ji}^{(s)}|\right\};\ s = 1, \cdots, S, \quad (2.9)$$

where $I\{\cdot\}$ is an indicator function. Lee and Liu (2015) also mentioned that the solution $\hat{\boldsymbol{\Omega}}^{(s)}$ is not necessarily positive definite, but it is positive definite with high probability in their numerical study.

In JEMP, we pay attention to the role of ν in (2.8). Lee and Liu (2015) argued that ν shows the complicity of the unique structures, and they set ν as S^{-1} or $S^{-1/2}$. According to their numerical examples, either value gives similar results. However, because a single value of ν implies equal weights for all subjects, the outcome may be disturbed if there exist outliers in terms of a pattern in precision matrices. Motivated by this, we propose JEMP with weighted aggregation in Section 3.1.

2.4 Joint Graphical Lasso

Danaher et al. (2014) suggested the Joint Graphical Lasso (JGL) method as an extension of the graphical lasso (Friedman et al., 2008) to the case of multiple classes based on maximizing a penalized Gaussian log-likelihood. Assuming that the $\sum_{s=1}^{S} n_s$ observations are independent and $\boldsymbol{x}_1^{(s)}, \cdots, \boldsymbol{x}_{n_s}^{(s)} \sim N(\boldsymbol{\mu}^{(s)}, \boldsymbol{\Sigma}^{(s)})$, the authors took the following penalized log-likelihood approach:

$$\max_{\{\boldsymbol{\Omega}\}} \left(\sum_{s=1}^{S} n_s \left[\log\left\{ \det\left(\boldsymbol{\Omega}^{(s)}\right)\right\} - \operatorname{tr}\left(\hat{\boldsymbol{\Sigma}}^{(s)} \boldsymbol{\Omega}^{(s)}\right)\right] - P\left(\{\boldsymbol{\Omega}\}\right)\right) \quad (2.10)$$

where $\{\mathbf{\Omega}\} = \mathbf{\Omega}^{(1)}, \cdots, \mathbf{\Omega}^{(S)}$ and subject to $\mathbf{\Omega}^{(1)}, \cdots, \mathbf{\Omega}^{(S)}$ are positive definite. Here $P(\{\mathbf{\Omega}\})$ denotes a convex penalty function. They introduced two choices of the penalty function $P(\cdot)$ in (2.10), which are the fused graphical lasso (FGL) and the group graphical lasso (GGL). First, the FGL has the form:

$$P(\{\mathbf{\Omega}\}) = \lambda_1 \sum_{s=1}^{S} \sum_{i \neq j} |\omega_{ij}^{(s)}| + \lambda_2 \sum_{s<s'} \sum_{i,j} |\omega_{ij}^{(s)} - \omega_{ij}^{(s')}|, \tag{2.11}$$

where λ_1 and λ_2 are tuning parameters. The FGL produces sparse estimates $\hat{\mathbf{\Omega}}^{(s)}$, $s = 1, \cdots, S$ when λ_1 is large, and most of the elements in $\hat{\mathbf{\Omega}}^{(s)}$, $s = 1, \cdots, S$ become identical when λ_2 is large (Tibshirani et al., 2005). Next, the GGL has the form:

$$P(\{\mathbf{\Omega}\}) = \lambda_1 \sum_{s=1}^{S} \sum_{i \neq j} |\omega_{ij}^{(s)}| + \lambda_2 \sum_{i \neq j} \left(\sum_{s=1}^{S} {\omega_{ij}^{(s)}}^2 \right)^{1/2}, \tag{2.12}$$

where λ_1 and λ_2 are tuning parameters again. Here the lasso penalty induces sparsity within each $\hat{\mathbf{\Omega}}^{(s)}$ and the group lasso penalty is applied to the (i, j)th element across all S precision matrices (Yuan and Lin, 2007), encouraging a similar pattern of sparsity across all subjects. In practice, the FGL in (2.11) provides very similar edge values throughout all the S matrices, while the GGL in (2.12) encourages only a shared pattern of sparsity, which results in a weaker form of similarity across the S precision matrices.

Since we are interested in estimating precision matrices not only limited to the group structure but also with the individual-wise uniqueness, we consider the GGL approach rather than FGL. As we did for JEMP in Section 2.3, we pay attention to the number of observations n_s in (2.10). If the number of observations is the same for all subjects as common in fMRI data, the JGL approach cannot adequately account for different contributions from different subjects. Although it considers the different sample size, still the outcome can be affected

by outliers with different pattern in precision matrices. We suggest JGL with regularized aggregation to mitigate the effect of such outliers in Section 3.2.

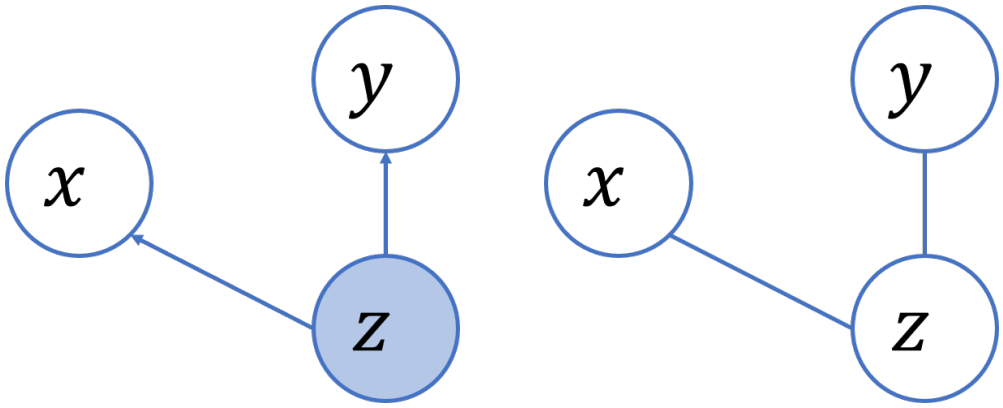
2.5 Regularized Aggregation

Wang et al. (2019) proposed a regularized aggregation method for statistical parametric maps in group-level analyses of fMRI data to find an optimal weight for aggregation, which aids in detecting and mitigating the effect of outlying subjects. In order to combine individual brain maps to produce accurate and robust group maps, the authors suggested a regularized unsupervised learning approach. Let $\boldsymbol{X}$ be a $p \times S$ design matrix, with p-dimensional random vectors from S subjects. The following optimization problem can determine subject-specific weight values $\boldsymbol{w} = (w^{(1)}, \cdots, w^{(S)})^T$:

$$\min_{\boldsymbol{y},\boldsymbol{w}} L(\boldsymbol{y}, \boldsymbol{w}) = \min_{\boldsymbol{y},\boldsymbol{w}} \left\{ \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2^2 + \lambda_0 \|\boldsymbol{w}\|_2^2 \right\} \quad \text{s.t.} \quad \sum_{s=1}^{S} w^{(s)} = 1, \quad w^{(s)} \geq 0; \quad s = 1, \cdots, S, \tag{2.13}$$

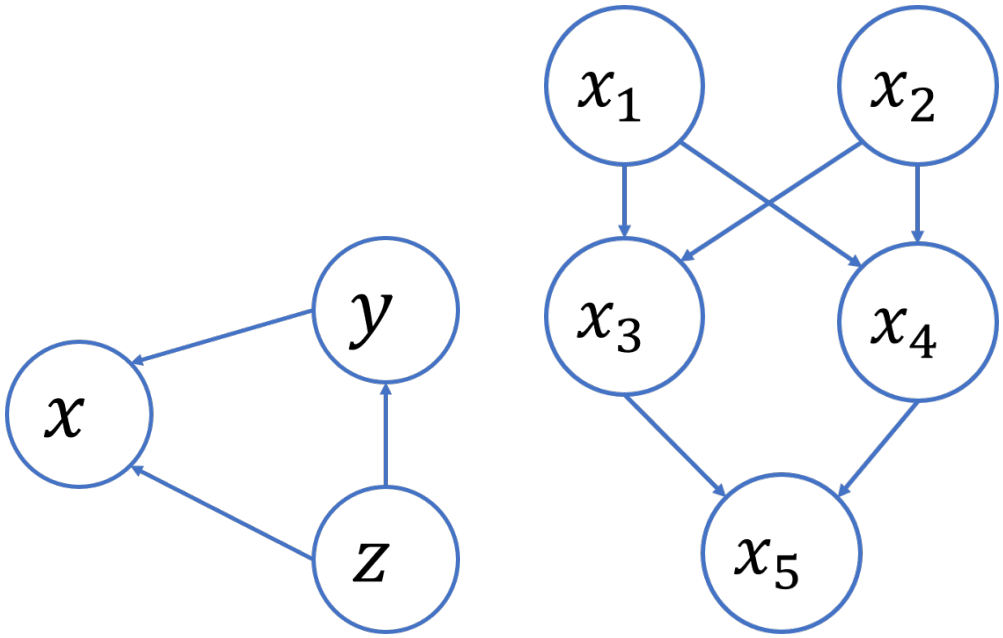
where $\boldsymbol{y}$ is an unobserved response vector, and $\lambda_0 > 0$ is a tuning parameter. Since $\boldsymbol{y}$ is unobserved, the optimization problem can be regarded as an unsupervised learning. In the algorithm, $\boldsymbol{y}$ is initialized as the median vector of $\boldsymbol{X}$ for robustness and is estimated with $\boldsymbol{w}$ iteratively. Wang et al. (2019) showed the following properties of the obtained weights from (2.13). The estimated weights can be interpreted as the individual effect of each subject to the whole group; they are similar to each other if subjects are similar, and they are small for the subjects that are dissimilar from the majority of the group. In other words, outlying subjects can be detected through the estimated weights. They also showed that the proposed method has a grouping property; if there exist multiple clusters in a group, the estimated weights are similar within each cluster but different between the clusters.

The main idea for this dissertation is to combine two different methodologies that jointly estimate inverse covariance matrices for individual subjects based on both Non-Gaussian and Gaussian assumptions, and aggregating those to provide representative and robust group graphs. As stated in Wang et al. (2019), aggregation approaches yield more interpretable result, generalized conclusion, and reduced computational demand of the analysis. Since outlying subjects can affect the estimated inverse covariance outcome, it is necessary to mitigate such outlying effect in aggregation. We introduce our methods in Chapter 3.



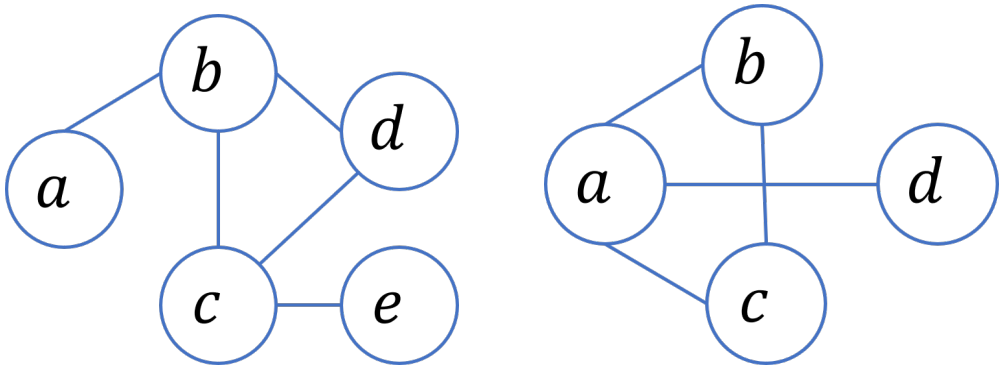
(a) A directed graph (b) An undirected graph

Figure 2.1: Examples for conditional independence. Each node represents a random variable and arrows / edges show probabilistic relationships.



(a) A directed graph of three variables (b) A directed acyclic graph (DAG)

Figure 2.2: Examples of directed graphical models. Each node represents a random variable and arrows show probabilistic relationships.



(a) An undirected graph (b) An undirected graph

Figure 2.3: Examples of undirected graphical models. Each node represents a random variable and edges show probabilistic relationships.

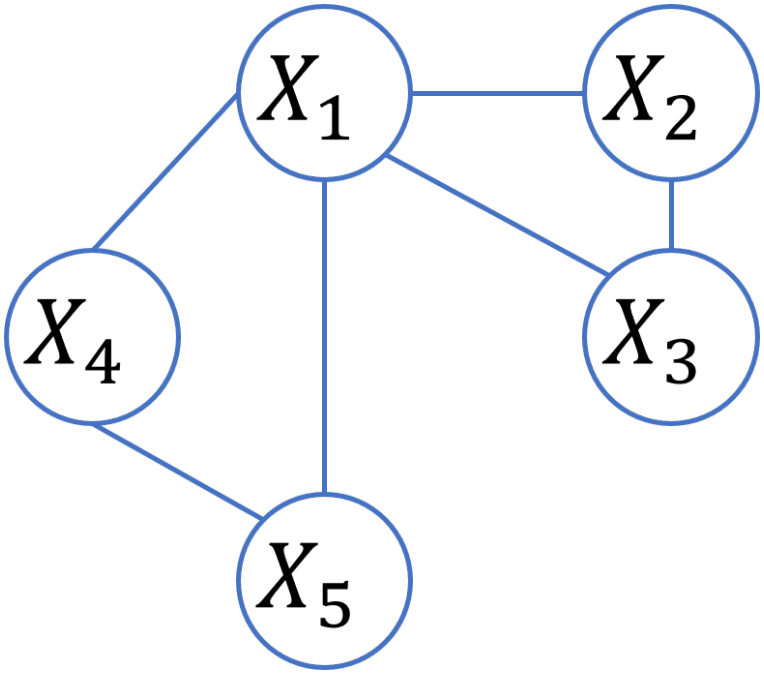


Figure 2.4: An undirected graph constructed from the inverse covariance matrix in equation (2.5).

Chapter 3

Methodology

In this chapter, we describe the proposed methods which can jointly estimate multiple precision matrices with regularized aggregation under both Non-Gaussian and Gaussian assumptions, inspired by Lee and Liu (2015), Danaher et al. (2014) and Wang et al. (2019). We also demonstrate pre-whitening procedure for real data analysis.

3.1 JEMP with RA

We describe the proposed joint estimation of multiple precision matrices (JEMP) with regularized aggregation (RA) method in this section. We also illustrate its detailed algorithm and tuning parameters selection.

For the precision matrices, we define the common structure $\boldsymbol{C}$ in a group and the unique structure $\boldsymbol{U}^{(s)}$ in the sth subject as

$$\boldsymbol{C} := \sum_{s=1}^{S} w^{(s)} \boldsymbol{\Omega}^{(s)}, \;\; \boldsymbol{U}^{(s)} := \boldsymbol{\Omega}^{(s)} - \boldsymbol{C}, \;\; \sum_{s=1}^{S} w^{(s)} = 1; \;\; s = 1, \cdots, S, \tag{3.1}$$

where $w^{(s)}$ is the weight for the sth subject, $\sum w^{(s)} = 1$, and $\sum_{s=1}^{S} w^{(s)} \boldsymbol{U}^{(s)} = 0$ by (3.1). We note that Lee and Liu (2015) consider a similar structure, but they use equal weights

$w^{(s)} = 1/S$. By defining the common structure as the weighted average of precision matrices, we expect a mitigated outlier effect. Then, the objective function can be written as:

$$\min\left\{\|\boldsymbol{C}\|_1 + \sum_{s=1}^{S} w^{(s)}\|\boldsymbol{U}^{(s)}\|_1\right\} \text{ s.t. } \left|\frac{1}{S}\sum_{s=1}^{S}\{\hat{\boldsymbol{\Sigma}}^{(s)}(\boldsymbol{C}+\boldsymbol{U}^{(s)}) - \boldsymbol{I}\}\right|_\infty \leq \lambda_1,$$
$$\left|\hat{\boldsymbol{\Sigma}}^{(s)}(\boldsymbol{C}+\boldsymbol{U}^{(s)}) - \boldsymbol{I}\right|_\infty \leq \lambda_2, \ \sum_{s=1}^{S} w^{(s)}\boldsymbol{U}^{(s)} = 0, \ \sum_{s=1}^{S} w^{(s)} = 1. \tag{3.2}$$

Note that the objective function in (3.2) does not contain any likelihood from the specific probability distribution; it solves the constraint l_1 minimization problem. This indicates that the estimation procedure is free from the Gaussian assumption. Since the first and second inequalities in the constraints do not change compared to (2.8), λ_1 and λ_2 have the same role. We symmetrize the estimates as in (2.9) because they still do not guarantee symmetry. The final estimates are positive definite with high probability in our numerical study. We determine weights $w^{(s)}$ using regularized aggregation introduced in Section 2.5.

3.1.1 Algorithm and Theoretical Property

We solve the optimization problems of (3.2) and (2.13) by a linear programming and a quadratic programming approach, respectively. In a similar way to Cai et al. (2011) and Lee and Liu (2015), the optimization problem in (3.2) can be decomposed into p separate problems:

$$\min\left\{\|\boldsymbol{c}_i\|_1 + \sum_{s=1}^{S} w^{(s)}\|\boldsymbol{u}_i^{(s)}\|_1\right\} \text{ s.t. } \left|\frac{1}{S}\sum_{s=1}^{S}\{\hat{\boldsymbol{\Sigma}}^{(s)}(\boldsymbol{c}_i+\boldsymbol{u}_i^{(s)}) - \boldsymbol{e}_i\}\right|_\infty \leq \lambda_1,$$
$$\left|\hat{\boldsymbol{\Sigma}}^{(s)}(\boldsymbol{c}_i+\boldsymbol{u}_i^{(s)}) - \boldsymbol{e}_i\right|_\infty \leq \lambda_2, \ \sum_{s=1}^{S} w^{(s)}\boldsymbol{u}_i^{(s)} = 0, \ \sum_{s=1}^{S} w^{(s)} = 1, \tag{3.3}$$

where $\{\boldsymbol{c}_i, \boldsymbol{u}_i^{(1)}, \cdots, \boldsymbol{u}_i^{(S)}\}$ is the ith column of $\{\boldsymbol{C}, \boldsymbol{U}^{(1)}, \cdots, \boldsymbol{U}^{(S)}\}$, and $\boldsymbol{e}_i$ is also the ith column of $\boldsymbol{I}$ for $i = 1, \cdots, p$. Cai et al. (2011) showed that the individual p minimization

problem in (3.3) is equivalent to the comprehensive problem in (3.2), but it is more computationally efficient. For the simulation and real data studies, we use the `fastclime` package (Pang et al., 2014) in `R` to achieve the optimal solutions. To solve the RA problem in (2.13), we use the `quadprog` package (Turlach and Weingessel, 2013) in `R`. The $p \times S$ design matrix $\boldsymbol{X}$ in (2.13) is composed of $\{\boldsymbol{c}_i + \boldsymbol{u}_i^{(1)}, \cdots, \boldsymbol{c}_i + \boldsymbol{u}_i^{(S)}\}$ for $i = 1, \cdots, p$, and both problems (3.2) and (2.13) can be optimized column-wise. During the iterations, the score vector $\boldsymbol{y}$ and the weights $w^{(s)}$ for $s = 1, \cdots, S$, are both updated until the likelihood $L(\boldsymbol{y}, \boldsymbol{w})$ becomes stable. The algorithm for JEMP with RA is provided in Algorithm 1.

Algorithm 1 JEMP with RA for finding optimal precision matrices and weights

1: Initialize $\boldsymbol{w}$ by homogeneous weights $(1/S)$
2: **if** $\lambda_1 \leq \lambda_2$ in (3.3) **then**
3: **for** Each column i **do**
4: Solve $\boldsymbol{c}_i, \boldsymbol{u}_i^{(1)}, \cdots, \boldsymbol{u}_i^{(S)}$ in (3.3)
5: Set $\boldsymbol{X}$ by $\{\boldsymbol{c}_i + \boldsymbol{u}_i^{(1)}, \cdots, \boldsymbol{c}_i + \boldsymbol{u}_i^{(S)}\}$
6: Solve $\boldsymbol{w}_i$ in (2.13)
7: **end for**
8: Symmetrize $\hat{\boldsymbol{\Omega}}^{(s)} := \hat{\boldsymbol{C}} + \hat{\boldsymbol{U}}^{(s)}; s = 1, \cdots, S$ in (2.9)
9: **end if**

As a result, we can obtain column-wise weights for each subject. Then, we obtain the final weight for each subject by averaging those column-wise weights.

In what follows, we evaluate the convergence rate of our proposed estimator JEMP with RA. For the property, we first construct the similar setting as in Cai et al. (2011) and Lee and Liu (2015). Assume the same size of samples, $n = n_1 = \cdots = n_S$, consider the following class of matrices,

$$\mathcal{U} := \{\boldsymbol{\Omega} : \boldsymbol{\Omega} \succ 0, \|\boldsymbol{\Omega}\|_{L_1} \leq C_M\},$$

and assume that $\boldsymbol{\Omega}^{(s)} \in \mathcal{U}$ for all $s = 1, \cdots, S$. Also we define $E(\boldsymbol{x}^{(s)}) = (\mu_1^{(s)}, \cdots, \mu_p^{(s)})^T$. The following Condtion 1 implies that the components of $\boldsymbol{x}^{(s)}$ are uniformly sub-Gaussian.

Condition 1. *There exists some* $0 < \eta < 1/4$ *such that* $E[\exp\{t(x_i^{(s)} - \mu_i^{(s)})^2\}] \leq K < \infty$ *for all* $|t| \leq \eta$ *and all* i, s *and* $S \log p/n \leq \eta$*, where* K *is a bounded constant.*

Then Theorem 1 guarantees that the convergence rate is the same as that of CLIME and JEMP estimator which are of order $(\log p/n)^{1/2}$.

Theorem 1. *Assume Condition 1 holds. Let* $\lambda_1 = \lambda_2 = 3C_M C_0 (\log p/n)^{1/2}$*, where* $C_0 = 2\eta^{-2}(2 + \tau + \eta^{-1}e^2K^2)^2$ *and* $\tau > 0$*. Then*

$$\max_{ij} \left(\sum_{s=1}^{S} w^{(s)} \left| \hat{\omega}_{ij}^{(s)} - \omega_{ij}^{(s)} \right| \right) \leq 6C_M^2 C_0 \left(\frac{\log p}{n} \right)^{1/2},$$

with probability greater than $1 - 4Sp^{-\tau}$*.*

In order to prove Theorem 1, we first state some results proved by Cai et al. (2011) in their Theorem 1.

Lemma 1. *Suppose Condition 1 holds. For any fixed* $s = 1, \cdots, S$*, with probability greater than* $1 - 4p^{-\tau}$*,*

$$\max_{ij} \left| \hat{\sigma}_{ij}^{(s)} - \sigma_{ij}^{(s)} \right| \leq C_0 \left(\frac{\log p}{n} \right)^{1/2},$$

where C_0 *is given in Theorem 1.*

Proof is provided below.

Proof of Theorem 1. It follows from Lemma 1 that

$$\max_{ij} \left| \hat{\sigma}_{ij}^{(s)} - \sigma_{ij}^{(s)} \right| \leq \lambda_2/(3C_M) \text{ for all } s = 1, \cdots, S, \tag{3.4}$$

with probability greater than $1 - 4Sp^{-\tau}$. All following arguments assume (3.4) holds. First, we have that

$$
\begin{aligned}
\left|\left(\hat{\boldsymbol{\Omega}}^{(s)}-\boldsymbol{\Omega}^{(s)}\right)\boldsymbol{e_j}\right|_\infty &= \left|\boldsymbol{\Omega}^{(s)}\left(\boldsymbol{\Sigma}^{(s)}\hat{\boldsymbol{\Omega}}^{(s)}-\boldsymbol{I}\right)\boldsymbol{e_j}\right|_\infty \le \left\|\boldsymbol{\Omega}^{(s)}\right\|_1\left|\left(\boldsymbol{\Sigma}^{(s)}\hat{\boldsymbol{\Omega}}^{(s)}-\boldsymbol{I}\right)\boldsymbol{e_j}\right|_\infty \\
&\le C_M\left|\left(\boldsymbol{\Sigma}^{(s)}\hat{\boldsymbol{\Omega}}^{(s)}-\hat{\boldsymbol{\Sigma}}^{(s)}\hat{\boldsymbol{\Omega}}^{(s)}+\hat{\boldsymbol{\Sigma}}^{(s)}\hat{\boldsymbol{\Omega}}^{(s)}-\boldsymbol{I}\right)\boldsymbol{e}_j\right|_\infty \\
&\le C_M\left\{\left|\left(\boldsymbol{\Sigma}^{(s)}-\hat{\boldsymbol{\Sigma}}^{(s)}\right)\hat{\boldsymbol{\Omega}}^{(s)}\boldsymbol{e}_j\right|_\infty+\left|\left(\hat{\boldsymbol{\Sigma}}^{(s)}\hat{\boldsymbol{\Omega}}^{(s)}-\boldsymbol{I}\right)\boldsymbol{e}_j\right|_\infty\right\} \\
&\le C_M\left\|\hat{\boldsymbol{\Omega}}^{(s)}\boldsymbol{e}_j\right\|_1\left|\boldsymbol{\Sigma}^{(s)}-\hat{\boldsymbol{\Sigma}}^{(s)}\right|_\infty+C_M\lambda_2 \\
&\le \left\|\hat{\boldsymbol{\Omega}}^{(s)}\boldsymbol{e}_j\right\|_1\lambda_2/3+C_M\lambda_2 \quad \forall\, s=1,\cdots,S.
\end{aligned}
$$

In addition, note that $\{\boldsymbol{C},\boldsymbol{U}^{(1)},\cdots,\boldsymbol{U}^{(S)}\}$ is a feasible solution of (3.2) since

$$
\begin{aligned}
\left|\boldsymbol{I}-\hat{\boldsymbol{\Sigma}}^{(s)}\left(\boldsymbol{C}+\boldsymbol{U}^{(s)}\right)\right|_\infty &= \left|\left(\boldsymbol{\Sigma}^{(s)}-\hat{\boldsymbol{\Sigma}}^{(s)}\right)\boldsymbol{\Omega}^{(s)}\right|_\infty \le \left\|\boldsymbol{\Omega}^{(s)}\right\|_1\left|\boldsymbol{\Sigma}^{(s)}-\hat{\boldsymbol{\Sigma}}^{(s)}\right|_\infty \\
&\le C_M\lambda_2/3C_M<\lambda_2=\lambda_1
\end{aligned}
$$

Hence, we have

$$
\begin{aligned}
\sum_{s=1}^S w^{(s)}\left|\left(\hat{\boldsymbol{\Omega}}^{(s)}-\boldsymbol{\Omega}^{(s)}\right)\boldsymbol{e}_j\right|_\infty &\le \sum_{s=1}^S w^{(s)}\left\{\left\|\hat{\boldsymbol{\Omega}}^{(s)}\boldsymbol{e}_j\right\|_1\lambda_2/3+C_M\lambda_2\right\} \\
&= \sum_{s=1}^S w^{(s)}\left\{\left\|\left(\hat{\boldsymbol{C}}+\hat{\boldsymbol{U}}^{(s)}\right)\boldsymbol{e}_j\right\|_1\lambda_2/3\right\}+C_M\lambda_2 \\
&\le \left\{\|\hat{\boldsymbol{c}}_j\|_1+\sum_{s=1}^S w^{(s)}\left\|\hat{\boldsymbol{u}}_j^{(s)}\right\|_1\right\}\lambda_2/3+C_M\lambda_2 \\
&\le \left\{\|\boldsymbol{c}_j\|_1+\sum_{s=1}^S w^{(s)}\left\|\boldsymbol{u}_j^{(s)}\right\|_1\right\}\lambda_2/3+C_M\lambda_2 \\
&\le 3C_M\lambda_2/3+C_M\lambda_2=2C_M\lambda_2=6{C_M}^2C_0\left(\frac{\log p}{n}\right)^{1/2}.
\end{aligned}
$$

By the inequality

$$\begin{aligned}\max_{ij}\left(\sum_{s=1}^{S} w^{(s)}\left|\hat{\omega}_{ij}^{(s)}-\omega_{ij}^{(s)}\right|\right) &\leq \max_{j}\sum_{s=1}^{S} w^{(s)}\left|\left(\hat{\mathbf{\Omega}}^{(s)}-\mathbf{\Omega}^{(s)}\right)\boldsymbol{e}_j\right|_{\infty} \\ &\leq 6{C_M}^2 C_0\left(\frac{\log p}{n}\right)^{1/2}.\end{aligned}$$

■

3.1.2 Tuning Parameters Selection

In our model, we have three tuning parameters; λ_0 from RA, as well as λ_1 and λ_2 from JEMP. We use a grid search restricted to $\lambda_0 > 0$ and $0 < \lambda_1 \leq \lambda_2$ to find the optimal combination. Although grid search is not powerful in computation, it is simple and can be conducted in parallel (Bergstra and Bengio, 2012). As a criterion of performance comparison, we consider the likelihood loss (LL) suggested by Cai et al. (2011):

$$LL = \sum_{s=1}^{S} \operatorname{tr}\left(\hat{\mathbf{\Sigma}}_v^{(s)}\hat{\mathbf{\Omega}}^{(s)}\right) - \log\left\{\det\left(\hat{\mathbf{\Omega}}^{(s)}\right)\right\}, \tag{3.5}$$

where $\hat{\mathbf{\Sigma}}_v^{(s)}$ is the covariance matrix of the sth subject from a validation set, $\hat{\mathbf{\Omega}}^{(s)}$ is the estimated precision matrix of the sth subject from a training set, and det stands for the determinant of a matrix. After comparing various combinations of tuning parameters values, we choose the combination that achieves the minimum value of LL in (3.5). In case that a validation set is not feasible, we slightly modify (3.5) similar to cross-validation:

$$LL^* = \operatorname{tr}\left(\bar{\hat{\mathbf{\Sigma}}}^{(v)}\bar{\hat{\mathbf{\Omega}}}^{(t)}\right) - \log\left\{\det\left(\bar{\hat{\mathbf{\Omega}}}^{(t)}\right)\right\}, \tag{3.6}$$

where $\bar{\hat{\mathbf{\Sigma}}}^{(v)}$ is the covariance matrix averaged from the randomly chosen validation set described below, and $\bar{\hat{\mathbf{\Omega}}}^{(t)}$ is the estimated precision matrix averaged from the remaining training

set. We first randomly pick some portions R of the data and assign those as a validation set and treat the remainder as a training set. Then, we evaluate LL^* for every combination of tuning parameters, find the optimal values which minimize LL^*, and repeat this process K times with differently chosen validation sets. The final optimal combination is selected by averaging K combinations of λ_0, λ_1, and λ_2 values. With the optimal combination, we obtain the final group precision matrix and build the corresponding graphical model using the complete data. In our simulation and real data analysis, we use $R = 0.2$ and $K = 10$.

3.2 JGL with RA

We propose the joint graphical lasso (JGL) with regularized aggregation (RA) method in this section. Different from JEMP, the method requires the Gaussian distribution assumption. The algorithm and tuning parameters selection for the method are also described.

With the Gaussian distribution assumption, the weighted penalized log-likelihood approach can be written as:

$$\max_{\{\mathbf{\Omega}\}} \left(\sum_{s=1}^{S} w^{(s)} \left[\log \left\{ \det \left(\mathbf{\Omega}^{(s)} \right) \right\} - \operatorname{tr} \left(\hat{\mathbf{\Sigma}}^{(s)} \mathbf{\Omega}^{(s)} \right) \right] - P\left(\{\mathbf{\Omega}\} \right) \right), \tag{3.7}$$

where the penalty has the form of GGL, which is

$$P\left(\{\mathbf{\Omega}\} \right) = \lambda_1 \sum_{s=1}^{S} \sum_{i \neq j} |\omega_{ij}^{(s)}| + \lambda_2 \sum_{i \neq j} \left(\sum_{s=1}^{S} {\omega_{ij}^{(s)}}^2 \right)^{1/2}. \tag{3.8}$$

Notice that Danaher et al.'s JGL with GGL penalty is a special case of $w^{(s)} = n_s$. If subjects are homogeneous, which means that if they have similar inverse covariance structure, then the estimated weights will also be similar. Additionally, if some subjects have different inverse covariance structure from the majority of the group, their weights will be estimated low. We expect a reduced effect of outlier by assigning subject-wise different weights in the objective

function in (3.7). The proposed JGL with RA method does not minimize the norm, and solves the weighted penalized likelihood function. Let $\hat{\boldsymbol{\Omega}}^{(s)}_{PMLE}$, $s = 1, \cdots, S$ denotes the penalized maximum likelihood estimators for each subject from the given objective function in (3.7). Then the proposed method allocates different weights to each subject by comparing median$(\hat{\boldsymbol{\Omega}}^{(1)}_{PMLE}, \cdots, \hat{\boldsymbol{\Omega}}^{(S)}_{PMLE})$ ($\boldsymbol{y}$ in equation (2.13)) with each $\hat{\boldsymbol{\Omega}}^{(s)}_{PMLE}$, $s = 1, \cdots, S$. Thus, the weights are smaller for the subjects with the estimated inverse covariances which are much deviated from the estimated median inverse covariance of the group. Since we do not change the GGL penalty in (3.8), the tuning parameters λ_1 and λ_2 have the same role as in (2.12).

3.2.1 Algorithm and Tuning Parameters Selection

Following Danaher et al. (2014), we use an alternating directions method of multipliers (ADMM) algorithm (Boyd et al., 2011) to solve the problem (3.7). Specifically, we implement the `JGL` package (Danaher, 2018) in `R` and modify it to solve the optimization problem. The objective function can be rewritten as:

$$\min_{\{\boldsymbol{\Omega}\},\{\boldsymbol{Z}\}} \left(-\sum_{s=1}^{S} w^{(s)} \left[\log \left\{ \det \left(\boldsymbol{\Omega}^{(s)} \right) \right\} - \operatorname{tr} \left(\hat{\boldsymbol{\Sigma}}^{(s)} \boldsymbol{\Omega}^{(s)} \right) \right] + P\left(\{\boldsymbol{Z}\} \right) \right),$$

where $\boldsymbol{Z}^{(s)} = \boldsymbol{\Omega}^{(s)}$ for $s = 1, \cdots, S$, $\{\boldsymbol{Z}\} = \boldsymbol{Z}^{(1)}, \cdots, \boldsymbol{Z}^{(S)}$ and subject to $\{\boldsymbol{\Omega}\}$, $\{\boldsymbol{Z}\}$ are positive definite.

Then the scaled augmented Lagrangian (Boyd et al., 2011) is:

$$\begin{aligned} L_\rho\left(\{\boldsymbol{\Omega}\}, \{\boldsymbol{Z}\}, \{\boldsymbol{A}\} \right) \;=\; & -\sum_{s=1}^{S} w^{(s)} \left[\log \left\{ \det \left(\boldsymbol{\Omega}^{(s)} \right) \right\} - \operatorname{tr} \left(\hat{\boldsymbol{\Sigma}}^{(s)} \boldsymbol{\Omega}^{(s)} \right) \right] + P\left(\{\boldsymbol{Z}\} \right) \\ & + \frac{\rho}{2} \sum_{s=1}^{S} \left\| \boldsymbol{\Omega}^{(s)} - \boldsymbol{Z}^{(s)} + \boldsymbol{A}^{(s)} \right\|_F^2 - \frac{\rho}{2} \sum_{s=1}^{S} \left\| \boldsymbol{A}^{(s)} \right\|_F^2, \end{aligned}$$

where $\{\boldsymbol{A}\} = \boldsymbol{A}^{(1)}, \cdots, \boldsymbol{A}^{(S)}$ are dual variables and $\rho > 0$ controls the step size and convergence criterion. We describe the detailed ADMM algorithm for JGLRA with RA:

1. Initialize $\boldsymbol{\Omega}^{(s)} = \boldsymbol{I}$, $\boldsymbol{Z}^{(s)} = \boldsymbol{A}^{(s)} = \boldsymbol{0}$, and $w^{(s)} = 1/S$ for $s = 1, \cdots, S$.

2. Choose a scalar $\rho > 0$.

3. For $i = 1, 2, 3, \cdots$ th iteration, update the variables until convergence:

   (a) Update $\boldsymbol{\Omega}_{(i)}^{(s)}$ for $s = 1, \cdots, S$ as the minimizer of

   $$-w^{(s)} \left[\log \left\{ \det \left(\boldsymbol{\Omega}^{(s)} \right) \right\} - \operatorname{tr} \left(\hat{\boldsymbol{\Sigma}}^{(s)} \boldsymbol{\Omega}^{(s)} \right) \right] + \frac{\rho}{2} \left\| \boldsymbol{\Omega}^{(s)} - \boldsymbol{Z}_{(i-1)}^{(s)} + \boldsymbol{A}_{(i-1)}^{(s)} \right\|_F^2 .$$

   Let $\boldsymbol{Q}\boldsymbol{D}\boldsymbol{Q}^{-1}$ be the eigendecomposition of $\hat{\boldsymbol{\Sigma}}^{(s)} - \rho \boldsymbol{Z}_{(i-1)}^{(s)}/w^{(s)} + \rho \boldsymbol{A}_{(i-1)}^{(s)}/w^{(s)}$, the solution can be obtained by $\boldsymbol{Q}\tilde{\boldsymbol{D}}\boldsymbol{Q}^{-1}$ (Witten and Tibshirani, 2009), where $\tilde{\boldsymbol{D}}$ is the diagonal matrix with jth diagonal element

   $$\frac{w^{(s)}}{2\rho} \left\{ -d_{jj} + \left(d_{jj}^2 + 4\rho/w^{(s)} \right)^{1/2} \right\} .$$

   (b) Update $\left\{ \boldsymbol{Z}_{(i)} \right\}$ as the minimizer of

   $$\frac{\rho}{2} \sum_{s=1}^{S} \left\| \boldsymbol{Z}^{(s)} - \boldsymbol{B}^{(s)} \right\|_F^2 + P\left(\{\boldsymbol{Z}\} \right)$$

   where

   $$\begin{aligned} \boldsymbol{B}^{(s)} &= \left(\boldsymbol{\Omega}_{(i)}^{(s)} + \boldsymbol{A}_{(i-1)}^{(s)} \right), \text{ and} \\ P\left(\{\boldsymbol{Z}\} \right) &= \lambda_1 \sum_{s=1}^{S} \sum_{i \neq j} |z_{ij}^{(s)}| + \lambda_2 \sum_{i \neq j} \left(\sum_{s=1}^{S} {z_{ij}^{(s)}}^2 \right)^{1/2} . \end{aligned}$$

   The solution for the diagonal elements can be obtained by $\hat{z}_{jj}^{(s)} = b_{jj}^{(s)}$ for all $j = 1, \cdots, p$ and $s = 1, \cdots, S$. By following Friedman et al. (2010b), the off-

diagonal elements take the form of

$$\hat{z}_{ij}^{(s)} = \mathcal{S}(b_{ij}^{(s)}, \lambda_1/\rho) \left(1 - \frac{\lambda_2}{\rho \left\{ \sum_{s=1}^{S} \mathcal{S}(b_{ij}^{(s)}, \lambda_1/\rho)^2 \right\}^{1/2}} \right)_+ ,$$

where $\mathcal{S}$ indicates the soft thresholding operator and $(x)_+$ denotes $\max(0, x)$.

(c) Update $\boldsymbol{A}_{(i)}^{(s)} = \boldsymbol{A}_{(i-1)}^{(s)} + \boldsymbol{\Omega}_{(i)}^{(s)} - \boldsymbol{Z}_{(i)}^{(s)}$ for $s = 1, \cdots, S$.

(d) Set $\boldsymbol{X}$ by vectorized $\left\{ \boldsymbol{\Omega}_{(i)}^{(1)}, \cdots, \boldsymbol{\Omega}_{(i)}^{(S)} \right\}$, and solve $\boldsymbol{w}_{(i)}$ using regularized aggregation in (2.13). Update each $w_{(i)}^{(s)}$ followed by $\boldsymbol{w}_{(i)}$.

Boyd et al. (2011) showed that the algorithm guarantees the convergence to the global optimum, and the positive definiteness constraint is enforced by the step 3(a) (Danaher et al., 2014). We employ $\rho = 1$ throughout the examples in this dissertation and use convergence criterion as

$$\sum_{s=1}^{S} \left\| \boldsymbol{\Omega}_{(i)}^{(s)} - \boldsymbol{\Omega}_{(i-1)}^{(s)} \right\|_1 / \sum_{s=1}^{S} \left\| \boldsymbol{\Omega}_{(i-1)}^{(s)} \right\|_1 < 10^{-5}.$$

Along with the general model selection criterion such as Akaike information criterion (AIC), Bayesian information criterion (BIC) and cross-validation, the network model selection can be improved when it is accompanied by empirical principles, such as network interpretability, consistency and a low false discovery rate (Meinshausen and Bühlmann, 2010; Li et al., 2013). In this sense, Danaher et al. (2014) recommended an application-driven selection of tuning parameters that can provide biologically reasonable interpretation with sufficient complexity to be interesting and sufficient sparsity to be interpretable. Following this idea, we apply LL^* suggested in Section 3.1.2 in equation (3.6) with the range of a grid which provides plausible interpretation. In the simulation study and the real data analysis, respectively in Chapter 4 and 5, the proposed JGL with RA method tends to select the

tuning parameters that produce more dense network rather than sparse within the range of the grid.

3.3 Pre-whitening

In the real data analysis, we consider a pre-whitening procedure to account for temporal correlation in fMRI data before applying the proposed methods. We employ a wavelet transform as a pre-whitening procedure, which can reduce temporal correlation in the voxel time series (Stoev et al., 2005). A wavelet decomposition can reduce temporal correlation so that the coefficients are much decorrelated than the corresponding time series (Flandrin, 1992; Kaplan and Kuo, 1993; Li et al., 2002). We apply linear binning technique described in Wand (1994) to adjust the time points to 128 (2^7) by using the `npsp` package (Fernandez-Casal, 2019) in `R`. Then we use the `wavethresh` package (Nason, 2016) in `R` to perform the wavelet transform. More details about the wavelet transform and the corresponding `R` package are available in the book written by Nason (2010).

Chapter 4

Simulation Studies

4.1 JEMP with RA

In this section, we describe a simulation study to show the properties of the proposed method and to compare its performance with JEMP (Lee and Liu, 2015) in various simulation settings. In all settings, we set $p = 10$ which can be considered as the number of regions of interest (ROIs) in fMRI data, and we set $S = 15$ as the number of subjects. Then, we repeat each simulation 100 times. For tuning parameters, we adopt 10 different values for λ_0, λ_1, and λ_2 in $0 < \lambda_0 \leq 2$ and $0.2 \leq \lambda_1 \leq \lambda_2 \leq 0.5$. The λ_1 and λ_2 values determine the sparsity of the estimated precision matrices. If the values are extremely small or large, the resulting matrices are too dense or sparse, respectively. Thus, we select a moderate range of values for each tuning parameter grid. Among $1,000$ different combinations of the tuning parameters values, we select the optimal combination, which minimizes LL^* as explained in Section 3.1.2. We use the same values of tuning parameters for JEMP.

4.1.1 Case 1: Homogeneous subjects

For a simulation study, we generate covariance matrices using the Gaussian distribution. We construct numerical data that resemble typical fMRI data, where most of the elements in the corresponding covariance matrices are positive. In specific, we create 15 independent 10×10 matrices $\{A^{(1)}, \cdots, A^{(15)}\}$ where each element is from $N(-0.01, 0.2^2)$. Then, the covariance matrices are generated by $\hat{\mathbf{\Sigma}}^{(s)} = A^{(s)^T} A^{(s)}; s = 1, \cdots, 15$ to guarantee the positive definiteness. The first simulation setting checks whether the proposed method yields similar weights when subjects are homogeneous.

Table 4.1 shows the means and standard deviations (in parentheses) of the estimated weights by JEMP with RA from 100 repetitions for homogeneous subjects. As can be seen in Table 4.1, our method successfully provides the estimated weights ranging from 0.0661 to 0.0670, which are close to the equal weight of $1/15 = 0.0667$. Also, the small standard deviations ranging from 0.0009 to 0.0040 indicate that the obtained weights are reliable.

We select one example from the 100 repetitions to demonstrate a group graph (a common structure) estimated by JEMP with RA and compare it with that of JEMP. Figure 4.1 displays two group graphs for homogeneous subjects. In the graphs, the edges are dropped if the corresponding $\hat{\omega}_{ij}$ is less than 0.3. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs. We observe that Figures 4.1(a) and 4.1(b) are the same, which indicates that our proposed method can provide an appropriate group graph for homogeneous subjects because the simple average of individual precision matrices (JEMP) is optimal in this case.

4.1.2 Case 2: Three outlying subjects

In this case, we consider three outlying subjects to see how the JEMP with RA method handles outliers. The 12 standard subjects are generated in the same way as Case 1, but for

the three outlying subjects, we set elements of $A^{(s)}$ from $N(-0.01, 0.04^2)$ and corresponding covariance matrix by $\hat{\mathbf{\Sigma}}^{(s)} = {A^{(s)}}^T A^{(s)}; s = 13, 14, 15$. The outlying subjects have much smaller variance than the standard subjects, which implies nearly flat signals in fMRI data.

Table 4.2 demonstrates the means and standard deviations (in parentheses) of the estimated weights by our method over 100 repetitions for the three outlying subjects case. The estimated mean weights of the three outlying subjects are much smaller than those of the 12 standard subjects, ranging from 0.0034 to 0.0042. On the other hand, the standard subjects have large assigned weights, ranging from 0.0811 to 0.0833. We can see that both standard and outlying groups show similar weights within each group, which suggests that the proposed method possesses a grouping property as also shown in Wang et al. (2019). This simulation demonstrates that the proposed JEMP with RA method can detect multiple outliers through the estimated weights.

Figure 4.2 displays one example of the JEMP with RA and that of JEMP group graphs which are picked from 100 repetitions for Case 2. In order to assess the effect of the outliers to the two methods, we add another group graph estimated by JEMP without the three outlying subjects. For that graph, we generate additional three standard subjects to eliminate sample size effect. The three graphs are drawn using the same λ_1 and λ_2 values. Figure 4.2 shows that outliers can greatly affect a group graph particularly for JEMP. For example, the Figures 4.2(a) and 4.2(b) show some associations with ROI 1 and ROI 2 while the JEMP group graph in Figure 4.2(c) displays none of them. It can be seen that Figure 4.2(a) is closer to the group graph from the JEMP without outliers in Figure 4.2(b) than Figure 4.2(c) from JEMP. Also, several associations in Figure 4.2(b) disappear in Figure 4.2(c) due to the effect of the outliers. This result shows that outliers can have a substantial impact on creating a group graph from individual ones, which can be misleading in interpreting the relationship between the nodes.

4.1.3 Case 3: Multiple clusters

In this setting, we generate three different clusters that are composed of subjects with different relationship structures. For $A^{(s)}; s = 1, \cdots, 15$ matrices, the first 8 subjects are from $N(-0.01, 0.2^2)$ as usual, but the next 4 subjects come from $N(-0.01, 0.09^2)$, and the last 3 subjects come from $N(-0.01, 0.04^2)$. Likewise, the covariance matrices are constructed by $\hat{\mathbf{\Sigma}}^{(s)} = A^{(s)^T} A^{(s)}; s = 1, \cdots, 15$ to ensure the positive definiteness. It is difficult to conclude that any subjects in the three clusters are outliers in this case.

Table 4.3 displays the means and standard deviations (in parentheses) of the estimated weights by JEMP with RA from 100 repetitions for the multiple clusters setting. We observe clear weight differences among the three clusters. The weights for the first cluster range between 0.0896 and 0.0906; those of the second cluster range between 0.0647 and 0.0661; and those of the last cluster range between 0.0047 and 0.0068. This result again demonstrates a grouping property; our method can identify the three clusters that are composed of different structures and assign different weights for each cluster, but the similar weights within each cluster.

Figure 4.3 compares the group graphs generated by JEMP with RA and JEMP from one example selected from 100 repetitions for the multiple clusters setting. We can see a significant disparity between the two group graphs. Several strong associations that appear in the JEMP with RA graph in Figure 4.3(a) disappear in the JEMP graph in Figure 4.3(b). By applying equal weights, JEMP keeps a small number of associations in the common structure. On the other hand, JEMP with RA uses weighted weights in the estimation process that captures more relationships in its group graph.

In conclusion, the proposed JEMP with RA method captures dissimilarities between subjects if any, and allocates larger weights to the majority cluster and smaller weights to the minority or outlying clusters in various simulation settings. Also, the proposed method assigns similar weights within each cluster because the subjects are homogeneous within a

cluster. Using the assigned weights, JEMP with RA produces a more representative and robust group graph than JEMP by lowering the outlying effects. In our simulation settings, the outlying subjects produce more sparse graph compared to the standard subjects. JEMP also tends to produce more sparse group graph than the proposed method due to the effect of such outliers.

4.2 JGL with RA

We perform various simulation experiments to evaluate the performance of the proposed JGL with RA method and to compare it with JGL (Danaher et al., 2014) under several different scenarios. As we did in Section 4.1, we set $p = 10$ as the number of ROIs and $S = 15$ as the number of subjects in all simulation settings. Then we repeat each simulation 100 times to check consistency among different simulation settings. For tuning parameters, we apply LL^* suggested in Section 3.1.2 in equation (3.6) with the range of a grid chosen to provide reasonable interpretation with adequate sparsity: 10 different values for λ_0, λ_1, and λ_2 in $0 < \lambda_0 \leq 2$ and $0.0026 \leq \lambda_1, \lambda_2 \leq 0.022$. Within the grid, the optimal combination of the tuning parameters which minimizes LL^* is selected, and we obtain results based on those tuning parameter values. For comparison with JGL, we use the same tuning parameters.

4.2.1 Case 1: Homogeneous subjects

Similar to the simulation study conducted in Section 4.1, we construct covariance matrices that resemble typical fMRI data using the Gaussian distribution. First we create 15 independent 10×10 matrices $A^{(s)} \sim N(-0.01, 0.08^2)$ and generate the corresponding sample covariance matrices by $\hat{\mathbf{\Sigma}}^{(s)} = A^{(s)^T} A^{(s)}$, $s = 1, \cdots, S$ to satisfy the positive definiteness condition. The first simulation setting is called homogeneous subjects, since all of the subjects in the group have the same covariance structure. In this setting, we show the properties of the

proposed method when the group is composed of homogeneous subjects. Figure 4.4 illustrates the covariance heat maps for Subject 1 and 15 in one example from the 100 repetitions in this setting. The element with high covariance is colored in red in the heat map. In Figure 4.4, the two heat maps show similar light-red colored tables since they are from the same distribution.

Table 4.4 describes the vectorized Euclidean distances between the median and each subject's estimated inverse covariance along with the estimated weights by JGL with RA for homogeneous subjects. The distances, means, and standard deviations (in parentheses) of the estimated weights are from 100 repetitions. In Table 4.4, the distances range from 1.7422 to 1.8856 with the marginal differences, and the estimated weights are close to the equal weight of $1/15 = 0.0667$. In addition, exceptionally small standard deviations (from 4e-04 to 1e-03) indicate that the proposed method provides consistent results through the 100 repetitions.

We choose one example from the 100 repetitions and provide group graphs by the proposed method and JGL from homogeneous subjects in Figure 4.5. In the construction of the graphs, we use 0.05 as a threshold for including edges. We use different size of nodes to depict the number of edges that are connected to the corresponding nodes, and also use different thickness of edges to represent the strength of relationships between two nodes. In Figure 4.5, the two graphs in Figures 4.5(a) and 4.5(b) are the same in terms of the number of edges and the strength of the associations. The result implies that the proposed method can assign similar weights to the subjects and produce a suitable group graph for homogeneous subjects, which is the same as the group graph that is estimated by JGL.

4.2.2 Case 2: Small covariance outlying subjects

Next, we investigate performance of the JGL with RA method for the effect of outliers who have a smaller covariance structure than the standard subjects. The standard 12 subjects are

generated from $A^{(s)} \sim N(-0.01, 0.2^2)$; $\hat{\mathbf{\Sigma}}^{(s)} = {A^{(s)}}^T A^{(s)}$, $s = 1, \cdots, 12$, while for the outlying subjects we set $A^{(s)} \sim N(-0.01, 0.04^2)$, and respective covariance matrix by $\hat{\mathbf{\Sigma}}^{(s)} = {A^{(s)}}^T A^{(s)}$, $s = 13, 14, 15$ to guarantee the positive definiteness. The outlying subjects represent the subjects who show flat signals throughout the experiment in fMRI data. Figure 4.6 displays two subjects' covariance heat maps from one case picked from the 100 repetitions in this setting. In Figure 4.6, we observe that the standard Subject 1 has the covariance structure with dark red colors whereas the outlying Subject 15 shows light red colors. The two heat maps depict a clear difference in the covariance structures between the standard and outlying subjects.

In Table 4.5, we report the vectorized Euclidean distances between the median and each subject's estimated inverse covariance, the means and standard deviations (in parentheses) of the estimated weights by JGL with RA from 100 repetitions for the small covariance outlying subjects case. For the standard subjects, the mean distances range between 3.4567 and 3.640 and the corresponding mean weights are assigned from 0.0758 to 0.0782. On the other hand, the mean distances and weights for the outlying subjects range from 10.3835 to 10.3983 and around 0.0254, respectively. The result implies that the weights become smaller when one's estimated inverse covariance is much diverged from the estimated median inverse covariance. The proposed JGL with RA method still shows a grouping property (Wang et al., 2019), since the estimated weights within each of the standard and outlying subjects are similar. From this simulation, we show that the proposed method assigns smaller weights to the outliers and mitigate their effects in the aggregation.

Figure 4.7 illustrates the group graphs estimated by the proposed method and JGL from one example picked from 100 repetitions for the small covariance outliers setting. We also add another group graph generated by JGL without the three small covariance outliers in Figure 4.7(b) to evaluate the effect of the outliers. In the estimation, the same tuning parameter values are used for comparison. We observe that the group graph estimated by the proposed

method in Figure 4.7(a) has the shape that is closer to the group graph by JGL without outliers in Figure 4.7(b) than Figure 4.7(c). The group graph estimated by JGL in Figure 4.7(c) is too sparse; it has insufficient structure to be interesting, does not well represent the data due to the effect of outliers, and show clear disparity with Figure 4.7(b). The result is similar to that of JEMP in Section 4.1; that is, the proposed JGL with RA method yields a more robust group graph estimate to the outliers that have abnormally sparse inverse covariance than the standard subjects. This simulation also demonstrates that such outlying effect should be reduced to create a graph that can better represent a group.

4.2.3 Case 3: Large covariance outlying subjects

We consider three outlying subjects with large covariance to examine how the JGL with RA method controls the effect of such outliers. In this case, we set $A^{(s)} \sim N(-0.01, 0.08^2)$; $s = 1, \cdots, 12$ for the 12 standard subjects and $A^{(s)} \sim N(-0.01, 0.25^2)$; $s = 13, 14, 15$ for the three outlying subjects. Then, the covariance matrices are created by $\mathbf{\Sigma}^{(s)} = {A^{(s)}}^T A^{(s)}$, $s = 1, \cdots, 15$ to ensure positive definiteness. The outlying subjects have much larger variance than the standard subjects, so they can be considered as the subjects showing over responsive signals. In Figure 4.8, we illustrate two covariance heat maps from one example from the 100 repetitions for the large covariance outliers setting. The two heat maps show clear distinction in terms of their colors, light-red for the standard subjects and dark-red for the outlying subjects, respectively.

Table 4.6 demonstrates the vectorized Euclidean distances between the median and each subject's estimated inverse covariance, the means, and standard deviations (in parentheses) of the estimated weights by our method from 100 repetitions for the large covariance outlying subjects setting. The 12 standard subjects have large assigned weights, ranging from 0.0792 to 0.0807, and they have relatively short distances, ranging between 2.0462 and 2.2047. On the contrary, the estimated mean weights and distances for the three outlying subjects are

much smaller and longer than those of the 12 standard subjects, ranging from 0.0144 to 0.0146 and from 15.7491 to 15.7758, respectively. The result again shows an inverse relationship between the estimated weights and distances. If a subject's estimated inverse covariance is deviated from the estimated median inverse covariance from the group, the estimated weight for that subject becomes small. We again note that the estimated weights for both standard and outlying groups are similar within each group; thus a grouping property holds for this case. Throughout this simulation, we show that the proposed JGL with RA method can identify multiple outliers by assigning different weights based on the distances from the median inverse covariance of the group.

We provide group graphs by the proposed method and JGL for the large covariance outlying subjects setting from one example selected from 100 repetitions in Figure 4.9. In a similar way to Figure 4.7(b), we assess the effect of outliers by adding Figure 4.9(b) which is another group graph estimated by JGL without the large covariance outliers. The result is opposite to the small covariance outlying subjects setting in Section 4.2.2: the group graph by JGL with RA and JGL without outliers respectively in Figures 4.9(a) and 4.9(b) are more sparse than the group graph by JGL in Figure 4.9(c). The group graph constructed by JGL estimation in Figure 4.9(c) is too dense to infer biologically plausible interpretation. On the other hand, the group graph estimated by the proposed JGL with RA method in Figure 4.9(a) is more similar to Figure 4.9(b) than Figure 4.9(c), by reducing the effect of outliers with dense inverse covariance. The result in this setting demonstrates that the proposed method can provide more interpretable outcome that is less affected by the outliers with large covariance.

4.2.4 Case 4: Multiple clusters

We generate a group consisting of three different clusters with different covariance structures to examine the property of the proposed JGL with RA method in this setting. Specifically,

we set elements in $A^{(s)}; s = 1, \cdots, 15$ matrices as follows: the first 10 subjects are from $N(-0.01, 0.08^2)$, the next 3 subjects are from $N(-0.01, 0.14^2)$, and the last 2 subjects come from $N(-0.01, 0.25^2)$. Again, we generate the corresponding covariance matrices by $\mathbf{\Sigma}^{(s)} = {A^{(s)}}^T A^{(s)}; s = 1, \cdots, 15$ to make those positive definite. The number of subjects and different covariance structures in each cluster are determined based on typical task-related fMRI data. Figure 4.10 demonstrates the covariance heat maps from each cluster for the multiple clusters setting. In Figure 4.10, Subject 1 in Cluster 1 shows light-red colored tables, Subject 12 in Cluster 2 displays mixed light- and medium-red colored tables, and Subject 15 in Cluster 3 presents dark-red colored tables. Although the number of subjects in Cluster 1 is larger than that in Clusters 2 or 3, we cannot specify any outliers in this case.

Table 4.7 describes the vectorized Euclidean distances from the median to the estimated inverse covariance for each subject, the means, and standard deviations (in parentheses) of the estimated weights by JGL with RA over 100 repetitions for the multiple clusters setting. The distances and weights can be clearly distinguished among the three clusters. The distances for each cluster range from 2.0462 to 2.2047, from 8.0783 to 8.2826, and from 15.7622 to 15.7625, respectively. Correspondingly, the assigned weights for the first cluster range between 0.0831 and 0.0847, those of the second cluster range between 0.0443 and 0.0449, and those of the last cluster range between 0.0134 and 0.0137. Likewise, the result indicates that our method assigns smaller weights to the subjects whose estimated inverse covariance is much deviated from the median estimated inverse covariance of the whole group. Moreover, the proposed method classifies the three clusters by allocating different weights for each cluster, but assigns similar weights within each cluster.

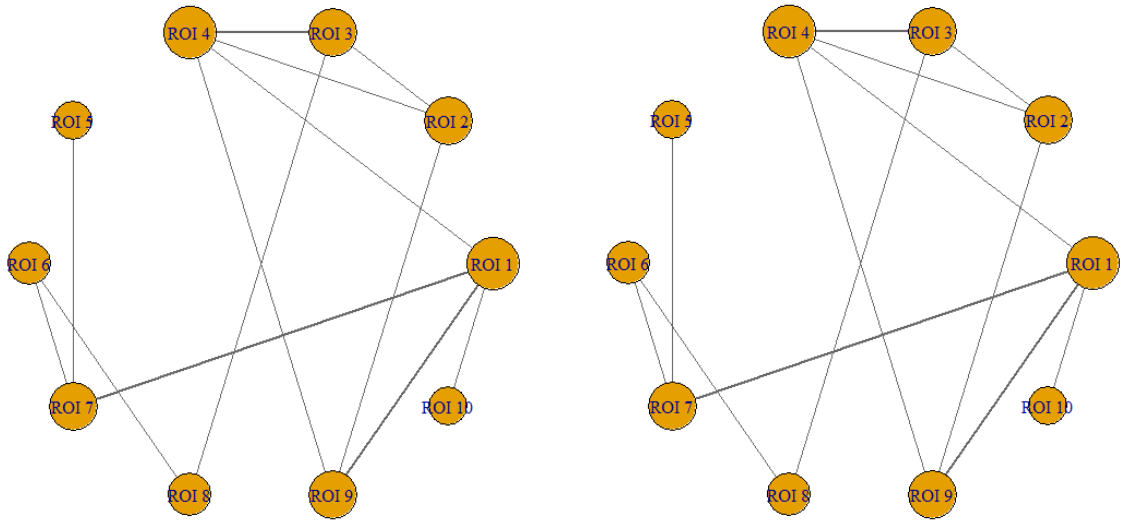
Figure 4.11 compares the group graphs constructed by JGL with RA and JGL from one example picked from 100 repetitions for the multiple clusters setting. The two group graphs in Figures 4.11(a) and 4.11(b) show clearly different shapes in terms of the number of edges and the strength of associations between nodes. Compare to the group graph generated by

JGL with RA in Figure 4.11(a), the group graph by JGL in Figure 4.11(b) is more dense and has stronger relationships. This result is due to the effect of the subjects in Clusters 2 and 3. Since the estimated inverse covariance matrices from those subjects are relatively more dense than those from the subjects in Cluster 1, Figure 4.11(b) has a dense shape, making it difficult to interpret the relationships. However, our method in Figure 4.11(a) provides a reasonable outcome by mitigating effect of the subjects in Clusters 2 and 3.

In summary, the proposed JGL with RA method compares the distances between the median and each subjects' estimated inverse covariance, and assigns different weights based on the distances. It results in allocating larger weights to the majority cluster and smaller weights to the minority or outlying cluster, and enables us to identify a different effect of each subject in group-level estimation. Moreover, our method assigns similar weights to the subjects within each cluster if the subjects are homogeneous. We further investigate the properties of the proposed methods with the saccade tasks fMRI data in Chapter 5.

Table 4.1: The estimated weights by JEMP with RA for homogeneous subjects. The means (and standard deviations) are calculated from 100 repetitions.

Subjects	1	2	3	4	5	6	7	8
w	0.0666	0.0666	0.0663	0.0664	0.0668	0.0666	0.0661	0.0669
(s.d.)	(0.0016)	(0.0015)	(0.0017)	(0.0040)	(0.0009)	(0.0014)	(0.0035)	(0.0017)
Subjects	9	10	11	12	13	14	15	-
w	0.0667	0.0668	0.0669	0.0668	0.0670	0.0668	0.0668	-
(s.d.)	(0.0020)	(0.0023)	(0.0014)	(0.0017)	(0.0014)	(0.0019)	(0.0015)	-



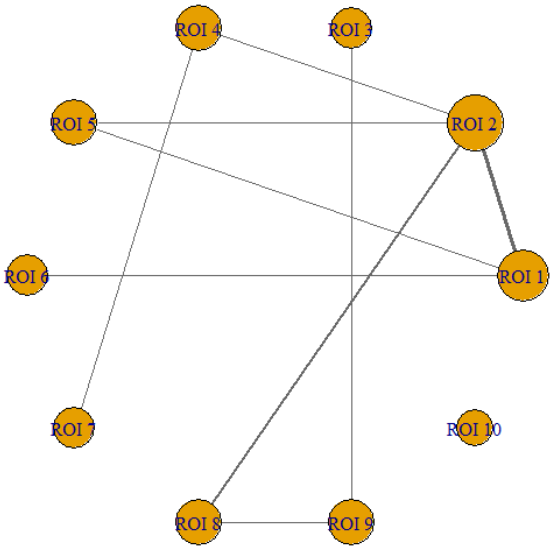
(a) JEMP with RA

(b) JEMP

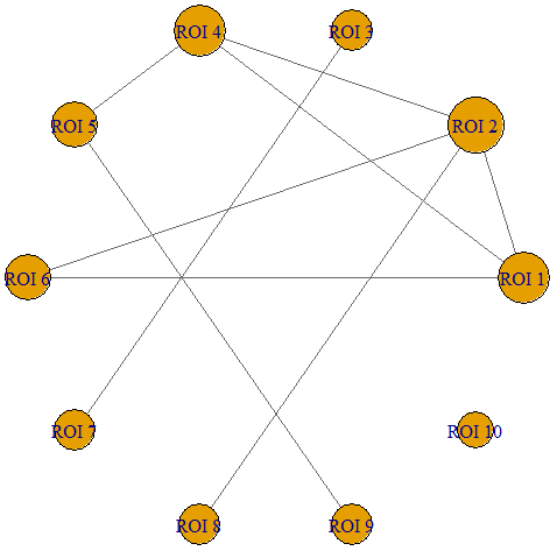
Figure 4.1: Group graphs by JEMP with RA and JEMP from homogeneous subjects. One example is picked from the 100 repetitions. The edges weaker than 0.3 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

Table 4.2: The estimated weights by JEMP with RA for the three outlying subjects setting. The means (and standard deviations) are from 100 repetitions.

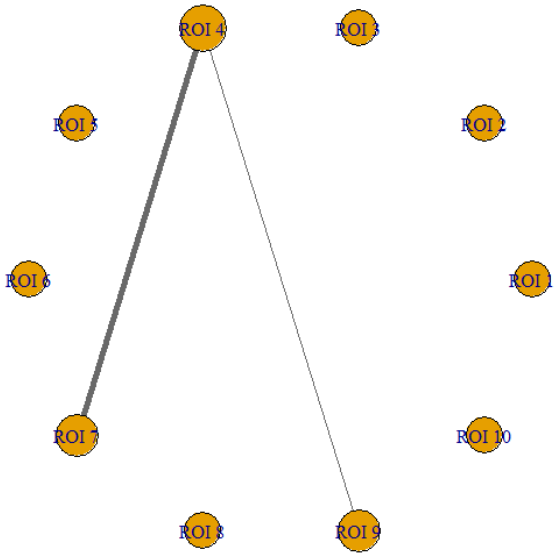
Subjects	1	2	3	4	5	6	7	8
w	0.0823	0.0821	0.0820	0.0833	0.0823	0.0822	0.0833	0.0827
(s.d.)	(0.0052)	(0.0047)	(0.0054)	(0.0044)	(0.0059)	(0.0056)	(0.0040)	(0.0065)
Subjects	9	10	11	12	**13**	**14**	**15**	-
w	0.0828	0.0818	0.0811	0.0831	**0.0042**	**0.0035**	**0.0034**	-
(s.d.)	(0.0051)	(0.0081)	(0.0090)	(0.0044)	**(0.0050)**	**(0.0047)**	**(0.0039)**	-



(a) JEMP with RA



(b) JEMP without outliers

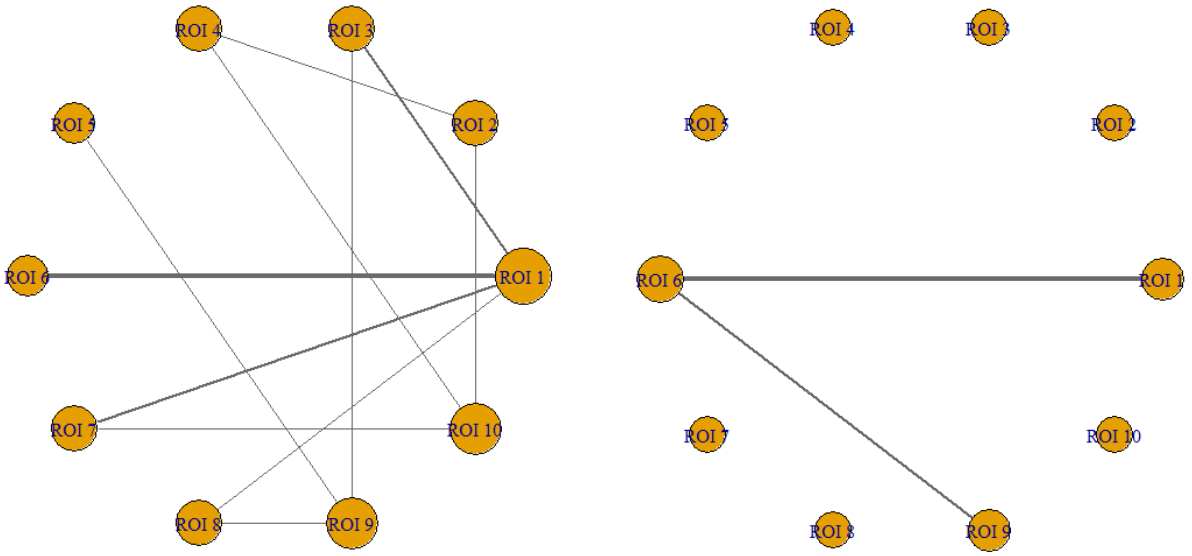


(c) JEMP

Figure 4.2: Group graphs by JEMP with RA and JEMP from the three outlying subjects setting. One example is picked from the 100 repetitions. The edges weaker than 0.3 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

Table 4.3: The estimated weights by JEMP with RA for the multiple clusters setting. The means (and standard deviations) are from 100 repetitions.

Subjects	1	2	3	4	5	6	7	8
w	0.0900	0.0896	0.0905	0.0903	0.0898	0.0902	0.0902	0.0906
(s.d.)	(0.0077)	(0.0059)	(0.0074)	(0.0070)	(0.0069)	(0.0065)	(0.0074)	(0.0063)
Subjects	*9*	*10*	*11*	*12*	**13**	**14**	**15**	-
w	*0.0657*	*0.0647*	*0.0650*	*0.0661*	**0.0059**	**0.0068**	**0.0047**	-
(s.d.)	*(0.0140)*	*(0.0153)*	*(0.0151)*	*(0.0140)*	**(0.0064)**	**(0.0074)**	**(0.0055)**	-



(a) JEMP with RA

(b) JEMP

Figure 4.3: Group graphs by JEMP with RA and JEMP from the multiple clusters setting. One example is picked from the 100 repetitions. The edges weaker than 0.3 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

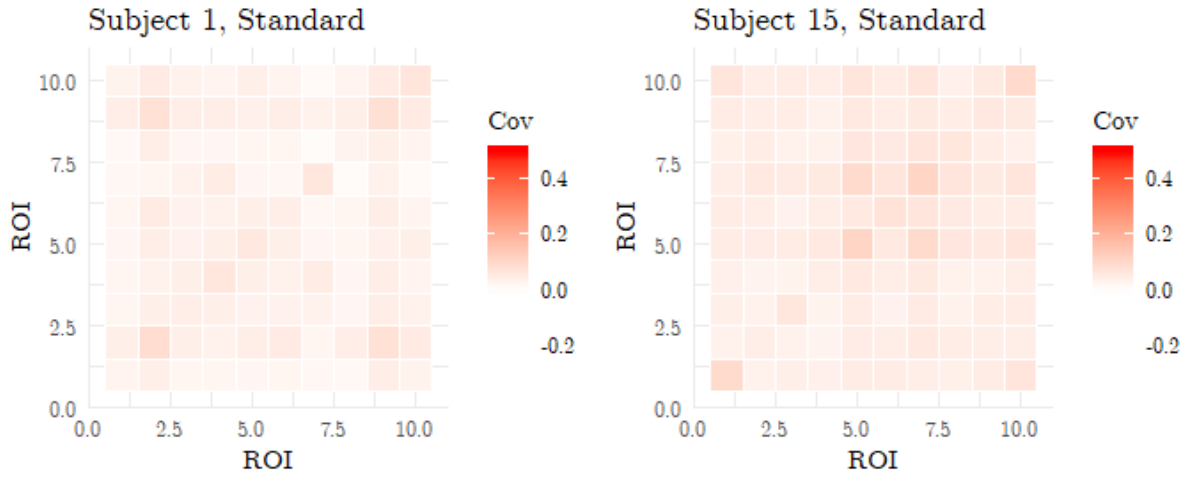
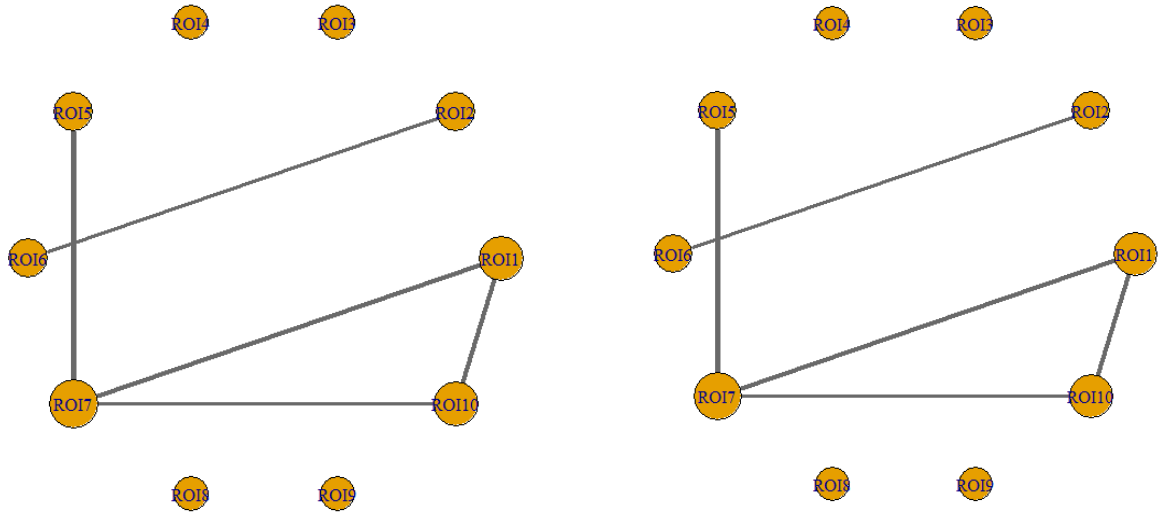


Figure 4.4: Covariance heat maps from homogeneous subjects. Covariance matrices from Subjects 1 and 15 in the one example from the 100 repetitions are provided. The horizontal and vertical axes are ROI indices.

Table 4.4: Distances from the median of the estimated inverse covariance and the estimated weights by JGL with RA for homogeneous subjects. Vectorized Euclidean distances between the median and each subject's estimated inverse covariance are provided with the estimated weights. The means (and standard deviations) are calculated from 100 repetitions.

Subjects	1	2	3	4	5	6	7	8
Distance	1.8161	1.8085	1.7422	1.8602	1.8172	1.7758	1.8172	1.8485
w	0.0668	0.0668	0.0668	0.0668	0.0667	0.0668	0.0669	0.0668
(s.d.)	(4e-04)	(4e-04)	(8e-04)	(4e-04)	(5e-04)	(1e-03)	(4e-04)	(6e-04)
Subjects	9	10	11	12	13	14	15	-
Distance	1.8314	1.8228	1.8856	1.7846	1.8010	1.8010	1.8683	-
w	0.0668	0.0668	0.0668	0.0668	0.0669	0.0669	0.0668	-
(s.d.)	(5e-04)	(5e-04)	(6e-04)	(4e-04)	(8e-04)	(4e-04)	(5e-04)	-



(a) JGL with RA (b) JGL

Figure 4.5: Group graphs by JGL with RA and JGL from homogeneous subjects. One example is picked from the 100 repetitions. The edges weaker than 0.05 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

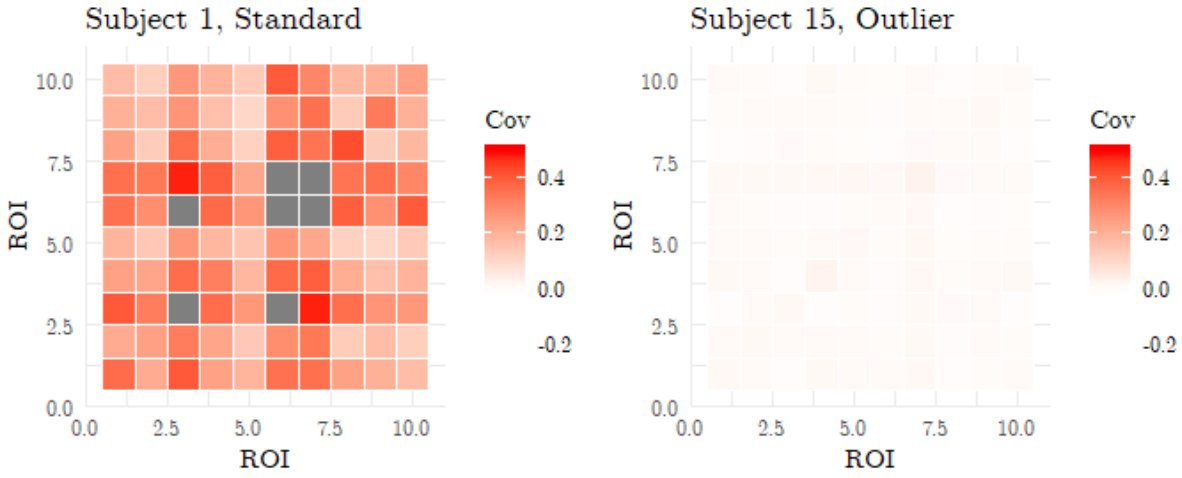
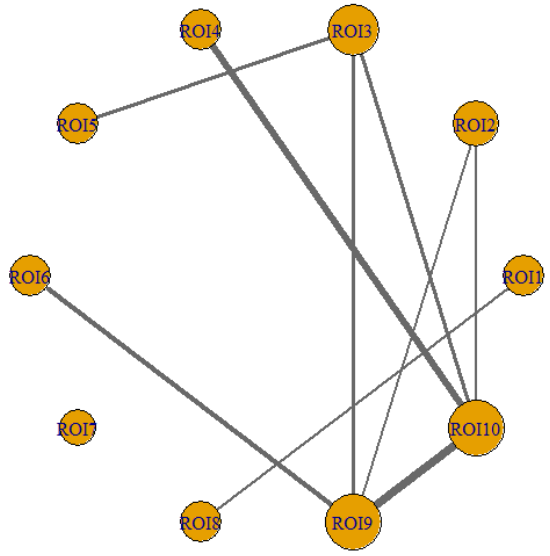


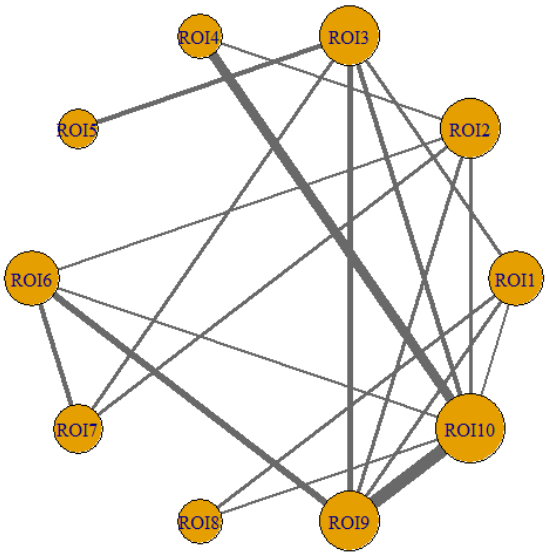
Figure 4.6: Covariance heat maps from the small covariance outliers setting. Covariance matrices from Subjects 1 and 15 in the one example from the 100 repetitions are provided. The horizontal and vertical axes are ROI indices.

Table 4.5: Distances from the median of the estimated inverse covariance and the estimated weights by JGL with RA from the small covariance outliers setting. Vectorized Euclidean distances between the median and each subject's estimated inverse covariance are provided with the estimated weights. The means (and standard deviations) are calculated from 100 repetitions.

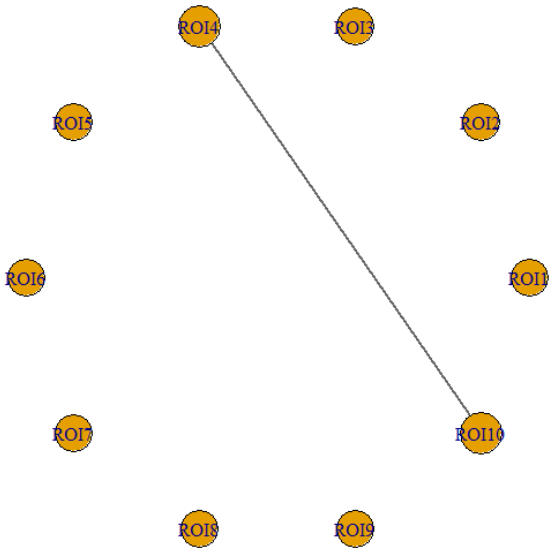
Subjects	1	2	3	4	5	6	7	8
Distance	3.6409	3.5368	3.6363	3.5034	3.5017	3.4567	3.6229	3.5817
w	0.0758	0.0781	0.0768	0.0782	0.0773	0.0770	0.0769	0.0765
(s.d.)	(0.0091)	(0.0078)	(0.0100)	(0.0088)	(0.0098)	(0.0092)	(0.0100)	(0.0101)
Subjects	9	10	11	12	**13**	**14**	**15**	-
Distance	3.5324	3.5585	3.6372	3.4931	**10.3983**	**10.3933**	**10.3835**	-
w	0.0780	0.0761	0.0763	0.0767	**0.0254**	**0.0254**	**0.0254**	-
(s.d.)	(0.0092)	(0.0113)	(0.0112)	(0.0094)	**(0.0057)**	**(0.0056)**	**(0.0057)**	-



(a) JGL with RA



(b) JGL without outliers



(c) JGL

Figure 4.7: Group graphs by JGL with RA and JGL from the small covariance outliers setting. One example is picked from the 100 repetitions. The edges weaker than 0.05 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

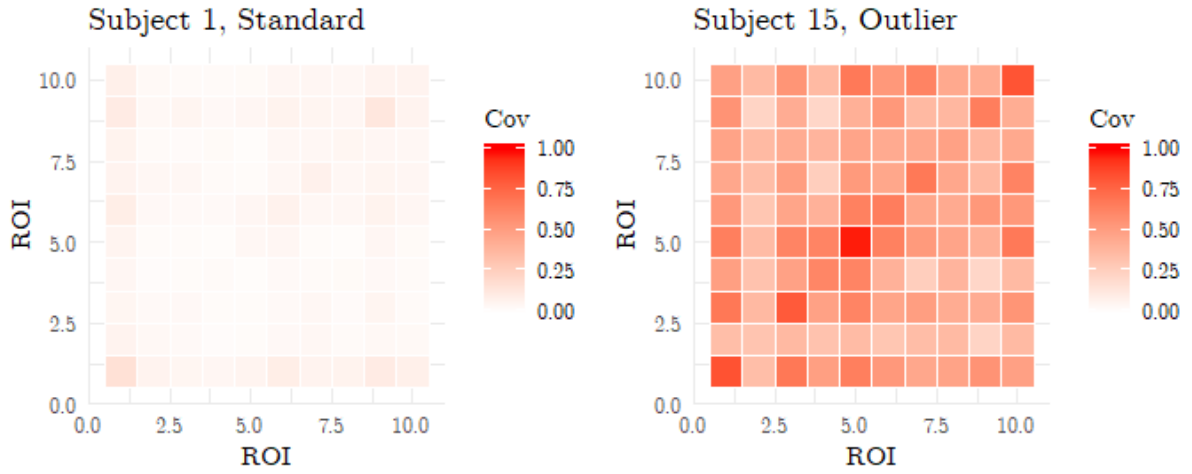
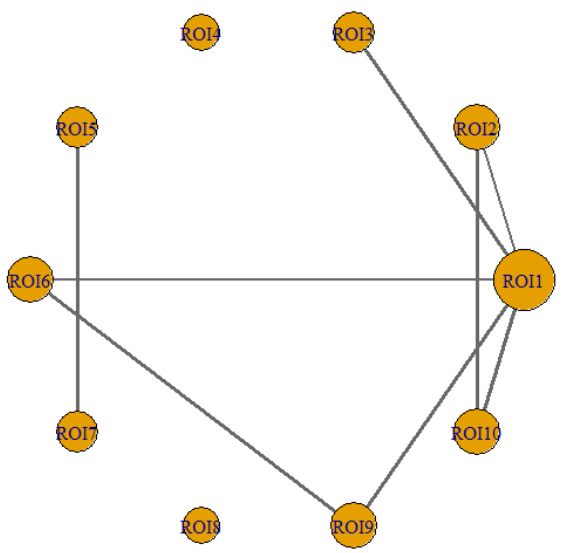


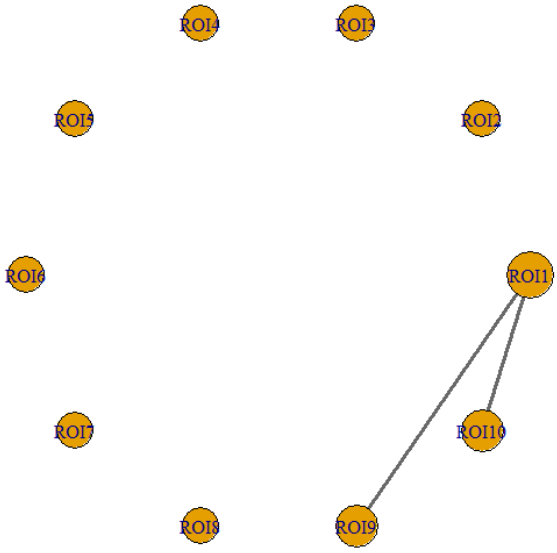
Figure 4.8: Covariance heat maps from the large covariance outliers setting. Covariance matrices from Subjects 1 and 15 in the one example from the 100 repetitions are provided. The horizontal and vertical axes are ROI indices.

Table 4.6: Distances from the median of the estimated inverse covariance and the estimated weights by JGL with RA from the large covariance outliers setting. Vectorized Euclidean distances between the median and each subject's estimated inverse covariance are provided with the estimated weights. The means (and standard deviations) are calculated from 100 repetitions.

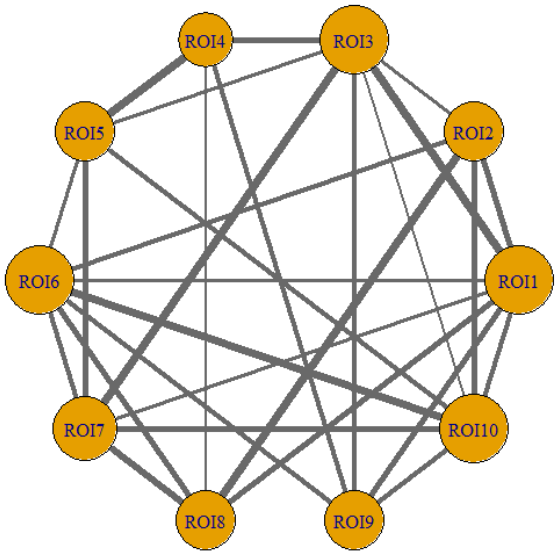
Subjects	1	2	3	4	5	6	7	8
Distance	2.0688	2.1609	2.0477	2.1646	2.1759	2.1554	2.0462	2.0590
w	0.0795	0.0793	0.0798	0.0792	0.0796	0.0794	0.0799	0.0807
(s.d.)	(0.0042)	(0.0045)	(0.0045)	(0.0050)	(0.0055)	(0.0046)	(0.0047)	(0.0039)
Subjects	9	10	11	12	**13**	**14**	**15**	-
Distance	2.1179	2.2034	2.0570	2.2047	**15.7758**	**15.7491**	**15.7680**	-
w	0.0802	0.0798	0.0797	0.0794	**0.0144**	**0.0146**	**0.0146**	-
(s.d.)	(0.0041)	(0.0042)	(0.0043)	(0.0052)	**(0.0053)**	**(0.0047)**	**(0.0047)**	-



(a) JGL with RA



(b) JGL without outliers



(c) JGL

Figure 4.9: Group graphs by JGL with RA and JGL from the large covariance outliers setting. One example is picked from the 100 repetitions. The edges weaker than 0.05 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

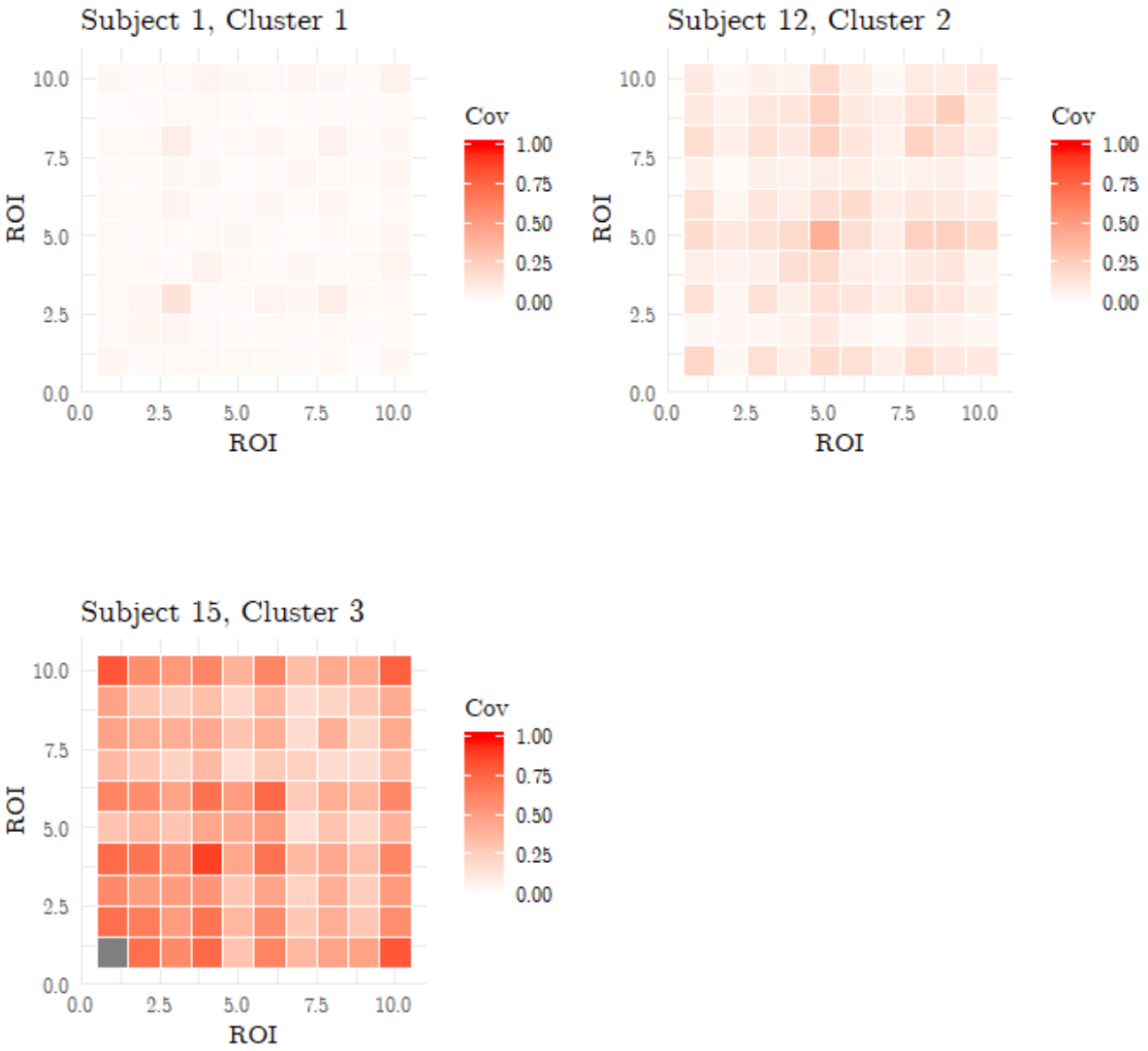
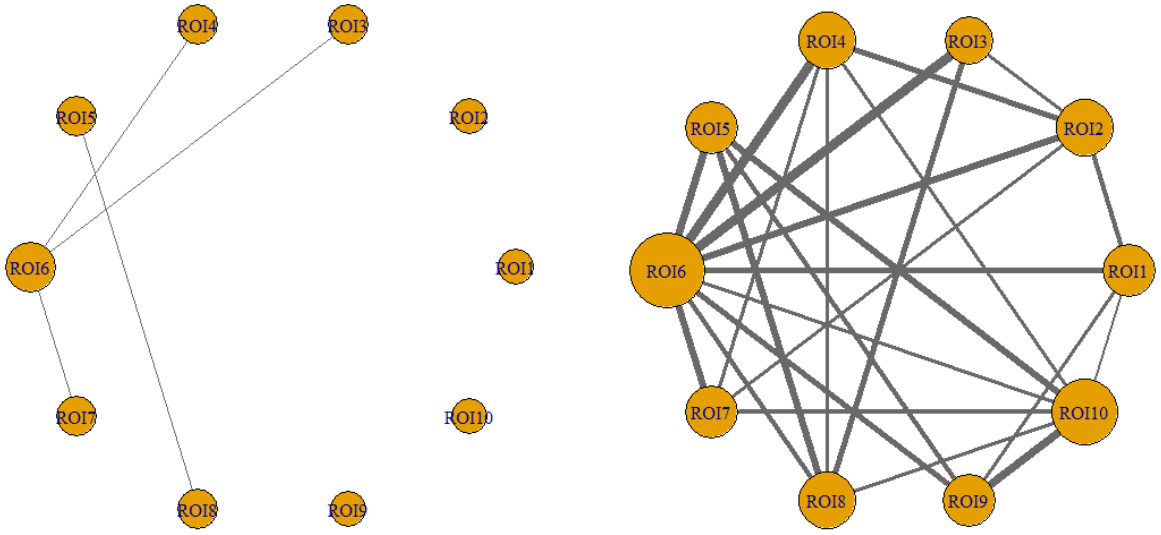


Figure 4.10: Covariance heat maps from the multiple clusters setting. Covariance matrices from Subjects 1, 12 and 15 in the one example from the 100 repetitions are provided. The horizontal and vertical axes are ROI indices.

Table 4.7: Distances from the median of the estimated inverse covariance and the estimated weights by JGL with RA from the multiple clusters setting. Vectorized Euclidean distances between the median and each subject's estimated inverse covariance are provided with the estimated weights. The means (and standard deviations) are calculated from 100 repetitions.

Subjects	1	2	3	4	5	6	7	8
Distance	2.6303	2.5194	2.5111	2.5204	2.4932	2.5469	2.5844	2.4412
w	0.0843	0.0833	0.0834	0.0831	0.0847	0.0844	0.0844	0.0838
(s.d.)	(0.0060)	(0.0043)	(0.0055)	(0.0052)	(0.0044)	(0.0052)	(0.0054)	(0.0057)
Subjects	9	10	*11*	*12*	*13*	**14**	**15**	-
Distance	2.4575	2.5454	*8.2826*	*8.2104*	*8.0783*	**15.7625**	**15.7622**	-
w	0.0836	0.0842	*0.0443*	*0.0444*	*0.0449*	**0.0137**	**0.0134**	-
(s.d.)	(0.0052)	(0.0054)	*(0.0079)*	*(0.0084)*	*(0.0079)*	**(0.0057)**	**(0.0058)**	-



(a) JGL with RA

(b) JGL

Figure 4.11: Group graphs by JGL with RA and JGL from the multiple clusters setting. One example is picked from the 100 repetitions. The edges weaker than 0.05 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

Chapter 5

Real Data Analysis

5.1 FMRI Data Description

We apply our JEMP with RA and JGL with RA methods to the saccade tasks fMRI data. The data set is composed of 32 right-handed healthy participants who were recruited from the University of Georgia Psychology Department online research pool, did not experience major psychiatric disorders or substance abuse, and had normal or corrected-to-normal vision. The detailed description of the data set can be found in Pierce and McDowell (2017).

Pierce and McDowell (2017) considered two types of the saccade tasks, prosaccade (PS) and antisaccade (AS). These tasks require rapid eye movement responses that require support of visual and motor neural circuitry. A PS task asked individuals to look toward a given stimulus. An AS task asked them to look to the mirror image (opposite direction) of the stimulus. The participants had five event-related saccade experiments with different proportions of AS tasks, 0% 25%, 50%, 75%, and 100%, respectively. Among 60 saccade tasks in each run, for example, the 25% run consisted of 15 AS tasks and 45 PS tasks. Additionally, there were block runs where each participant was given 8 blocks of saccade tasks. The experiments were composed of 4 AS and 4 PS blocks where each block contained 10 corresponding saccade tasks. The fMRI data were acquired by a 3-T GE Signa Excite

HDx system (General Electric Medical Systems, Milwaukee, WI) at the University of Georgia Bio-Imaging Research Center. The participants could identify the stimuli displayed by a projection screen at his or her feet using the Presentation software (Neuro-behavioral Systems, Albany, CA). The IView X MRI-LR system (Senso-Motoric Instruments, Teltow, Germany) was used to sample and record right eye pupil position at 60Hz. As a baseline, each participant was scanned using a high-resolution T1-weighted anatomical image (echo time = 3 msec, flip angle = 20°, field of view = 240 mm × 240 mm, matrix size = 256 × 256, 150 axial slices, in-slice resolution = 0.94 × 0.94 mm, slice thickness = 1.2 mm, scan time = 6 min 32 sec), followed by the functional T2*-weighted scans (echo time = 30 msec, repetition time = 2000 msec, flip angle = 90°, field of view = 220 mm × 220 mm, matrix size = 64 × 64, 33 interleaved oblique slices aligned with the AC–PC plane, in-slice resolution = 3.4 × 3.4 mm, slice thickness = 4 mm, slice gap = 0 mm, 4 dummy volumes for magnet stabilization, 158 volumes, scan time = 5 min 24 sec).

After the initial (pre-test) scan session, the participants practiced PS and AS tasks for four days. They came back after the practice for a post-test fMRI session with the same scan order as the pre-test session. The collected data underwent general preprocessing steps including slice-timing adjustment, motion correction, spatial standardization, spatial smoothing, and voxel-wise scaling. Additionally, a custom brain mask was created and compared with each individual's brain, then the voxels that fell outside of the masked brain were removed from the data in order to analyze the same brain regions within the group. To sum up, we had the data of 32 participants in two sessions (pre-test and post-test) for 6 runs (5 probabilities and 1 block). Then, we used ROIs (nodes) to construct a brain network. We considered 11 ROIs that are known to support saccadic performance, including the prefrontal cortex (PFC), the supplementary eye field (SEF), the medial frontal eye fields (mFEF), the lateral frontal eye fields (lFEF), the inferior frontal cortex (IFC), the thalamus (Thal), the basal ganglia (BG), the inferior parietal lobule (IPL), the precuneus (Precun), the cuneus (Cun), and the middle

occipital gyrus (MOG) (Pierce and McDowell, 2017). We averaged the voxels within each ROI for each subject. Hence, we eventually had 11 ROIs with 128 time points for any single subject, after the wavelet transformation as mentioned in Section 3.3.

5.2 JEMP with RA

We apply JEMP with RA to each scan session within each run. Similar to the simulation study in Section 4.1, we estimate weights for subject, produce group graphs, and compare the graphs with those provided by JEMP. We use LL^* in (3.6) to select the tuning parameters.

5.2.1 Event-related Experiments

First, we perform the analysis on the event-related experiments to investigate outlying effect. Noting that the group is composed of 32 subjects, we check whose estimated weight is smaller than $1/32 = 0.0313$ to identify possible outliers.

Figure 5.1 shows the weights estimated by JEMP with RA for all of the 5 probability (0%, 100%, 25%, 50%, and 75%) runs. In the figure, the pre- and post-test sessions of the event experiments are presented in the first and second column, respectively. In each panel, we use different colors, line types, and marks to distinguish the 5 different runs. For the 0% run in the pre-test session (red solid line) in Figure 5.1(a), we observe that Subjects 8, 25 and 27 have relatively smaller weights (0.0201, 0.0098, and 0.0197, respectively) than the others. In the post-test session for the 0% run in Figure 5.1(b), Subject 25 and 27 still show small weights (0.0216 and 0.0175, respectively) while Subject 8 recovers the weight similar to the others (0.0357). From the results, Subject 25 and 27 can be viewed as potential outliers for the 0% run. We also observe that Subject 7, 15, 20, and 21 have small weights (0.0153, 0.0115, 0.0129, and 0.0191, respectively). Such outliers can influence on the common structure of the entire group.

The estimated weights for the 100% run are presented (olive short-dashed lines) in Figures 5.1(a) and 5.1(b). In the two sessions, we note that Subject 27 again displays small weights (0.0091 for pre-test and 0.0043 for post-test, respectively) than the others. Furthermore, the weights in Figure 5.1(a) indicate that there are multiple clusters in the pre-test session, some subjects (1, 5, 6, 12, 13, 19, and 23) show large weights and some (3, 8, 11, 20, 25, 27, 30, and 32) show small weights. Although the differences among the weights are not large, we can detect multiple clusters in the pre-test session. In contrast, the post-test session result in Figure 5.1(b) shows more evenly distributed weights among the subjects.

The estimated weights by JEMP with RA for the 25% run in the pre- and post-test sessions are presented (green dashed lines) in Figure 5.1(a) and 5.1(b), respectively. Similar to the 0% run, Subject 27 in pre-test session presents weight of 0.0126, which is relatively smaller than the others in the group. Although the possible outlying subject does not have extremely small weight close to 0, there is clear difference in the weights from the other subjects. On the contrary, Subject 27 in the post-test session shows weight of 0.0259, which is the minimum among other subjects in the group, but not exceptionally different from the others. This implies that there is no extreme outlier in the post-test session for this run.

In the pre-test session for the 50% run (blue dashed line), we notice that Subjects 8, 22, and 27 display small weights (0.0231, 0.0239, and 0.0199, respectively). Since the other subjects show similar weights, we can consider these three subjects as outliers. On the other hand, we observe that all of the subjects in the group show the similar weight distribution in the post-test session except Subject 27 (0.0192).

The purple dotted lines in Figures 5.1(a) and 5.1(b) show the estimated weights for the 75% run in the pre- and post-test sessions, respectively. In the pre-test session, we note that a small weight (0.0097) is assigned to Subject 27. We also find small weights cluster (Subjects 9, 11, 20, and 21, the minimum weights=0.0188 from Subject 9) and large weights cluster (Subjects 1, 2, 12, 13, and 19, the maximum weights=0.0401 from Subject

19). Similarly in the post-test session, the weight of Subject 27 is still lower compared to the other weights (0.0050). Moreover, we note that Subjects 8, 9, and 29 show relatively small weights (the minimum weight = 0.0144 from Subject 9), while Subjects 1, 10, 12, 13, and 19 have considerably large weights (the maximum weight = 0.0391 from Subject 10). From this observation, we conclude that there are possibly three clusters for this particular case. In summary, the proposed method detects possible outlying subjects across different sessions and runs. For example, Subject 27 is consistently categorized as an outlier throughout the runs. In what follows, we investigate their impacts by comparing the group graphs created by JEMP with RA and JEMP.

We create group graphs to visualize the associations between ROIs, estimated by JEMP with RA and JEMP for different sessions and runs. We also summarize any changes from pre-test to post-test in the relationships between ROIs by the proposed method. In Table 5.1, we summarize the changes in associations after the practice for the 0% run. We only report the edges ($\hat{\omega}_{ij}$) whose absolute differences between the two sessions are greater than 0.5. It is noticeable that Cun has some stronger edges, and lFEF has some weaker edges after the practice. The similarity in sparsity level after the practice can be observed in Table 5.1 in the sense that there are similar number of stronger and weaker associations.

Figure 5.2 illustrates the group graphs estimated by JEMP with RA and JEMP for the 0% run. We note that all of the graphs show very strong associations between the Thal-BG, Thal-Cun, lFEF-SEF, lFEF-mFEF and MOG-Cun regions. Generally, the JEMP with RA group graphs capture more relationships between ROIs than the JEMP group graphs as also observed in the simulation study. For example, the pre-test JEMP with RA group graph in Figure 5.2(a) displays IFC-MOG, lFEF-Cun, mFEF-Thal, and mFEF-Precun, but the one by JEMP in Figure 5.2(b) shows no edges among those relationships. Because we detect some outliers in the pre-test in Figure 5.1(a) and in the post-test in Figure 5.1(b) for the 0%

run, the group graphs in both sessions from the two methods show some differences in terms of the number and the strength of associations.

Table 5.2 summarizes the changes of associations for the 100% run. The result shows the similar number of associations that become weaker and stronger after the practice. Specifically, we observe that lFEF-Cun association becomes stronger, and Thal-BG and Precun-Cun relationships become weaker.

Corresponding to Table 5.2, we illustrate the group graphs for the 100% run in Figure 5.3. In the pre-test session in Figures 5.3(a) and 5.3(b), JEMP with RA tends to capture slightly more associations than JEMP. On the other hand, the two group graphs in the post-test session in Figures 5.3(c) and 5.3(d) have an identical shape owing to the uniformly distributed weights in Figure 5.1(b). Also, the strength of Thal-BG relationship in the group graph from the post-test session in Figure 5.3(c) is weaker than that from the pre-test session in Figure 5.3(a) as observed in Table 5.2.

Table 5.3 summarizes the changes in associations for the 25% run. In contrast to Table 5.1, we observe weaker relationships than stronger ones. In particular, IPL becomes more weakly connected after the practice, which is different from the 0% run.

Figure 5.4 shows the group graphs estimated by JEMP with RA and JEMP for the 25% run. All of the graphs show strong associations between Thal-BG, lFEF-mFEF, lFEF-SEF, mFEF-SEF, and MOG-Cun. Similar to the result from 0% runs, the JEMP with RA group graphs capture more associations than those generated by the JEMP in the pre-test session. Besides, we notice that the group graphs in the post-test session in Figures 5.4(c) and 5.4(d) are identical. This is consistent with the evenly distributed weights in Figure 5.1(b). We also observe that PFC-Precun, SEF-Cun, IFC-IPL, and Thal-BG relationships become weaker or even disappear after the practice as indicated in Table 5.3.

In Table 5.4, we find that the SEF-mFEF and mFEF-lFEF associations get stronger, while the PFC-Precun and SEF-Precun relationships get weaker. Precun shows more associations that become weaker than those that get stronger after the practice.

Figure 5.5 demonstrates the group graphs estimated by JEMP with RA and JEMP for the 50% run. In the top panel, we observe that the pre-test group graphs display different relationships. For example, the IFC is connected with SEF, the IPL is associated with Thal, and the IFC and lFEF are not linked in Figure 5.5(c). The different structures of the two graphs stem from the weight differences shown in Figure 5.1(a). In contrast, the group graphs in post-test session in the bottom panel are almost identical. This is consistent with the similar weights shown in Figure 5.1(b).

Table 5.5 describes the changes in association for the 75% run. We find that more regions become weakly connected such as mFEF-lFEF, IFC-IPL, Thal-BG, and Precun-Cun after the practice. In particular, PFC and SEF show four more associations that get weaker after the practice while there are two and three stronger ones, respectively.

Lastly, Figure 5.6 displays the JEMP with RA and JEMP group graphs for the 75% run. Since the estimated weights in the pre-test session in Figure 5.1(a) suggest that there are multiple clusters, the two graphs in Figures 5.6(a) and 5.6(b) display different structures. In specific, we observe that IPL is linked with Thal, IFC, and lFEF in the JEMP with RA group graph while the JEMP group graph does not provide any of those associations. Likewise, Figures 5.6(c) and 5.6(d) show the different edge appearance and thickness, due to the clustered weight distribution in the post-test session in Figure 5.1(b). We can see that JEMP produces more sparse graphs than those from JEMP with RA.

5.2.2 Block Runs

Additionally, we analyze the block runs data using the proposed JEMP with RA and JEMP (Lee and Liu, 2015). Similar to the event experiments, we check the estimated weights to identify potential outliers, and provide the group graphs.

We report the estimated weights from JEMP with RA for the pre- and post-test sessions in Figures 5.7(a) and 5.7(b), respectively. If all of the subjects are homogeneous, each subject's weight should be assigned as $1/32 = 0.0313$. In Figure 5.7(a), most subjects have the weights slightly greater than 0.0313 because of the small weight of Subject 27 (0.0086). In this sense, Subject 27 can be considered as an outlier in the pre-test session. For the post-test session, we also observe that Subject 27 has a small estimated weight (0.0138) than the other subjects. Even though the difference is not as large as in the pre-test session, it can affect aggregation of individual precision matrices. Furthermore, Subjects 2, 7, 15, 21, and 25 present relatively small weights (the minimum weights=0.0217 from Subject 25). We notice that Subject 27 in the block runs corresponds to Subject 27 in the event-related experiments, who was consistently identified as a potential outlier in the previous analyses.

Table 5.6 summarizes the changes in association after the practice for the block runs. In the table, we observe that more associations become weaker than stronger after the practice. Especially, the associations with Cun get weaker after the practice, contrast to the event-related experiments.

We also illustrate the group graphs generated by JEMP with RA and JEMP for the block runs in the pre- and post-test sessions in Figure 5.8. It can be seen that the differences between the two graphs are marginal. Also, the strong associations such as Thal-BG, Thal-SEF, lFEF-mFEF, SEF-Precun, and MOG-Cun are captured in both graphs for both sessions. We also observe that the post-test group graph is more sparse than the pre-test as can be seen in Table 5.6, but some of the associations such as PFC-lFEF become more solid after the practice.

In fact, most of the runs in the event-related experiments and block runs can be considered as multiple clusters. Consistent with our simulation setting in Section 4.1.3, we show that the suggested JEMP with RA is more robust than JEMP, especially to the subjects who tend to have more sparse graphs than the standard subjects. JEMP is affected by such different clusters and results in presenting more sparse graphs than the proposed method throughout the experiments.

To sum up, we compare group graphs generated by the proposed method and JEMP based on graph kernels. A graph kernel quantifies the similarity between graphs using kernel methods (Vishwanathan et al., 2010). It measures the similarity of pairs of graphs, the more similar graph pairs we have, the higher value is provided. Among various kernel families, we employ the Weisfeiler-Lehman (W-L) graph kernel (Shervashidze et al., 2011) which shows good performance in terms of the effectiveness and the efficiency (Sugiyama et al., 2017). For the computation of the graph kernel, we use `R` package `graphkernels` (Sugiyama, 2018). In Figure 5.9, we demonstrate the graph kernel values for each test session and run. For example, "pre0" indicates the 0% run of the pre-test session. We observe that "post25" and "post50" graphs yield the highest value of 66, while "pre0" and "post0" graphs present low values of 14 and 13, respectively. The high value of the W-L graph kernel means that a graph pair is more similar to each other, so it implies that the JEMP with RA assigns uniform weights to the subjects and provides the group graph that has a similar shape to that from JEMP. On the other hand, the small W-L graph kernel value is computed when a graph pair shows a clear difference. If there are some outliers in the group, the estimated weights by our method are diverged and we can observe a large difference between the group graphs by the proposed method and JEMP. The result is consistent with the estimated weights in Figure 5.1 in the sense that 0% run in both sessions show some outliers, whereas 25% and 50% run in the post-test session display almost uniformly distributed weights. It illustrates

the group graph can be highly affected by outliers, which results in distinctive group graphs from our method and JEMP.

We also compare the pre- and post-test session group graphs from our method based on W-L graph kernels. Figure 5.10 shows the graph kernel values for each run. We observe that the 100% run in the event-related experiments shows the lowest score of 11, while the block runs show the highest score of 14. Although the difference is marginal, all of the runs show relatively small W-L graph kernels. From the result, we can infer that there is a practice effect which induces changes of brain networks, especially for the 100% run.

5.3 JGL with RA

In this section, we apply the proposed JGL with RA method to the saccade task fMRI data. The estimated weights for subjects, the vectorized Euclidean distances from the median to the estimated inverse covariance for each subject, and group graphs are provided for each scan session within each run. We also add the graphs generated by JGL for comparison. For the tuning parameters for each run, we use LL^* suggested in (3.6) with the range of a grid that provides sufficient sparsity.

5.3.1 Event-related Experiments

First, we evaluate the performance of our method on the 32 subjects who conducted event-related experiments. If the subjects in each run are homogeneous, the weight for each subject will be estimated around $1/32 = 0.0313$. Therefore, we consider the subjects who have assigned weights extremely smaller than 0.0313 as outlying subjects. We also note that if several subjects in the group have such abnormal weights, the group can be considered as multiple clusters.

Figure 5.11 describes the estimated weights and distances by JGL with RA from the two sessions (pre- and post-test) for the 5 probability runs (0%, 100%, 25%, 50%, and 75%) in each panel. In the figure, we present the estimated weights and distances respectively in the first and second row, and the pre- and post-test sessions of the event-related experiments in different columns. In each panel, different colors, line types, and marks are used to show the outcomes from the 5 different runs at once. The red solid lines in Figure 5.11 represent the estimated weights and distances by JGL with RA for the 0% run in the pre- and post-test sessions, respectively. In the pre-test session in Figures 5.11(a) and 5.11(c), we observe that Subjects 13 and 19 have apparently smaller weights (0.0115 and 0.0128, respectively) and longer distances (9.3624 and 9.1507, respectively) than the other subjects. Additionally, Subjects 22, 26, 28, and 31 have low assigned weights (0.0201, 0.0207, 0.0175, and 0.0222, respectively) with long distances (6.0249, 5.6708, 5.8846, and 4.3572, respectively). Similar to the simulation study demonstrated in Section 4.2, the result indicates an inverse relationship between the estimated weights and distances. We also note that it is hard to confirm that the subjects with small estimated weights as outliers in this group, since they are not a few, and the group also consists some subjects with medium and large weights such as Subjects 8 and 11 (0.0402 and 0.0401, respectively). Therefore, it is more reasonable to consider the group as multiple clusters according to JGL with RA. In the post-test session in Figures 5.11(b) and 5.11(d), although relatively smaller weights (0.0248 and 0.0257, respectively) and longer distances (6.1241 and 5.5413, respectively) are assigned to Subjects 12 and 31, we observe that the differences among the subjects become smaller than the pre-test session.

For the 100% run in the pre-test session (olive short-dashed lines) in Figures 5.11(a) and 5.11(c), similar to the 0% run, Subjects 13 and 19 again display extremely smaller weights (0.0102 and 0.0062, respectively) and longer distances (7.7545 and 9.2987, respectively) than the others. The two subjects show recovered weights (0.0278 and 0.0241, respectively) and shorten distances (3.3695 and 5.3326, respectively) in the post-test session respectively in

Figures 5.11(b) and 5.11(d), but those are still smaller and longer than the others. Generally, the distributions for the weights and distances for the 100% run are similar with those for the 0% run. We detect multiple clusters for both pre- and post-test sessions in this run, since the group contains multiple subjects with small, medium, and large weights.

The estimated weights and distances for the 25% run are presented by green dashed lines in Figure 5.11. In the pre-test session in Figures 5.11(a) and 5.11(c), we note that 5 subjects (1, 12, 13, 18, and 19) have the estimated weights smaller than 0.02 with the distances longer than 5.3. At the same time, we observe that 4 subjects (8, 10, 20, and 30) display the weights larger than 0.04 and the distances around 2.5. Likewise, the result implies an inverse relationship of the weights and distances, and it also shows that the group can be considered as multiple clusters. We notice that those clusters can have a substantial effect on the construction of group graph. In Figures 5.11(b) and 5.11(d), while Subjects 12 and 19 still show relatively smaller weights (0.0218 and 0.0198, respectively) and longer distances (6.5591 and 7.43, respectively) than the others, Subjects 1, 13, and 18 recover their weights (0.0263, 0.0240, and 0.0322, respectively) and distances (3.2038, 4.5433, and 2.2297, respectively) closer to the standard level after the practice. The result shows that the estimated weights are more evenly distributed among the subjects than the pre-test session for this run.

For the 50% run (blue dashed line) in Figure 5.11, we observe more clear multiple clusters in both sessions. In the pre-test session, some subjects (1, 11, 12, 13, 15, and 16) form a cluster with small weights and long distances (the minimum weight = 0.0152 from Subject 13, and the longest distance = 4.9943 from Subject 13) and some (6, 8, 9, 20, 22, 30, and 31) compose a cluster with large weights and short distances (the maximum weight = 0.0400 from Subject 8, and the shortest distance = 1.2879 from Subject 9). Similarly in the post-test session, we find a cluster (Subjects 2, 4, 10, 11, 13, 17, and 19) with small weights and long distances (the minimum weight = 0.0134 from Subject 4, and the longest distance = 6.6193

from Subject 19), and a cluster (Subjects 1, 6, 8, 9, 20, 21, 25, 28, and 30) that has large weights and short distances (the maximum weight = 0.0470 from Subject 21, and the shortest distance = 1.1827 from Subject 1).

In the pre-test session for the 75% run (purple dotted line) in Figure 5.11(a), we observe that the weights of Subjects 2, 12, 13, 19, and 22 are lower (0.0163, 0.0154, 0.0176, 0.0102, and 0.0188, respectively) compared to the other weights. The distances presented in Figure 5.11(c) show that the corresponding subjects have longer distances (6.5133, 6.6269, 6.0462, 8.3160, and 5.8395, respectively). On the other hand, there is no extreme outlier in the post-test session for this run in Figures 5.11(b) and 5.11(d). Although there are possibly some clusters for this case, the estimated weights are relatively similar among subjects, range between 0.0247 and 0.0359. Overall, we show that the proposed method identifies potential outliers or multiple clusters throughout the runs in the event-related experiments. Consistent with the simulation study in Section 4.2, the estimated weights and distances between the median and each subject's inverse covariance move in opposite directions in all cases. Moreover, the groups in most of the runs turn out to be composed of multiple clusters; some subjects in the group show small weights, while the others present medium or large weights. Subsequently, we examine the effect of those clusters detected by the proposed JGL with RA method on a group graph structure that can induce different interpretations.

Table 5.7 describes changes of the relationships in ROIs by JGL with RA method between the pre- and post-test sessions for the 0% run. We compare absolute differences between the two sessions, and include the edges ($\hat{\omega}_{ij}$) where the differences are greater than 0.02. If the absolute differences are greater than 0.06, we emphasize those regions with an asterisk. In Table 5.7, we observe that the SEF and mFEF have some edges that become stronger after the practice. On the contrary, there are no regions that have weaker relationships. The result implies that the saccade tasks practice strengthens the associations in the visual fields of brain, especially in the SEF and mFEF regions for the 0% run.

We illustrate the group graphs generated by the proposed method and JGL for the 0% run in Figure 5.12. In the figure, the pre- and post-test sessions are presented in different rows, and the group graphs estimated by our method and JGL are displayed in the first and second column, respectively. We observe that the group graphs constructed by our method in both pre- and post-test sessions are more sparse than those from JGL. For example, in the pre-test session in Figure 5.12(a), many edges with IFC, lFEF, and Precun regions captured in the group graph by JGL in Figure 5.12(b) disappear. By noting that the group can be viewed as multiple clusters in Figure 5.11, the result is consistent with the simulation conducted in Section 4.2.4; the proposed method reduces the effect of outlying subjects whose estimated inverse covariance are dense, and produces more sparse group graph than those by JGL. The difference between the two graphs in the post-test session in Figures 5.12(c) and (d) are marginal, since the weights are more evenly distributed than the pre-test session in Figure 5.11. In comparison between the pre- and post-test sessions, consistent with Table 5.7, the group graph in Figure 5.12(c) has more and the stronger edges than that in Figure 5.12(a). On the contrary, the JGL group graph for the post-test session in Figure 5.12(d) is more sparse than that for the pre-test session in Figure 5.12(b), which results in a reversal interpretation due to the effect of outliers.

Table 5.8 summarizes the changes of associations for the 100% run. Similar to the 0% run, we find that most of the associations become stronger after the practice. In particular, the edges with the PFC, SEF, lFEF, and Precun regions get stronger relationships after the practice, whereas we cannot observe any regions with the associations that become weaker.

Figure 5.13 demonstrates the group graphs for the 100% run. We note that very strong associations such as lFEF-SEF, mFEF-SEF, SEF-PFC, and SEF-Precun are observed in all of the graphs. In the group graphs in Figures 5.13(a) and 5.13(c), JGL with RA tends to provide more sparse group graphs than JGL. Likewise, such sparsity is caused by the multiple clusters detected in the pre-test in Figure 5.11(a) and in the post-test in Figure

5.11(b) for the 100% run. Corresponding to Table 5.8, the strength of several relationships such as lFEF-SEF, lFEF-Precun, and SEF-PFC in the JGL with RA group graph from the post-test session in Figure 5.8(c) is stronger than that from the pre-test session in Figure 5.8(a).

We summarize the changes in associations for 25% run in Table 5.9. The result shows a similar number of relationships with the SEF region that become stronger after the practice. Particularly, we observe that the SEF-Thal association becomes stronger than the others. On the other hand, the PFC-Precun association gets weaker in this run.

Figure 5.14 displays the group graphs created by JGL with RA and JGL for the 25% run. We again observe that the group graphs estimated by our method in Figures 5.14(a) and 5.14(c) are more sparse than those by JGL in Figures 5.14(b) and 5.14(d), especially in the pre-test session. The proposed method yields more interpretable brain network compared to the dense group graph by JGL such as Figure 5.14(b). As indicated in Table 5.9, the group graph for the post-test session in Figure 5.14(c) shows more associations than that for the pre-test session in Figure 5.14(a). In the post-test session, the associations such as SEF-IPL, SEF-BG, and SEF-Thal are captured in the graph and the lFEF-SEF, SEF-Precun, and SEF-IFC associations become stronger.

We present a summary of the changes in associations for the 50% run in Table 5.10. In the table, we observe that all of the relationships regarding to SEF region get strongly connected after the practice. Especially, the SEF-IFC, SEF-Thal, and SEF-IPL associations show greater increments compared to the others. We can say that the practice reinforces the associations mainly with SEF, based on the fact that there is no relationship that becomes weaker in this run.

Figure 5.15 illustrates the group graphs for the 50% run. As observed in Table 5.10, the group graph estimated by JGL with RA in the post-test session in Figure 5.15(c) is more dense than that from the pre-test session in Figure 5.15(a) and captures all of the

associations regarding to SEF region. Likewise, our method provides more sparse graphs as in Figures 5.15(a) and 5.15(c) than those estimated by JGL as in Figures 5.15(b) and 5.15(d) by reducing the effect of the subjects who provide dense estimated graphs.

Table 5.11 describes the changes in associations for the 75% run. We find that the edges with PFC, SEF, lFEF become stronger after the practice, while the mFEF-Precun relationship becomes weakly connected. The result indicates that the network becomes more dense after the practice, since it contains more strongly connected edges rather than weakly connected ones.

Lastly, we provide the group graphs estimated by JGL with RA and JGL for the 75% run in Figure 5.16. In the pre-test session in Figures 5.16(a) and 5.16(b), the proposed method produces more sparse and interpretable group graphs than that generated by JGL. The JGL group graph in 5.16(b) has a shape that is close to the fully connected graph, which results in impractical interpretation. In the post-test session in Figures 5.16(c) and 5.16(d), although there are minor edge differences, still the group graph by our method is more sparse than that by JGL. The result also coincides with Table 5.11, since the group graph in the post-test session in Figure 5.16(c) has more dense shape than that in the pre-test session in Figure 5.16(a). Overall, the proposed JGL with RA method provides more sparse graph than the JGL group graph in the event-related experiments, due to the effect that is caused by the multiple clusters. We also find that the relationships especially with the SEF region become stronger after the practice.

5.3.2 Block Runs

We investigate the block runs based on the proposed JGL with RA method and JGL. In the same way as the event-related experiments, the estimated weights, distances from the median to each subject's estimated inverse covariance, and group graphs are reported to assess the performance of our method. Since the group is composed of the same 32 subjects

in the event-related experiments, we detect outliers whose estimated weights are smaller than $1/32 = 0.0313$.

Figure 5.17 demonstrates the estimated weights and distances between the median and each subject's estimated inverse covariance from JGL with RA for the pre- and post-test sessions in the block runs. For the estimated weights in the pre-test session in Figure 5.17(a), we observe that Subjects 2, 12, 13, and 19 have smaller assigned weights (0.0258, 0.0257, 0.0223, and 0.0250, respectively) than the other subjects. The distances for the corresponding subjects in Figure 5.17(c) are longer than the others (4.7343, 4.3782, 6.3479, and 5.2167), which again implies an inverse relationship between the weights and distances. We also find the inverse relationship in the post-test session in Figures 5.17(b) and 5.17(d); Subjects 12, 19, 24, and 32 show lower estimated weights (0.0229, 0.0225, 0.0258, and 0.0258, respectively) and longer distances (4.7293, 5.0764, 4.2460, and 4.0981, respectively) than the others. At the same time, we note that Subjects 2, 7, and 25 have large assigned weights (0.0379, 0.0389, and 0.0405, respectively). The result indicates that the group is composed of multiple clusters similar to that in the event-related experiments. In both sessions, the weight difference is not as large as in the event-related experiments since there is no extreme outlier whose estimated weight is close to zero. However, we need to take multiple clusters in the group into account on aggregation of individual precision matrices.

Table 5.12 summarizes the changes in associations after the practice for the block runs. In contrast to the event-related experiments, we observe that the associations regarding SEF region become weaker after the practice. We also find weaker associations than stronger ones in the block runs. The result may be induced by different experimental designs between the event-related experiments and block runs.

We present the group graphs estimated by JGL with RA and JGL in the pre- and post-test sessions for the block runs in Figure 5.18. Similar to the event-related experiments, the JGL with RA group graphs for both sessions in Figures 5.18(a) and 5.18(c) are more sparse than

the JGL group graphs in Figures 5.18(b) and 5.18(d). As we have seen in Figure 5.17, the difference between the two methods' group graphs is not that large since there is no extreme outlier. Consistent with Table 5.12, the post-test group graph in Figure 5.18(c) becomes more sparse than the pre-test group graph in Figure 5.18(a). However, we also note that some associations with the lFEF region are captured after the practice, such as lFEF-Thal and lFEF-Cun.

In summary, the proposed JGL with RA method provides more sparse group graphs in both event-related experiments and block runs than those estimated from JGL. We show that our method compares the distances between the median and each subject's estimated inverse covariance, and assign smaller weights to the subject whose estimated inverse covariance is much deviated from the median inverse covariance. Such differently assigned weights form multiple clusters in most of the runs. As we have shown in the simulation study in Section 4.2.4, the weights lead us to the robust estimation based on the reduced effect of the subjects whose estimated inverse covariances are more dense than the others.

In Figure 5.19, we compare the estimated group graphs by our method and JGL using graph kernels (Vishwanathan et al., 2010). As introduced in Section 5.2, we use the Weisfeiler-Lehman (W-L) graph kernel (Shervashidze et al., 2011) to quantify the similarity of pairs of graphs. The higher W-L graph kernel value indicates the more similar graph pairs. In the figure, we observe that the graph pair for the 25% run in the post-test session of the event-related experiments produces the highest value of 22, whereas the graph pairs for the 100%, 25%, and 75% runs in the pre-test session of the event-related experiments yield the lowest values of 11. The result complies with the estimated weights in Figure 5.11; the estimated weights among the subjects for the 100%, 25%, and 75% runs in the pre-test session are more diverse than the other runs, while the weights for the 25% run in the post-test session show relatively less deviation. We also note that the value becomes 66 when a graph pair has the exactly same form in Figure 5.19. Since all of the values in Figure 5.19 are much smaller

than 66, we can say that JGL with RA identifies outliers or multiple clusters and provides different group graphs from those by JGL in all of the runs in the saccade tasks fMRI data.

Additionally, Figure 5.20 illustrates the W-L graph kernel values for the pre- and post-test session group graph pairs constructed from the proposed JGL with RA method. In the figure, we notice that the graph pair for the 0% run in the event-related experiments shows the highest value of 17, while the pair for the 75% run in the event-related experiments has the lowest value of 12. The result implies that the pre- and post-test group graphs for the 75% run in the event-related experiments are the most dissimilar, and more affected by the practice than the other runs.

5.4 Comparisons

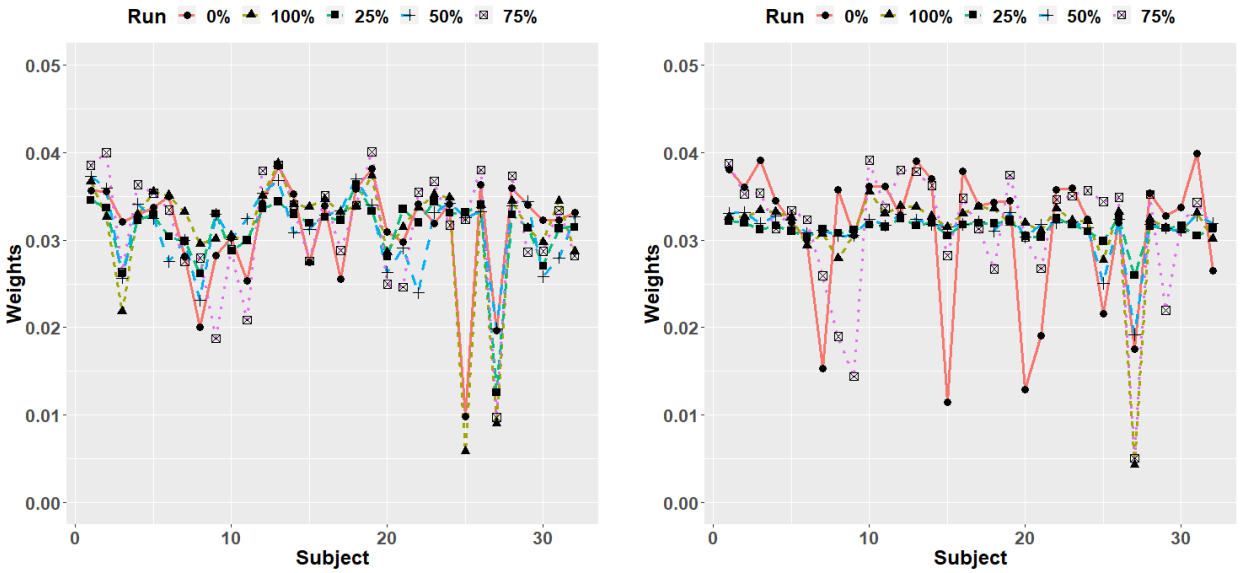
We compare the results from JEMP with RA and JGL with RA in this section. The differences between two methods are largely due to the Gaussian assumption. The JEMP with RA estimator does not need the Gaussian distributional assumption, whereas the JGL with RA estimator does. To verify the differences, we investigate which subjects are identified as outliers by the two methods. As an example, Figure 5.21 illustrates the covariance heat maps from the 100% run for the pre-test session in the event-related experiments. We provide the heat maps from Subjects 12, 16, 19, and 25 in the figure. In both methods, Subjects 12 and 16 can be considered as standard subjects, since they show the estimated weights which are close to $1/32 = 0.0313$ (0.0351 and 0.0347 from JEMP with RA, and 0.0323 and 0.0318 from JGL with RA, respectively). However, Subject 19 has small assigned weight (0.0062) by JGL with RA (0.0374 from JEMP with RA), and Subject 25 displays small weight (0.0059) by JEMP with RA (0.0426 from JGL with RA). The reason for such differently assigned weights can be found in the covariance heat maps in Figure 5.21. In the figure, we observe that Subjects 12 and 16 show medium-red colors, Subject 19 displays dark-red colors, and Subject 25 presents light-red color coded heat map. The proposed JEMP with RA method

employs the l_1-minimization of inverse covariance matrix, thus it penalizes small covariances more than large ones. This estimation property affects the aggregation, which results in assigning small weights to the subjects whose covariances are much smaller than the others. On the other hand, JGL with RA finds the penalized maximum likelihood estimator of inverse covariance, and compares the distances from roughly the median to each subject's estimated inverse covariance in aggregation. In our saccade tasks fMRI data, the method tends to assign lower weights to the subjects with large covariance, since the distances between the rough median and those subjects' estimated inverse covariances are larger than the others. We have seen that the group in the data generally has a form of multiple clusters, and the results are consistent with those from the simulation studies in Sections 4.1.3 and 4.2.4.

The different estimation properties of the two proposed methods also affect constructions of their group graphs. In Section 5.2, we have seen that JEMP with RA captures more associations than JEMP in the group graphs. This means that the JEMP with RA is more robust to the subjects who have sparse precision matrices (small covariances) than JEMP. On the contrary, JGL with RA produces more sparse group graphs than JGL in Section 5.3. Since the JGL with RA penalizes the subjects who have the estimated precision matrices much deviated from the median, it results in providing robust group graphs to the subjects with more dense structures (large covariances) in the data.

Therefore, the two proposed methods yield different results in changes of associations after the practice. From the results of JEMP with RA in Section 5.2, we have observed that the associations with Cun become stronger, while those with lFEF become weaker after the practice. Moreover, we find that the numbers of edges that become stronger and weaker after the practice are similar in general, which indicates that the sparsity between the two sessions remains at the same level. On the other hand, the JGL with RA detects that the relationships with SEF get stronger after the practice in most of the runs. We also notice that more associations become stronger than weaker in JGL with RA.

We observe that the common relationships such as lFEF-mFEF, lFEF-SEF, mFEF-SEF, SEF-PFC, SEF-Precun, and SEF-Thal are found by both methods in most of the runs. On the other hand, the strong relationships presented in JEMP with RA such as Thal-BG and MOG-Cun disappear in JGL with RA. To validate the reason that some relationships are found in both methods, some are very strong in one, and no relationship in the other, we conduct an additional analysis (results are provided in Appendices A). In specific, we standardize the data and compute covariances before applying the two proposed methods. Then we employ our methods to estimate the weights and group graphs using the standardized covariances. In the results, we are not able to observe a clear weight difference among subjects due to the standardization. Consequently, the group graphs estimated by JEMP with RA and JEMP, and JGL with RA and JGL are similar to each other. The JEMP with RA produces group graphs consistent with those before the standardization. However, we find that the JGL with RA group graphs show similar shapes to the JEMP with RA ones after the standardization. The associations appeared in the JEMP with RA group graphs such as Thal-BG, MOG-Cun, lFEF-mFEF, lFEF-SEF, and mFEF-SEF are also captured in the JGL with RA graphs in most of the runs. From this result, we can conclude that JGL with RA is more affected by the standardization than JEMP with RA. The standardized covariance structure used in JGL with RA influences the result due to the Gaussian assumption, and we can say that JEMP with RA provides more consistent outcomes than JGL with RA for this data. We also note that typical pre-processed fMRI data contain the sets of data which are already on the same scale. Hence, we recommend not to standardize the data if the research objective is to aggregate individual-wise information based on different patterns among subjects.



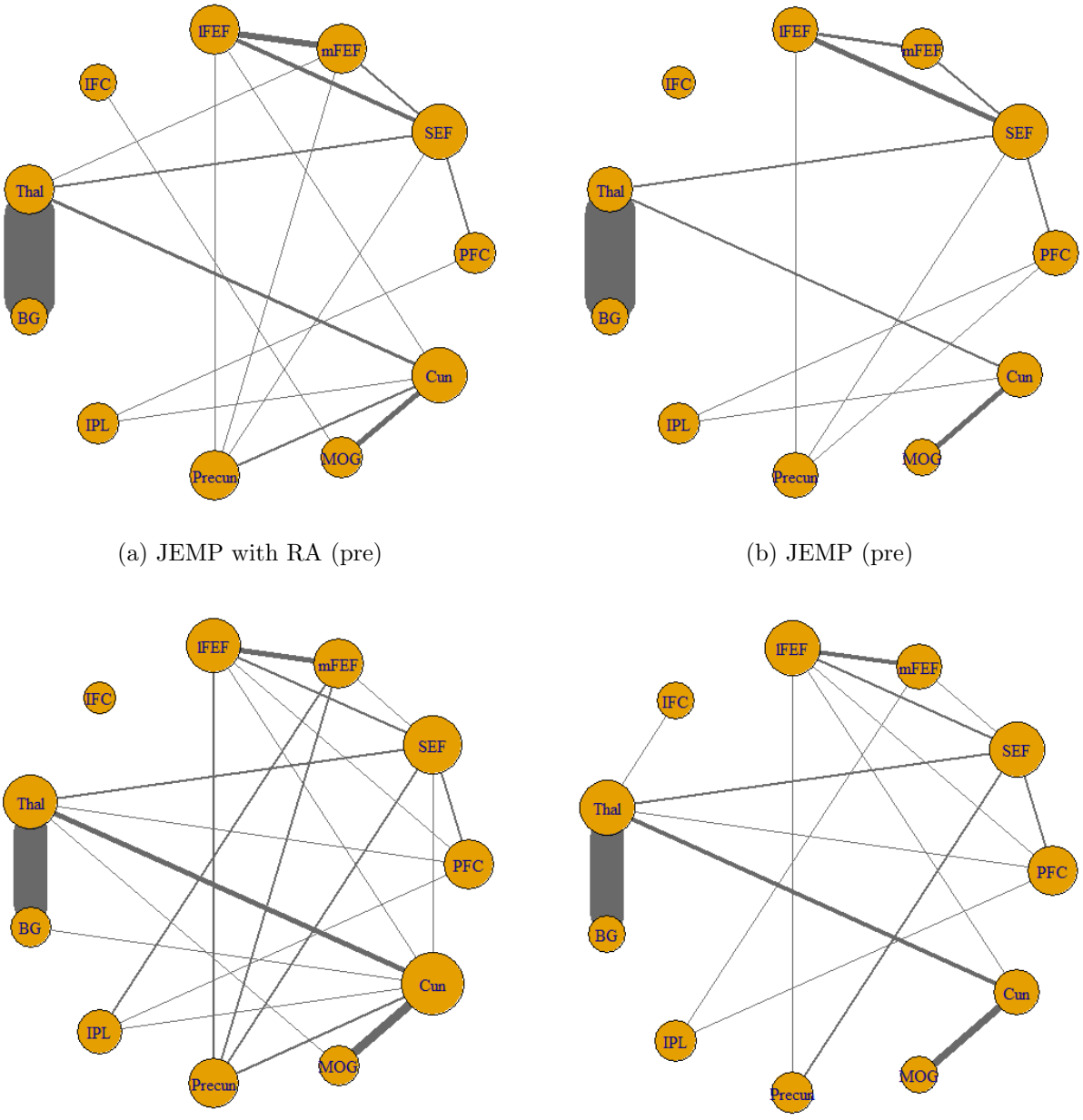
(a) Event experiments (Pre)

(b) Event experiments (Post)

Figure 5.1: Comparison of the estimated weights by JEMP with RA from two sessions for the 5 probability runs in each panel. The horizontal and vertical axes are subject index and the estimated weights, respectively.

Table 5.1: Change in relationships after the practice by JEMP with RA for the 0% run. The only regions which have absolute differences in $\hat{\omega}_{ij}$ greater than 0.5 between the pre- and post-test sessions are included. Especially, the regions with an asterisk (*) have absolute differences in $\hat{\omega}_{ij}$ greater than 1.5.

Stronger (Pre < Post)		Weaker (Pre > Post)	
PFC	lFEF	PFC	-
SEF	-	SEF	mFEF, lFEF
mFEF	IPL*, Precun	mFEF	lFEF, Thal
lFEF	Precun	lFEF	-
IFC	-	IFC	-
Thal	Cun*	Thal	BG*
BG	Cun	BG	-
IPL	-	IPL	-
Precun	-	Precun	Cun
MOG	Cun*	MOG	-



(a) JEMP with RA (pre)

(b) JEMP (pre)

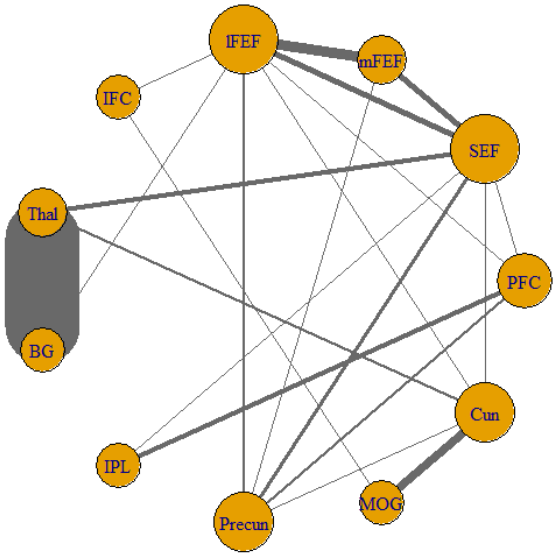
(c) JEMP with RA (post)

(d) JEMP (post)

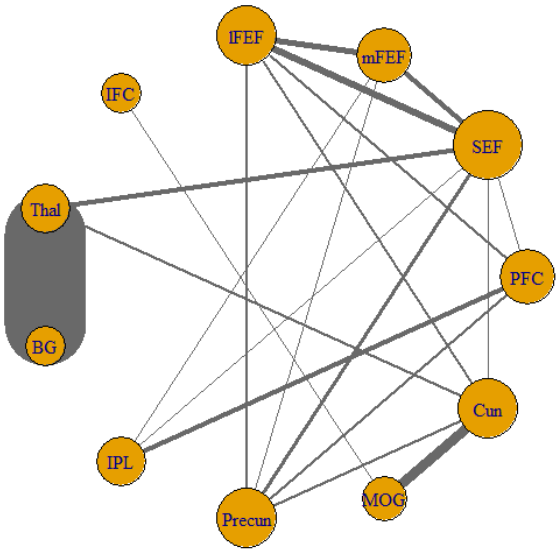
Figure 5.2: Group graphs by JEMP with RA and JEMP for the 0% run. The edges weaker than 0.3 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

Table 5.2: Change in relationships after the practice by JEMP with RA for the 100% run. The only regions which have absolute differences in $\hat{\omega}_{ij}$ greater than 0.5 between the pre- and post-test sessions are included. Especially, the regions with an asterisk (*) have absolute differences in $\hat{\omega}_{ij}$ greater than 1.5.

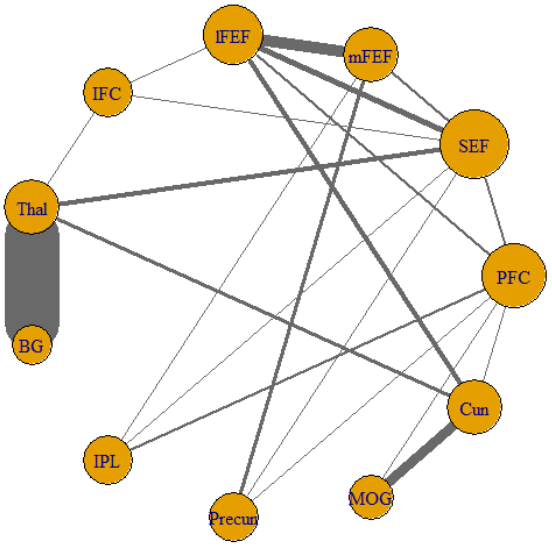
Stronger (Pre < Post)		Weaker (Pre > Post)	
PFC	SEF, lFEF, MOG	PFC	IPL, Precun
SEF	IFC, IPL	SEF	mFEF*, Precun, Cun
mFEF	IPL, Precun	mFEF	-
lFEF	Cun*	lFEF	Precun
IFC	-	IFC	MOG
Thal	-	Thal	BG*
BG	-	BG	-
IPL	-	IPL	-
Precun	-	Precun	Cun
MOG	Cun	MOG	-



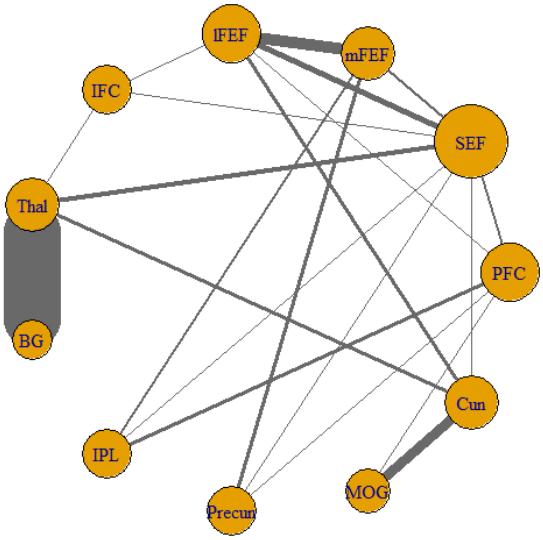
(a) JEMP with RA (pre)



(b) JEMP (pre)



(c) JEMP with RA (post)

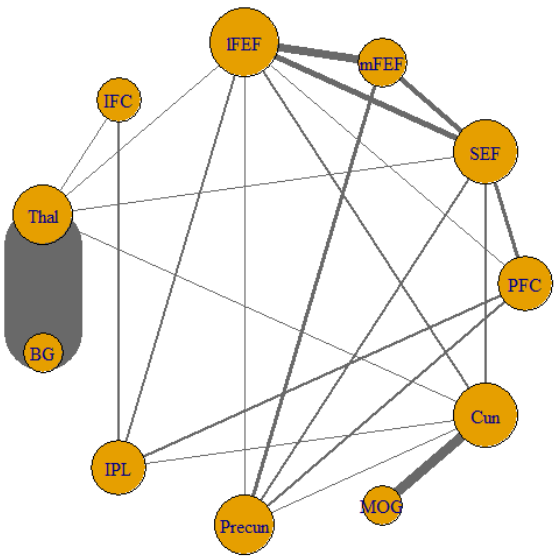


(d) JEMP (post)

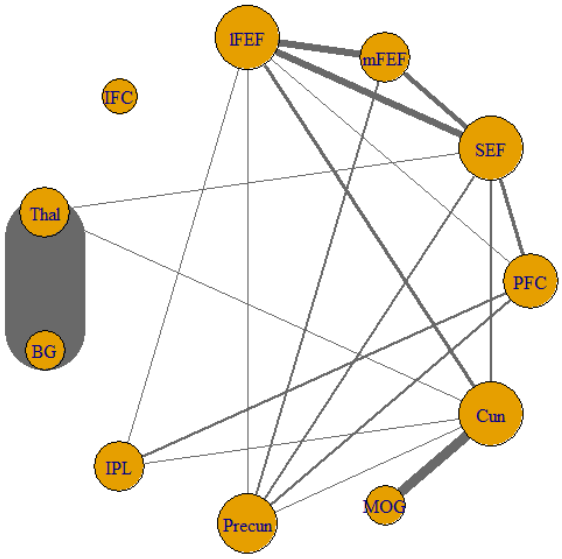
Figure 5.3: Group graphs by JEMP with RA and JEMP for the 100% run. The edges weaker than 0.3 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

Table 5.3: Change in relationships after the practice by JEMP with RA for the 25% run. The only regions which have absolute differences in $\hat{\omega}_{ij}$ greater than 0.5 between the pre- and post-test sessions are included. Especially, the regions with an asterisk (*) have absolute differences in $\hat{\omega}_{ij}$ greater than 1.5.

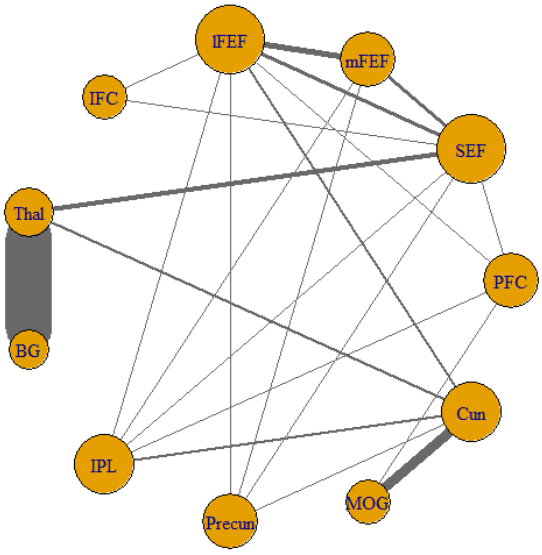
Stronger (Pre < Post)		Weaker (Pre > Post)	
PFC	MOG	PFC	SEF, lFEF, IPL, Precun*
SEF	Thal*	SEF	lFEF, Cun*
mFEF	IPL	mFEF	lFEF, Precun
lFEF	IFC, Precun	lFEF	IPL
IFC	-	IFC	IPL*
Thal	Cun	Thal	BG*
BG	-	BG	-
IPL	Cun	IPL	-
Precun	-	Precun	-
MOG	-	MOG	-



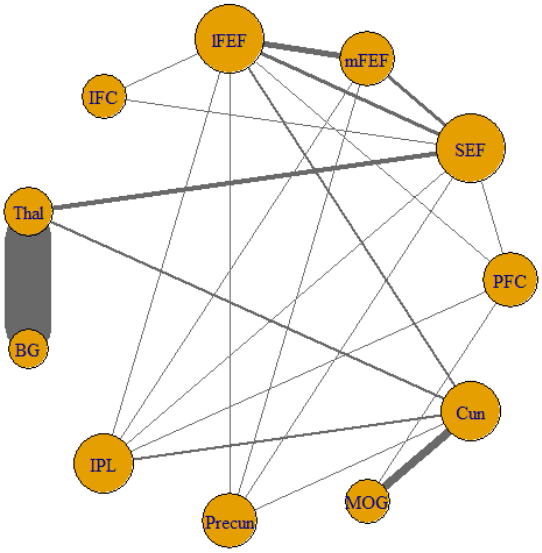
(a) JEMP with RA (pre)



(b) JEMP (pre)



(c) JEMP with RA (post)

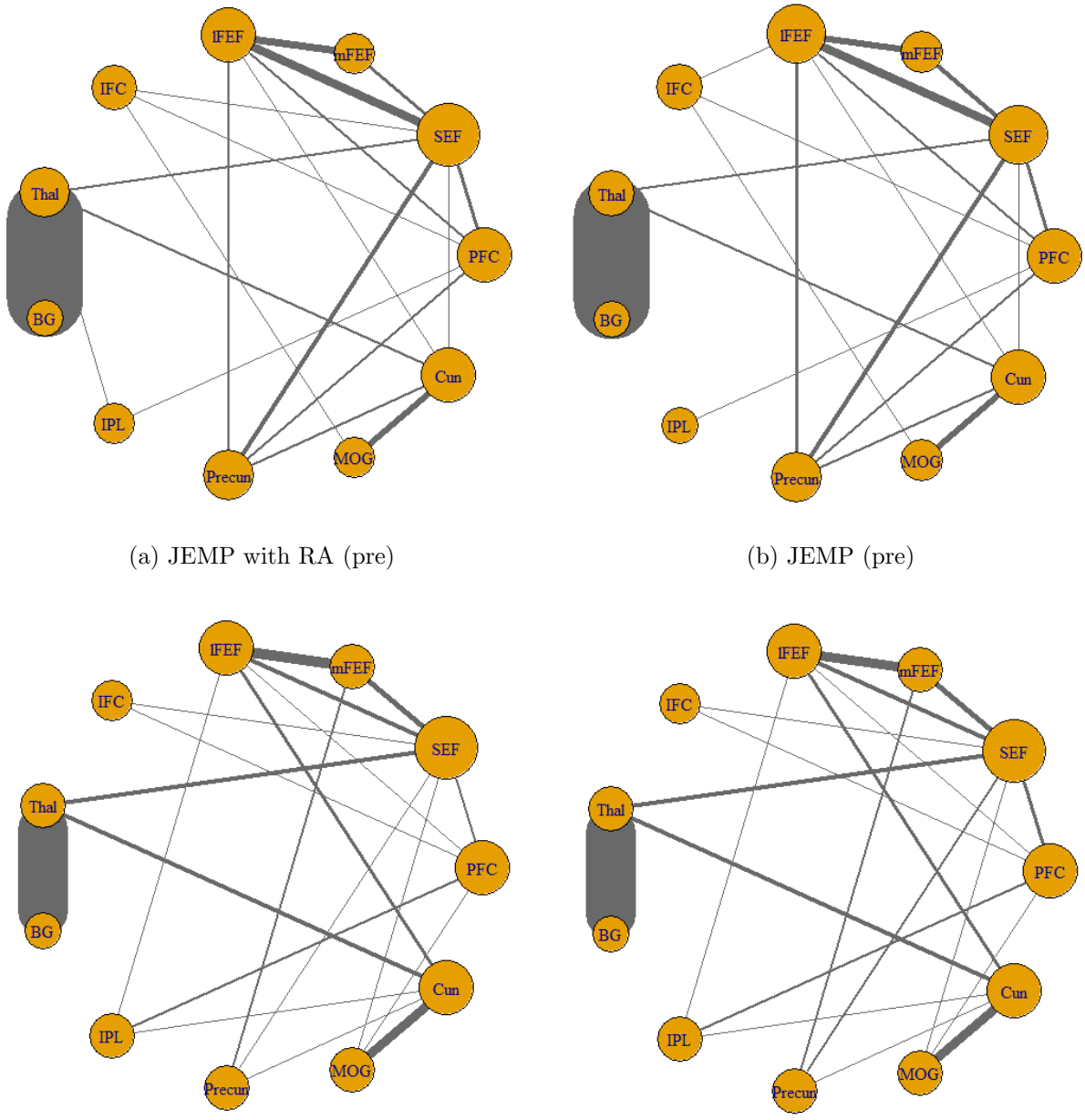


(d) JEMP (post)

Figure 5.4: Group graphs by JEMP with RA and JEMP for the 25% run. The edges weaker than 0.3 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

Table 5.4: Change in relationships after the practice by JEMP with RA for the 50% run. The only regions which have absolute differences in $\hat{\omega}_{ij}$ greater than 0.5 between the pre- and post-test sessions are included. Especially, the regions with an asterisk (*) have absolute differences in $\hat{\omega}_{ij}$ greater than 1.5.

Stronger (Pre < Post)		Weaker (Pre > Post)	
PFC	IPL, MOG	PFC	SEF, Precun*
SEF	mFEF*, Thal	SEF	lFEF*, Precun*, Cun
mFEF	lFEF*, Precun*	mFEF	-
lFEF	IPL, Cun	lFEF	Precun*
IFC	-	IFC	MOG
Thal	Cun	Thal	BG*
BG	-	BG	-
IPL	-	IPL	-
Precun	-	Precun	Cun
MOG	Cun	MOG	-



(a) JEMP with RA (pre)

(b) JEMP (pre)

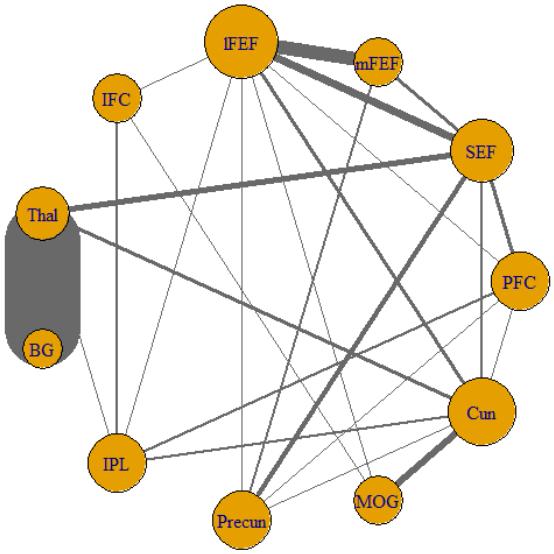
(c) JEMP with RA (post)

(d) JEMP (post)

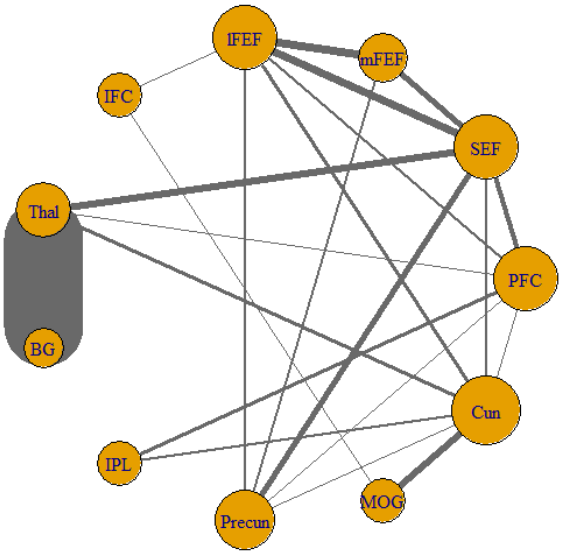
Figure 5.5: Group graphs by JEMP with RA and JEMP for the 50% run. The edges weaker than 0.3 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

Table 5.5: Change in relationships after the practice by JEMP with RA for the 75% run. The only regions which have absolute differences in $\hat{\omega}_{ij}$ greater than 0.5 between the pre- and post-test sessions are included. Especially, the regions with an asterisk (*) have absolute differences in $\hat{\omega}_{ij}$ greater than 1.5.

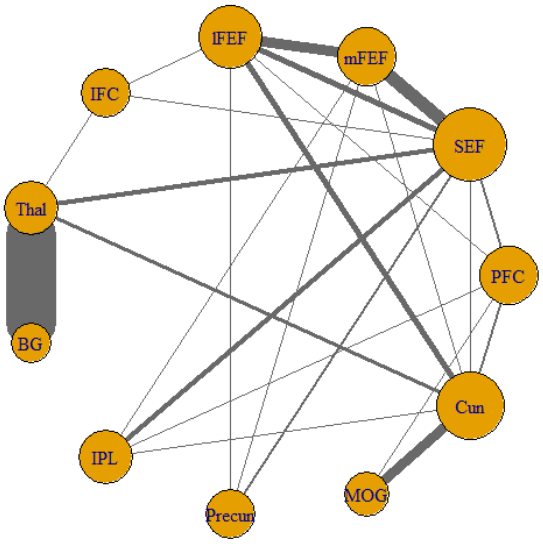
Stronger (Pre < Post)		Weaker (Pre > Post)	
PFC	MOG, Cun	PFC	SEF, lFEF, IPL, Precun
SEF	mFEF*, IFC, IPL*	SEF	lFEF, Thal, Precun, Cun
mFEF	Cun	mFEF	lFEF*, Precun
lFEF	Cun	lFEF	IFC, IPL
IFC	-	IFC	IPL*
Thal	-	Thal	BG*, IPL
BG	-	BG	-
IPL	-	IPL	Cun
Precun	-	Precun	Cun
MOG	Cun	MOG	-



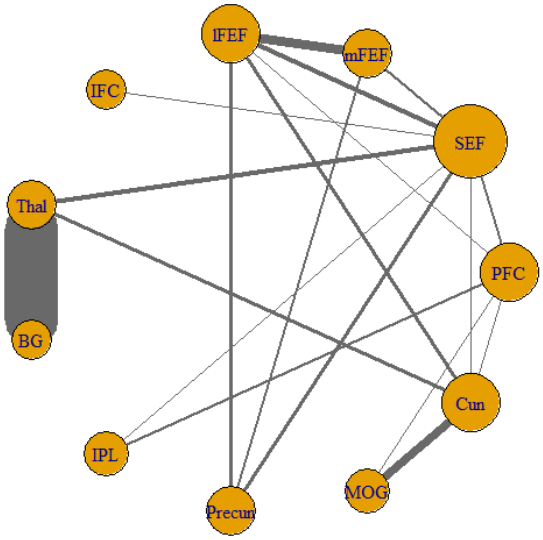
(a) JEMP with RA (pre)



(b) JEMP (pre)

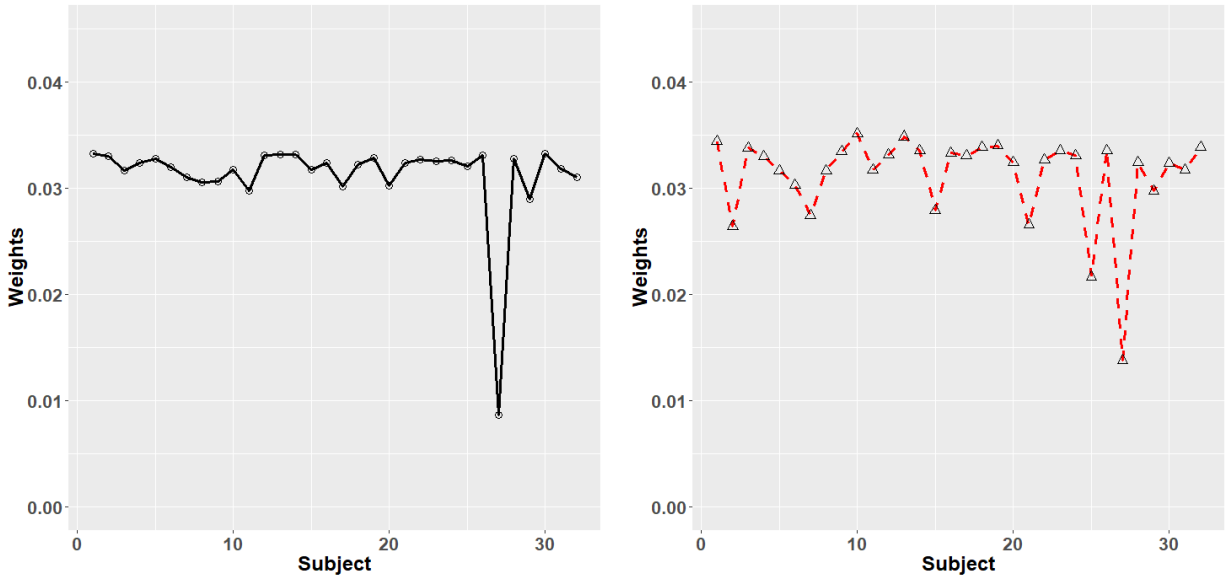


(c) JEMP with RA (post)



(d) JEMP (post)

Figure 5.6: Group graphs by JEMP with RA and JEMP for the 75% run. The edges weaker than 0.3 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.



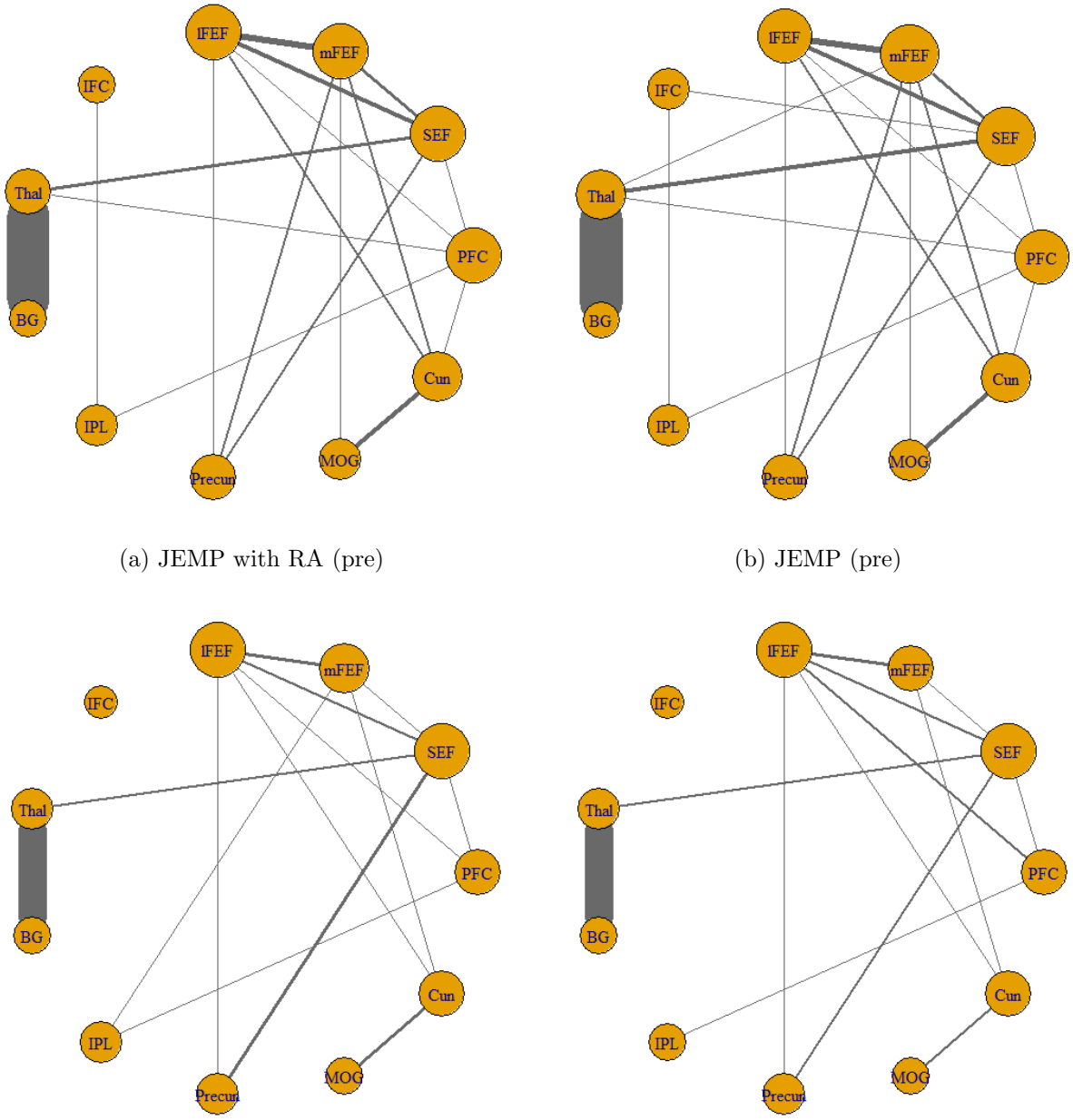
(a) Block Runs (Pre)

(b) Block Runs (Post)

Figure 5.7: Comparison of the estimated weights by JEMP with RA for the block runs in pre- and post-test sessions. In each panel, the horizontal and vertical axes are subject index and the estimated weights respectively.

Table 5.6: Change in relationships after the practice by JEMP with RA for the block runs. The only regions which have absolute differences in $\hat{\omega}_{ij}$ greater than 0.5 between the pre- and post-test sessions are included. Especially, the regions with an asterisk (*) have absolute differences in $\hat{\omega}_{ij}$ greater than 1.5.

Stronger (Pre < Post)		Weaker (Pre > Post)	
PFC	lFEF	PFC	SEF, Thal, IPL, Cun
SEF	-	SEF	mFEF*, lFEF, Thal
mFEF	-	mFEF	lFEF*, Precun*
lFEF	-	lFEF	-
IFC	-	IFC	IPL
Thal	-	Thal	BG*
BG	-	BG	-
IPL	-	IPL	-
Precun	-	Precun	-
MOG	-	MOG	Cun



(a) JEMP with RA (pre)

(b) JEMP (pre)

(c) JEMP with RA (post)

(d) JEMP (post)

Figure 5.8: Group graphs by JEMP with RA and JEMP for the block runs. The edges weaker than 0.3 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

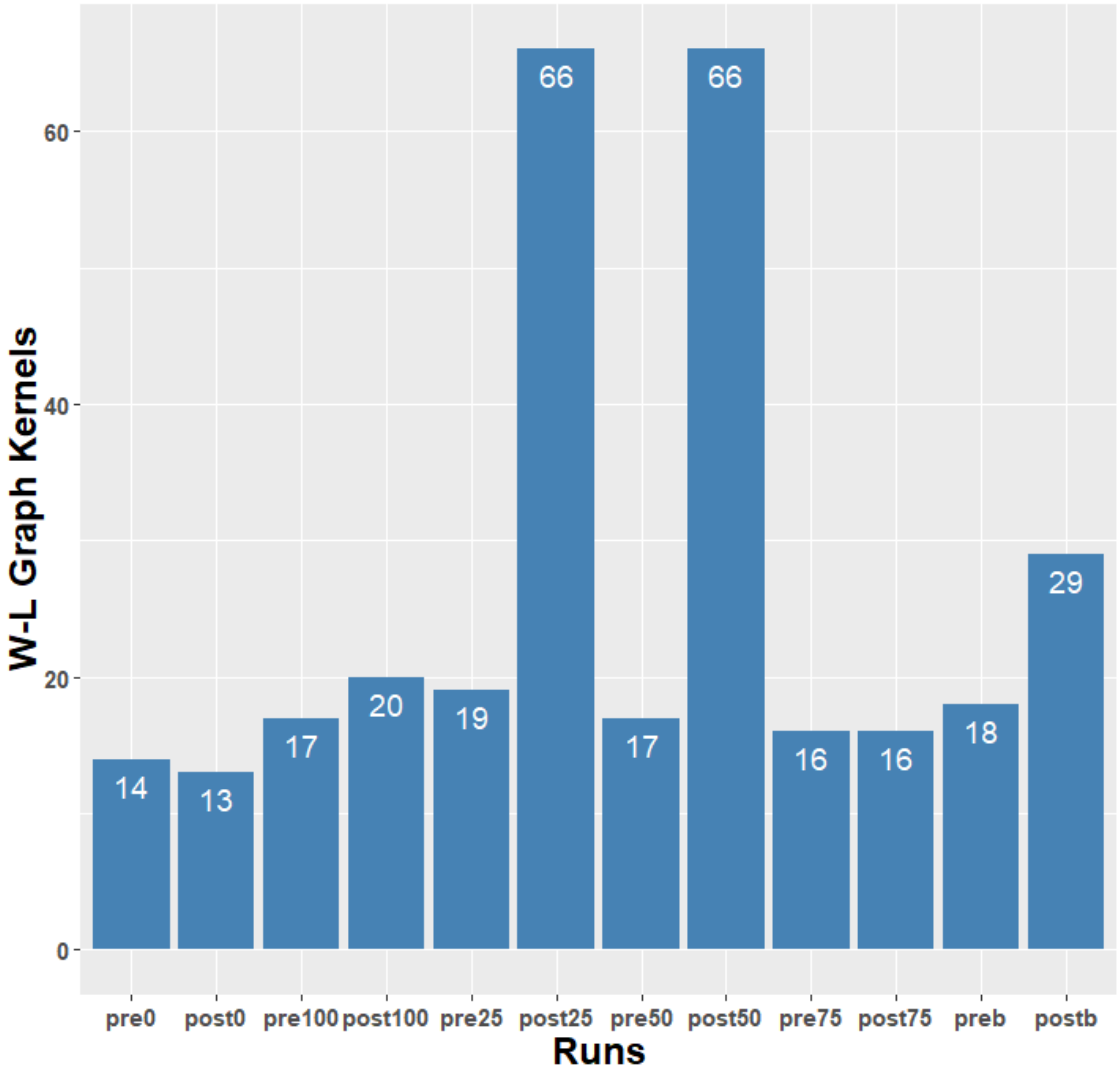


Figure 5.9: Comparison of JEMP with RA and JEMP group graphs based on Weisfeiler-Lehman graph kernels for each session and run. The horizontal and vertical axes are index of runs (e.g. "pre0" is for the 0% run in the pre-test session of the event-related expriments, and "postb" represents the post-test session of the block runs) and W-L graph kernel values respectively.

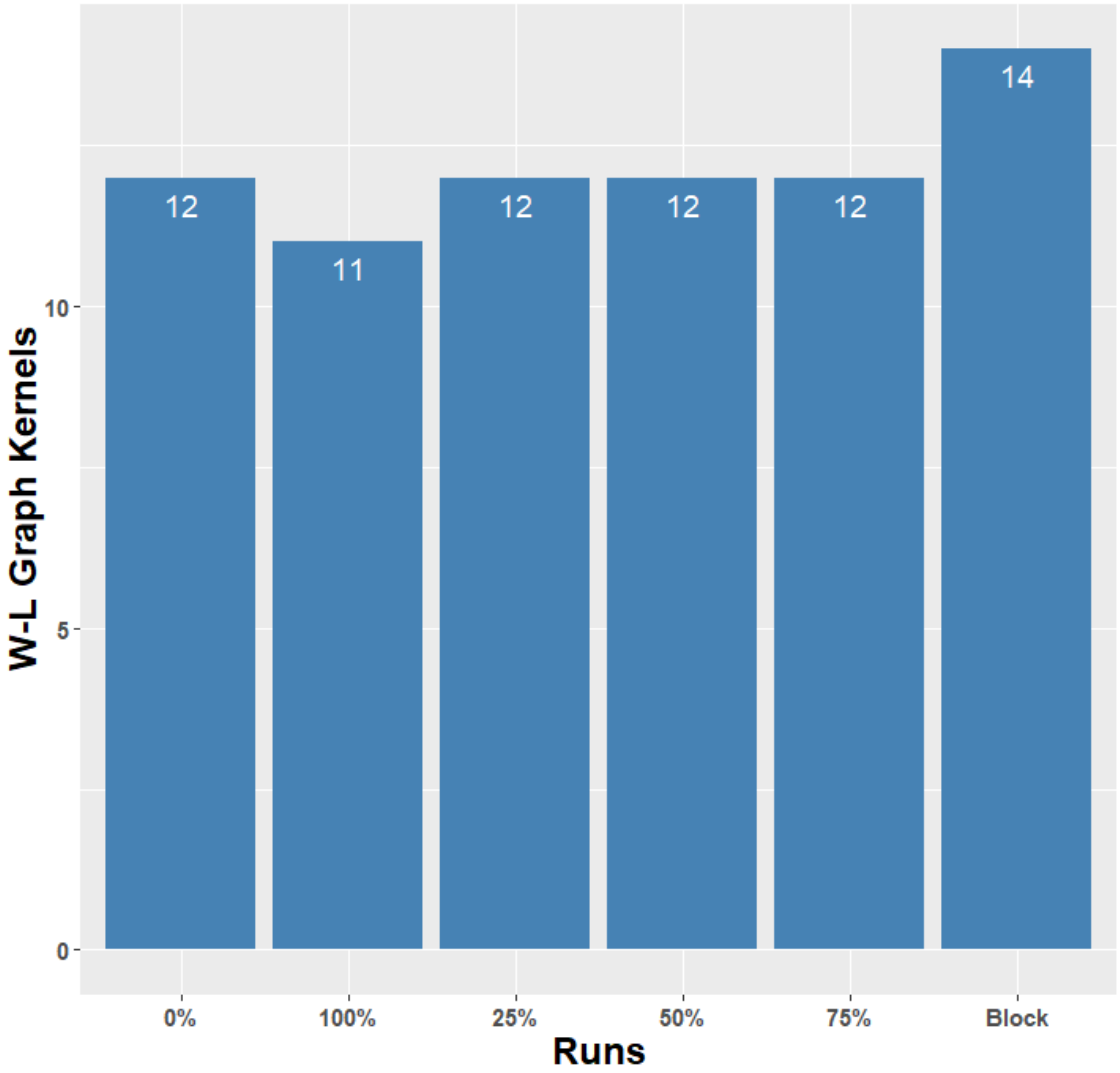
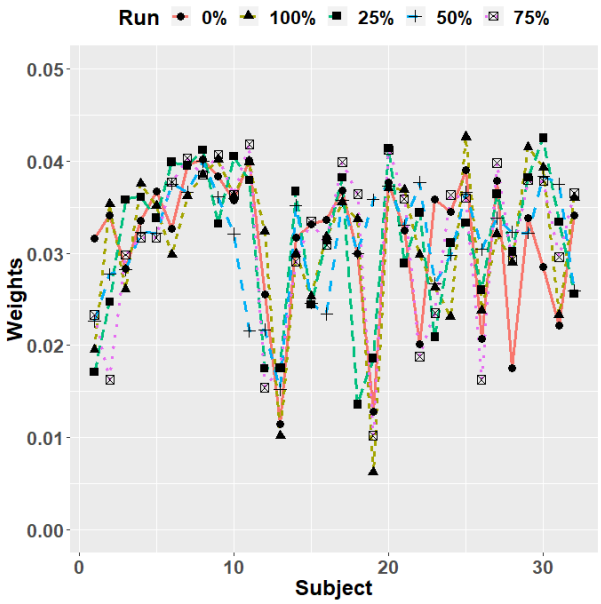
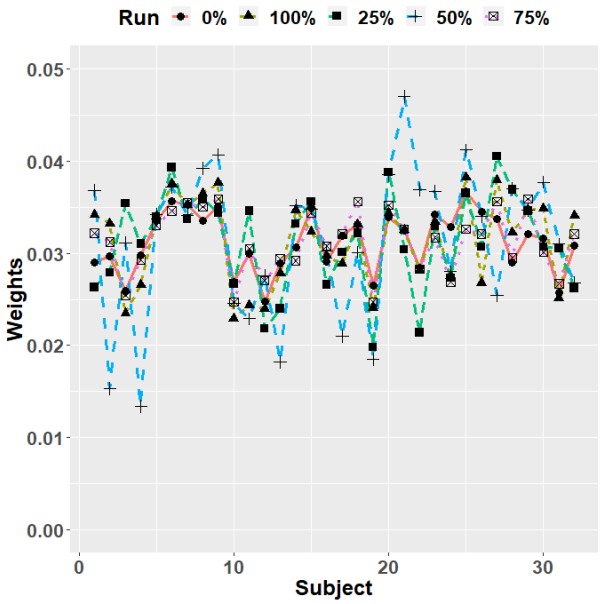


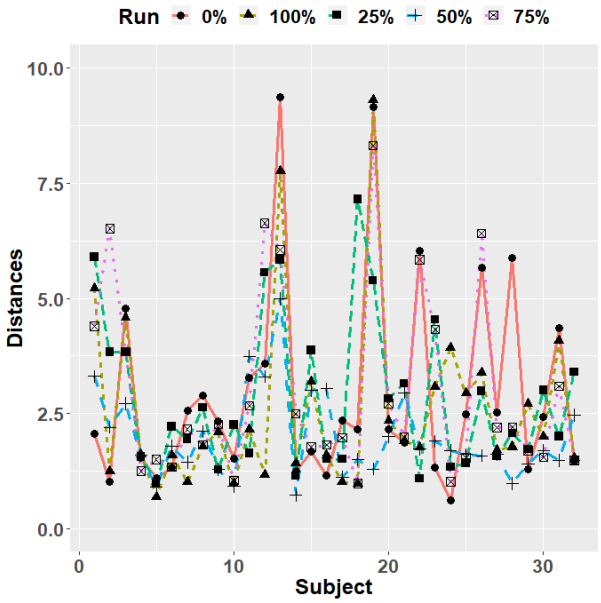
Figure 5.10: Comparison of JEMP with RA group graphs for the pre- and post-test session based on Weisfeiler-Lehman graph kernels. The horizontal and vertical axes are index of runs and W-L graph kernel values respectively.



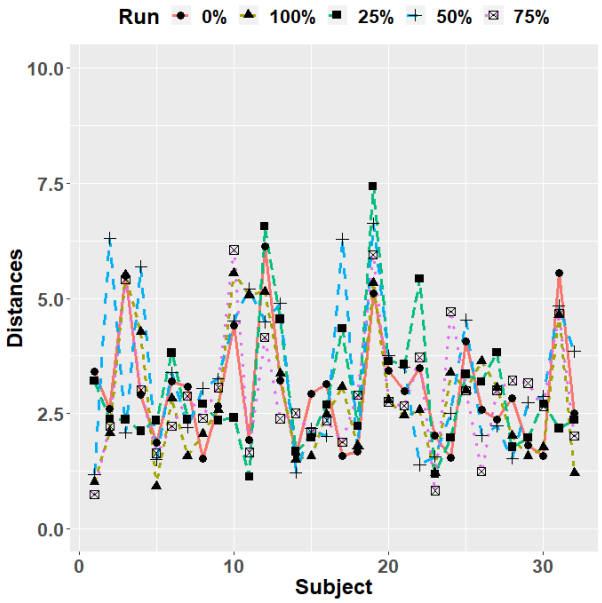
(a) Event experiments weights (Pre)



(b) Event experiments weights (Post)



(c) Event experiments distances (Pre)

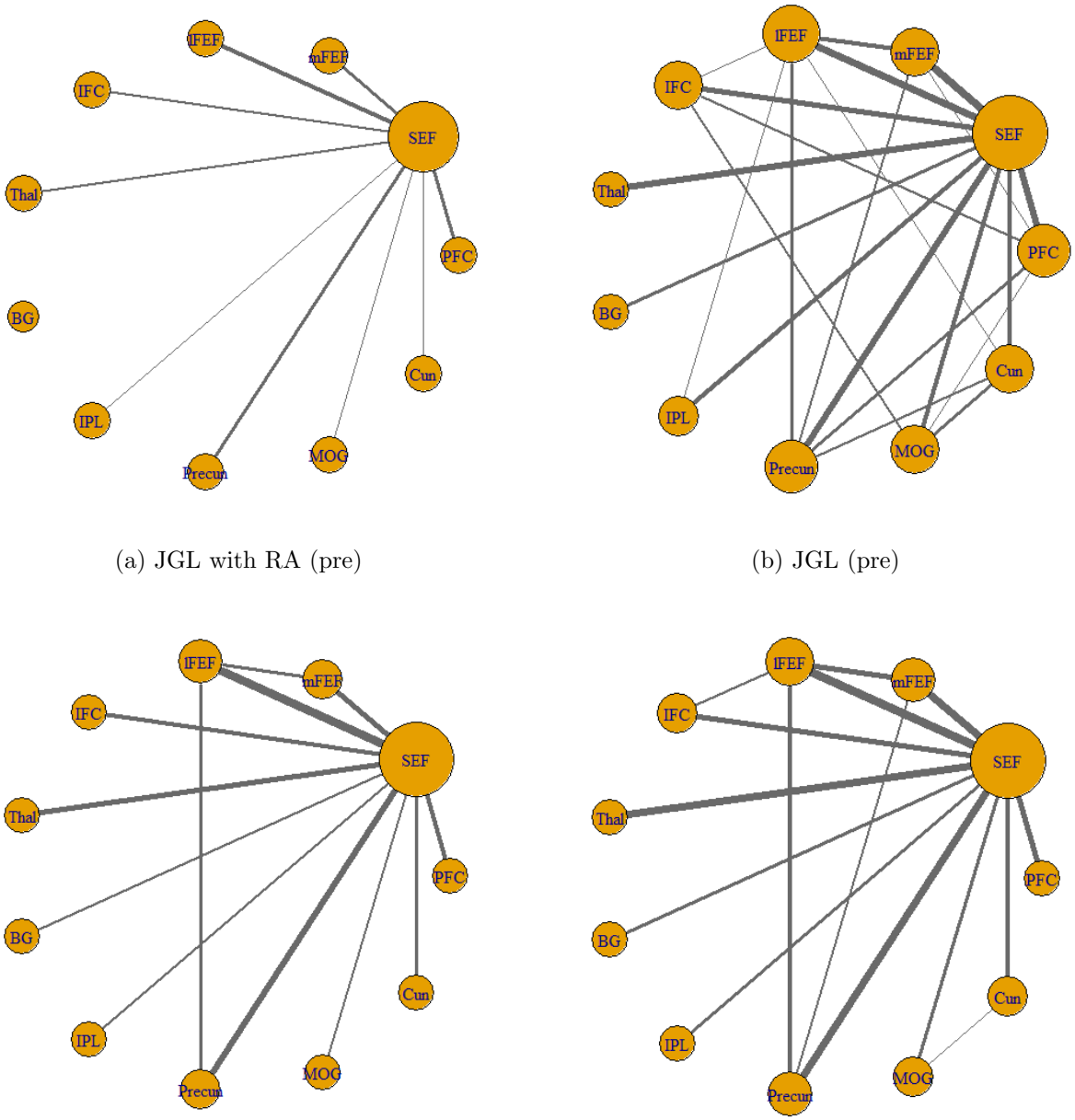


(d) Event experiments distances (Post)

Figure 5.11: Comparison of the estimated weights and distances by JGL with RA from two sessions for the 5 probability runs in each panel. The horizontal axe represents subject index, and the vertical axes are the estimated weights and the distances, respectively.

Table 5.7: Change in relationships after the practice by JGL with RA for the 0% run. The only regions which have absolute differences in $\hat{\omega}_{ij}$ greater than 0.02 between the pre- and post-test sessions are included. Especially, the regions with an asterisk (*) have absolute differences in $\hat{\omega}_{ij}$ greater than 0.06.

Stronger (Pre < Post)		Weaker (Pre > Post)	
PFC	-	PFC	-
SEF	mFEF, lFEF*, IFC, Thal*, BG, IPL, Precun, MOG, Cun	SEF	-
mFEF	Precun*	mFEF	-
lFEF	-	lFEF	-
IFC	-	IFC	-
Thal	-	Thal	-
BG	-	BG	-
IPL	-	IPL	-
Precun	-	Precun	-
MOG	-	MOG	-



(a) JGL with RA (pre)

(b) JGL (pre)

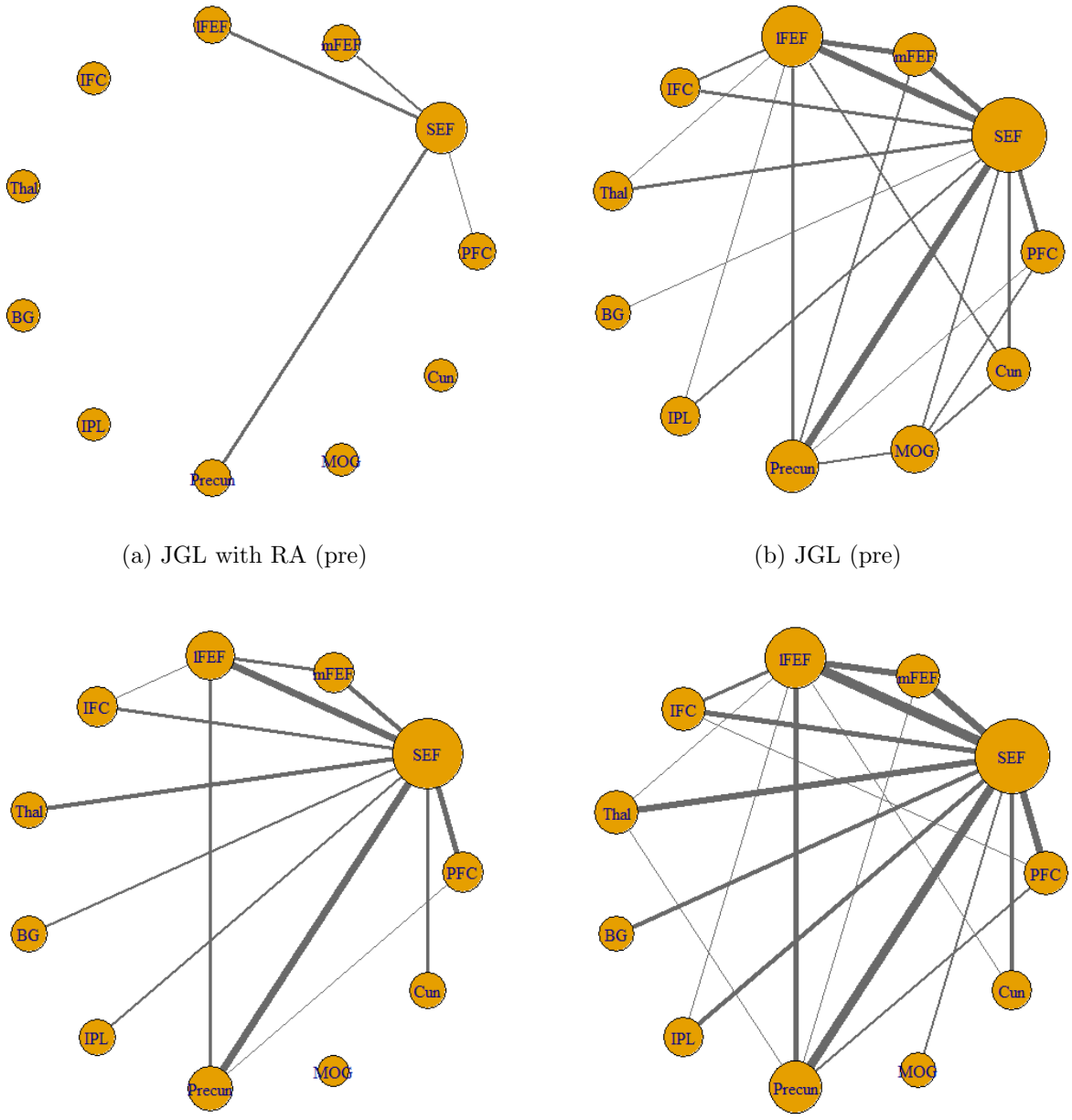
(c) JGL with RA (post)

(d) JGL (post)

Figure 5.12: Group graphs by JGL with RA and JGL for the 0% run. The edges weaker than 0.03 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

Table 5.8: Change in relationships after the practice by JGL with RA for the 100% run. The only regions which have absolute differences in $\hat{\omega}_{ij}$ greater than 0.02 between the pre- and post-test sessions are included. Especially, the regions with an asterisk (*) have absolute differences in $\hat{\omega}_{ij}$ greater than 0.06.

Stronger (Pre < Post)		Weaker (Pre > Post)	
PFC	SEF*, Precun	PFC	-
SEF	mFEF, lFEF*, IFC*, Thal*, BG, IPL, Precun*, Cun*	SEF	-
mFEF	lFEF*	mFEF	-
lFEF	IFC, Precun*	lFEF	-
IFC	-	IFC	-
Thal	-	Thal	-
BG	-	BG	-
IPL	-	IPL	-
Precun	-	Precun	-
MOG	-	MOG	-



(a) JGL with RA (pre)

(b) JGL (pre)

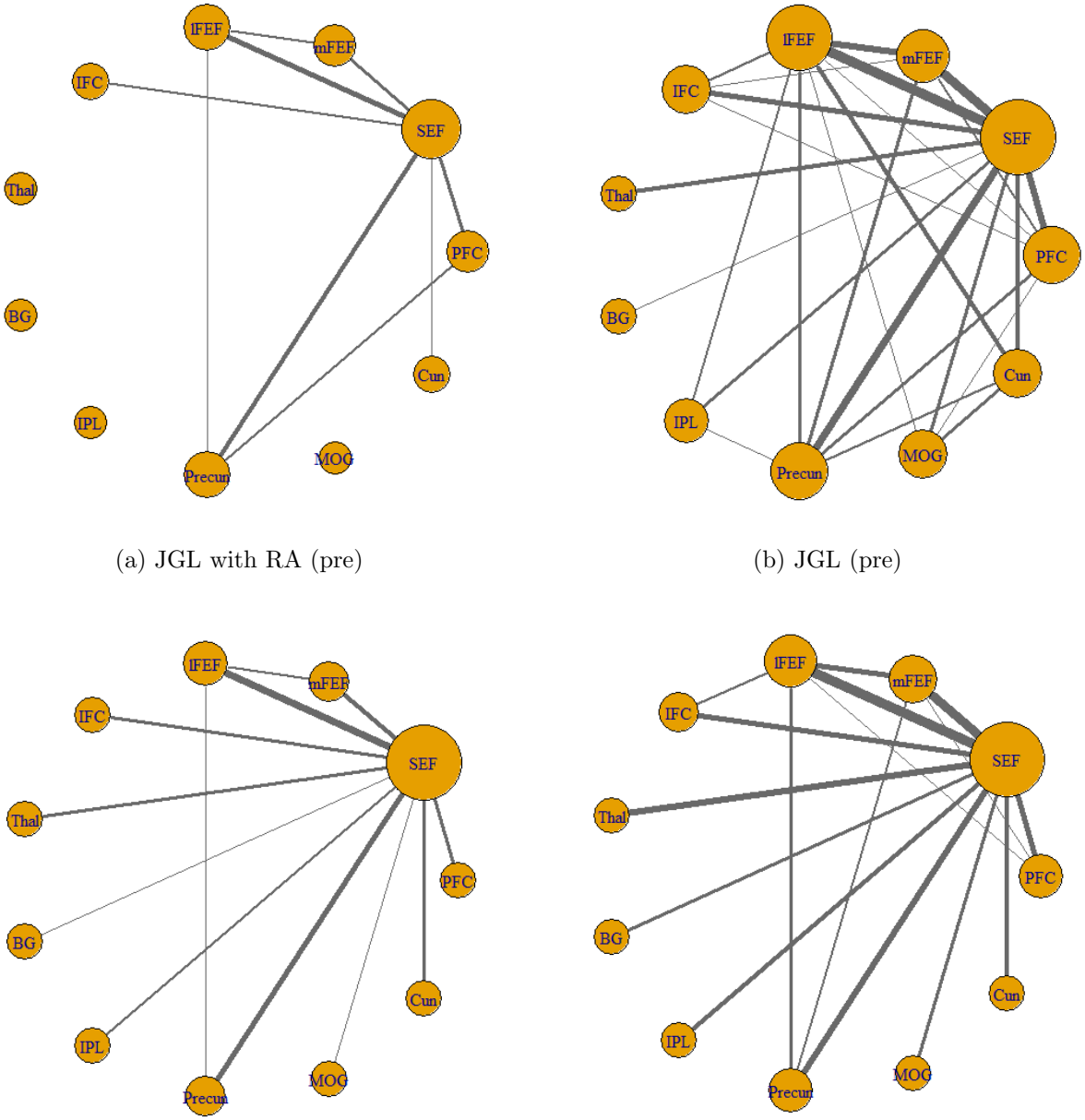
(c) JGL with RA (post)

(d) JGL (post)

Figure 5.13: Group graphs by JGL with RA and JGL for the 100% run. The edges weaker than 0.03 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

Table 5.9: Change in relationships after the practice by JGL with RA for the 25% run. The only regions which have absolute differences in $\hat{\omega}_{ij}$ greater than 0.02 between the pre- and post-test sessions are included. Especially, the regions with an asterisk (*) have absolute differences in $\hat{\omega}_{ij}$ greater than 0.06.

Stronger (Pre < Post)		Weaker (Pre > Post)	
PFC	-	PFC	Precun
SEF	mFEF, lFEF, IFC, Thal*, BG, IPL, Precun, MOG, Cun	SEF	-
mFEF	-	mFEF	-
lFEF	-	lFEF	-
IFC	-	IFC	-
Thal	-	Thal	-
BG	-	BG	-
IPL	-	IPL	-
Precun	-	Precun	-
MOG	-	MOG	-



(a) JGL with RA (pre)

(b) JGL (pre)

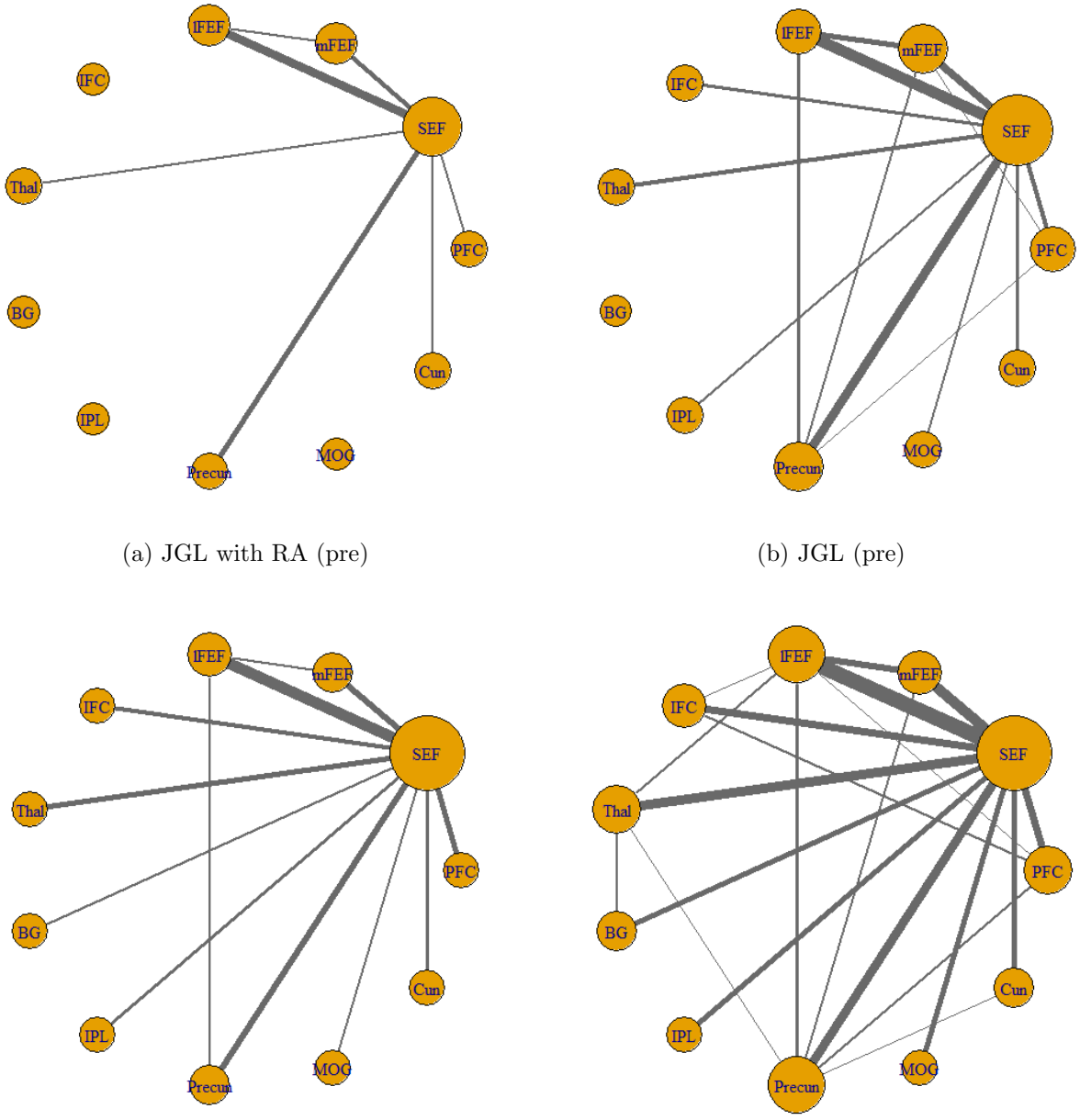
(c) JGL with RA (post)

(d) JGL (post)

Figure 5.14: Group graphs by JGL with RA and JGL for the 25% run. The edges weaker than 0.03 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

Table 5.10: Change in relationships after the practice by JGL with RA for the 50% run. The only regions which have absolute differences in $\hat{\omega}_{ij}$ greater than 0.02 between the pre- and post-test sessions are included. Especially, the regions with an asterisk (*) have absolute differences in $\hat{\omega}_{ij}$ greater than 0.06.

Stronger (Pre < Post)		Weaker (Pre > Post)	
PFC	SEF	PFC	-
SEF	mFEF, lFEF, IFC*, Thal*, BG, IPL*, Precun, MOG, Cun	SEF	-
mFEF	-	mFEF	-
lFEF	Precun	lFEF	-
IFC	-	IFC	-
Thal	-	Thal	-
BG	-	BG	-
IPL	-	IPL	-
Precun	-	Precun	-
MOG	-	MOG	-



(a) JGL with RA (pre)

(b) JGL (pre)

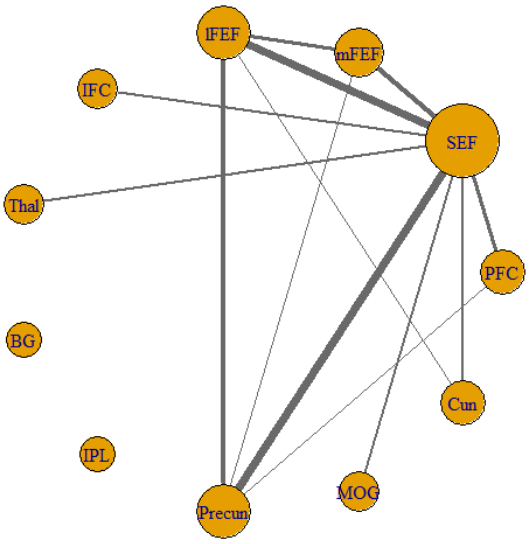
(c) JGL with RA (post)

(d) JGL (post)

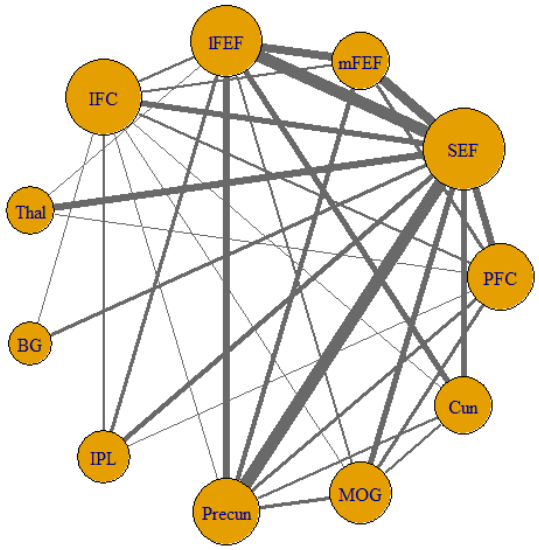
Figure 5.15: Group graphs by JGL with RA and JGL for the 50% run. The edges weaker than 0.03 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

Table 5.11: Change in relationships after the practice by JGL with RA for the 75% run. The only regions which have absolute differences in $\hat{\omega}_{ij}$ greater than 0.02 between the pre- and post-test sessions are included. Especially, the regions with an asterisk (*) have absolute differences in $\hat{\omega}_{ij}$ greater than 0.06.

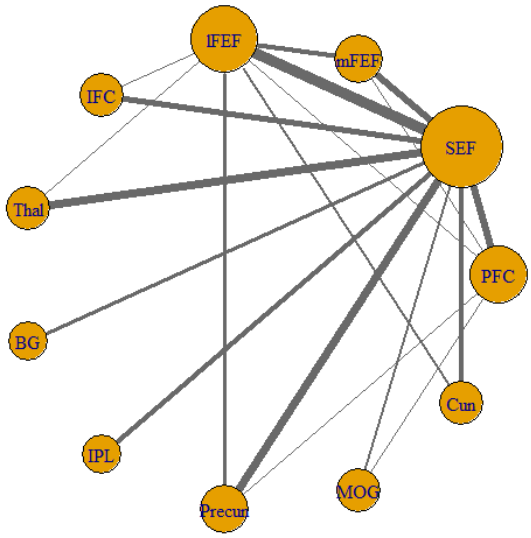
Stronger (Pre < Post)		Weaker (Pre > Post)	
PFC	SEF*, mFEF, lFEF, MOG	PFC	-
SEF	mFEF, lFEF, IFC*, Thal*, BG*, IPL*, Cun	SEF	-
mFEF	-	mFEF	Precun
lFEF	IFC, Thal	lFEF	-
IFC	-	IFC	-
Thal	-	Thal	-
BG	-	BG	-
IPL	-	IPL	-
Precun	-	Precun	-
MOG	-	MOG	-



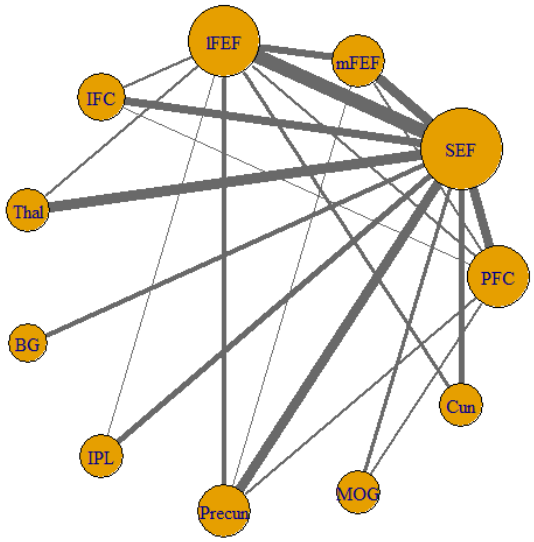
(a) JGL with RA (pre)



(b) JGL (pre)

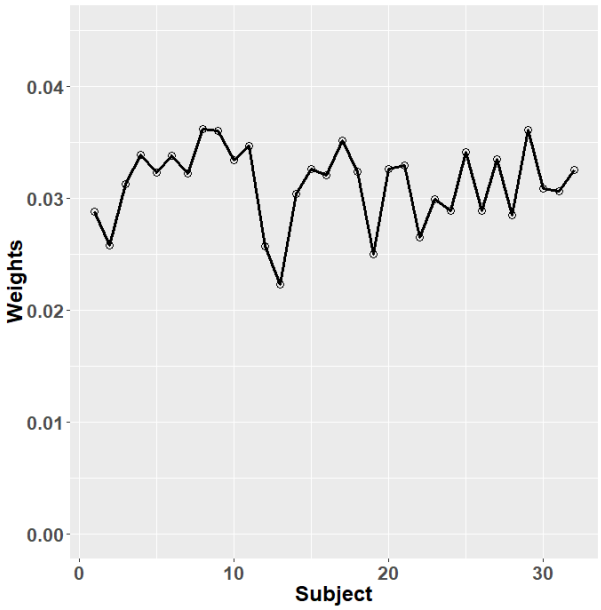


(c) JGL with RA (post)

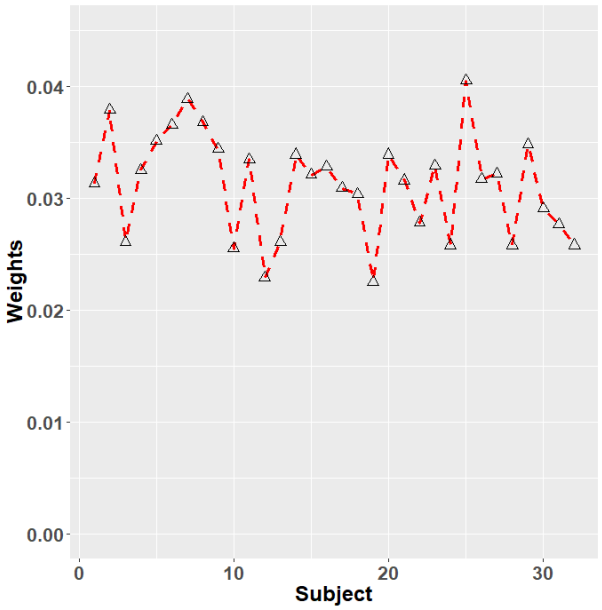


(d) JGL (post)

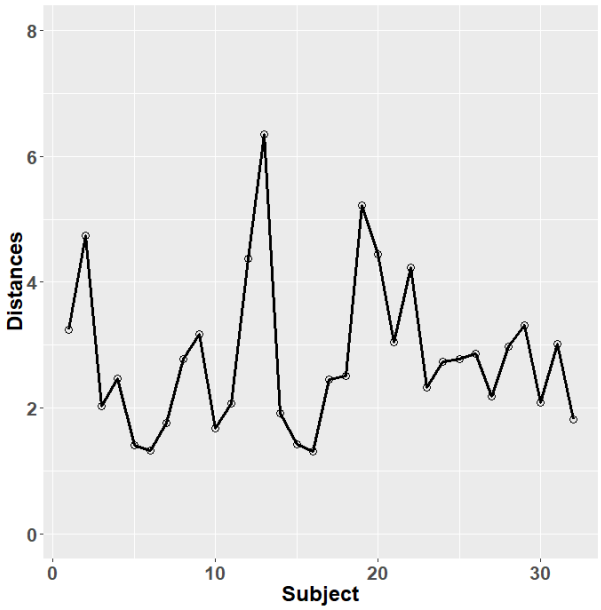
Figure 5.16: Group graphs by JGL with RA and JGL for the 75% run. The edges weaker than 0.03 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.



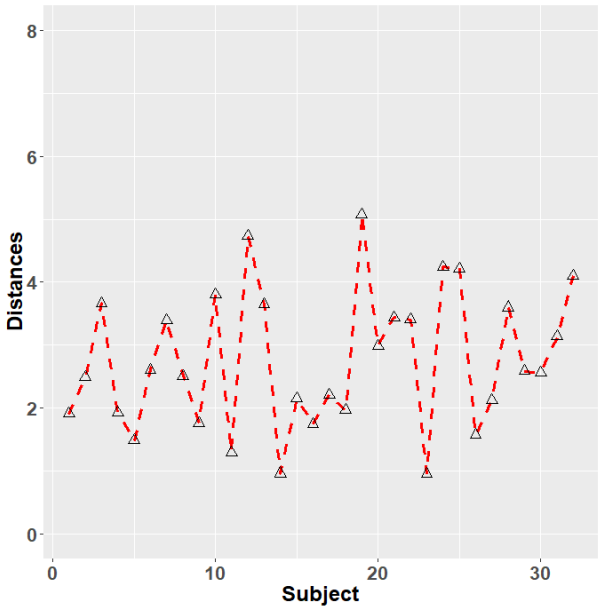
(a) Block Runs weights (Pre)



(b) Block Runs weights (Post)



(c) Block Runs distances (Pre)

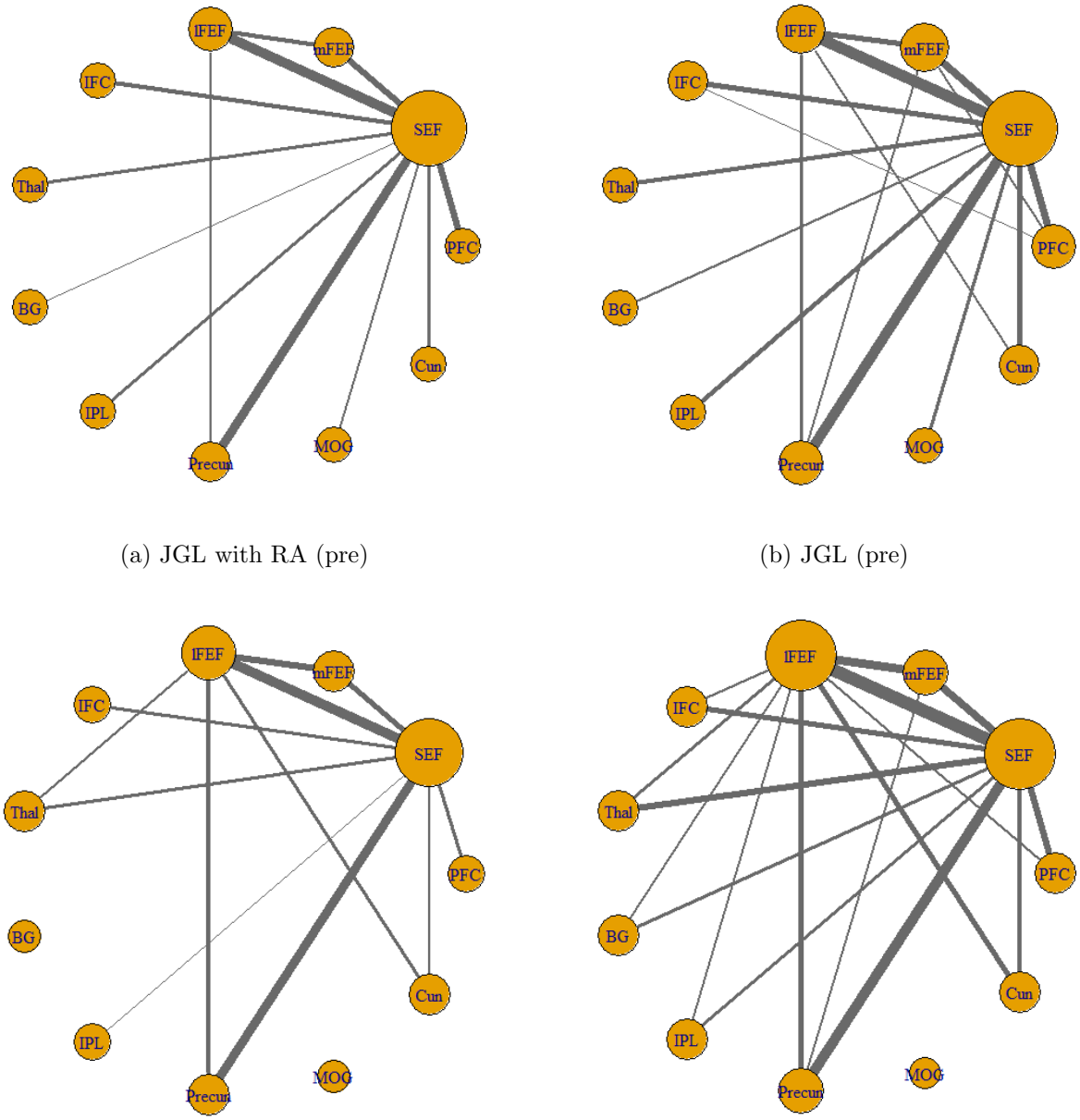


(d) Block Runs distances (Post)

Figure 5.17: Comparison of the estimated weights and distances by JGL with RA for the block runs in pre- and post-test sessions. The horizontal axe represents subject index, and the vertical axes are the estimated weights and the distances, respectively.

Table 5.12: Change in relationships after the practice by JGL with RA for the block runs. The only regions which have absolute differences in $\hat{\omega}_{ij}$ greater than 0.02 between the pre- and post-test sessions are included. Especially, the regions with an asterisk (*) have absolute differences in $\hat{\omega}_{ij}$ greater than 0.06.

Stronger (Pre < Post)		Weaker (Pre > Post)	
PFC	-	PFC	SEF
SEF	-	SEF	IFC, BG, IPL Precun, MOG, Cun
mFEF	lFEF	mFEF	-
lFEF	Thal, Precun, Cun*	lFEF	-
IFC	-	IFC	-
Thal	-	Thal	-
BG	-	BG	-
IPL	-	IPL	-
Precun	-	Precun	-
MOG	-	MOG	-



(a) JGL with RA (pre)

(b) JGL (pre)

(c) JGL with RA (post)

(d) JGL (post)

Figure 5.18: Group graphs by JGL with RA and JGL for the block runs. The edges weaker than 0.03 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

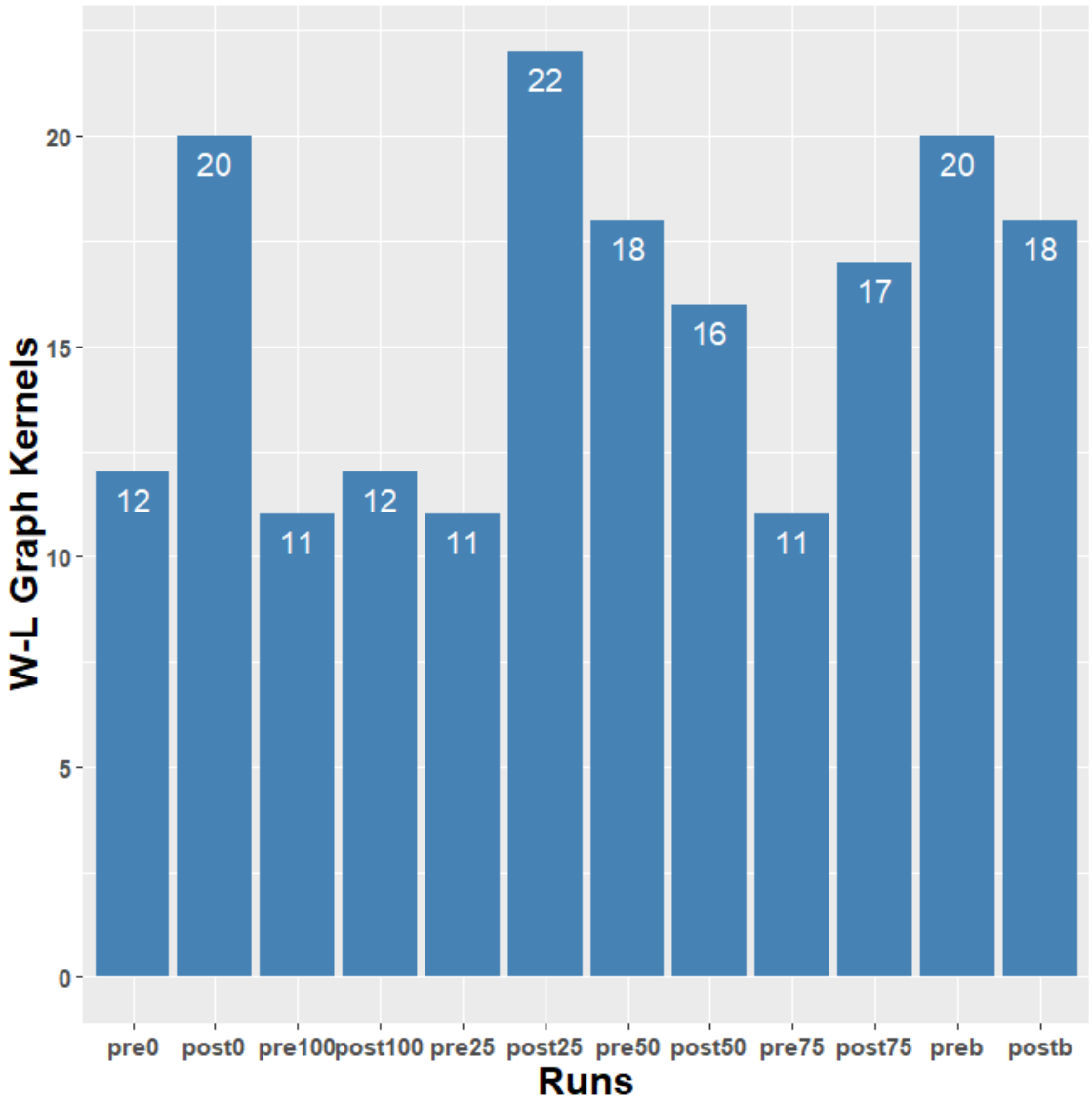


Figure 5.19: Comparison of JGL with RA and JGL group graphs based on Weisfeiler-Lehman graph kernels for each session and run. The horizontal and vertical axes are index of runs (e.g. "pre0" is for the 0% run in the pre-test session of the event-related expriments, and "postb" represents the post-test session of the block runs) and W-L graph kernel values respectively.

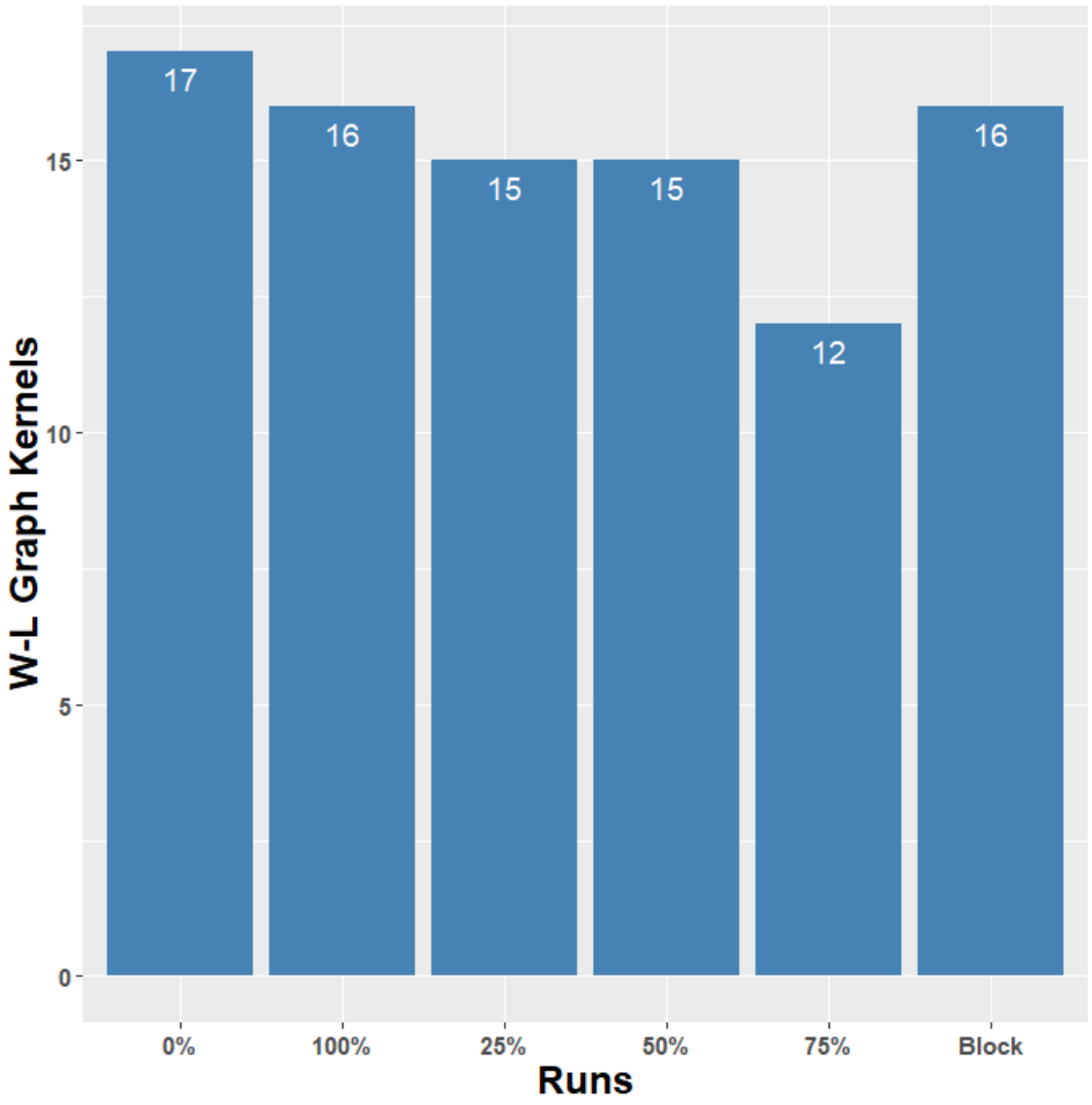


Figure 5.20: Comparison of JGL with RA group graphs for the pre- and post-test session based on Weisfeiler-Lehman graph kernels. The horizontal and vertical axes are index of runs and W-L graph kernel values respectively.

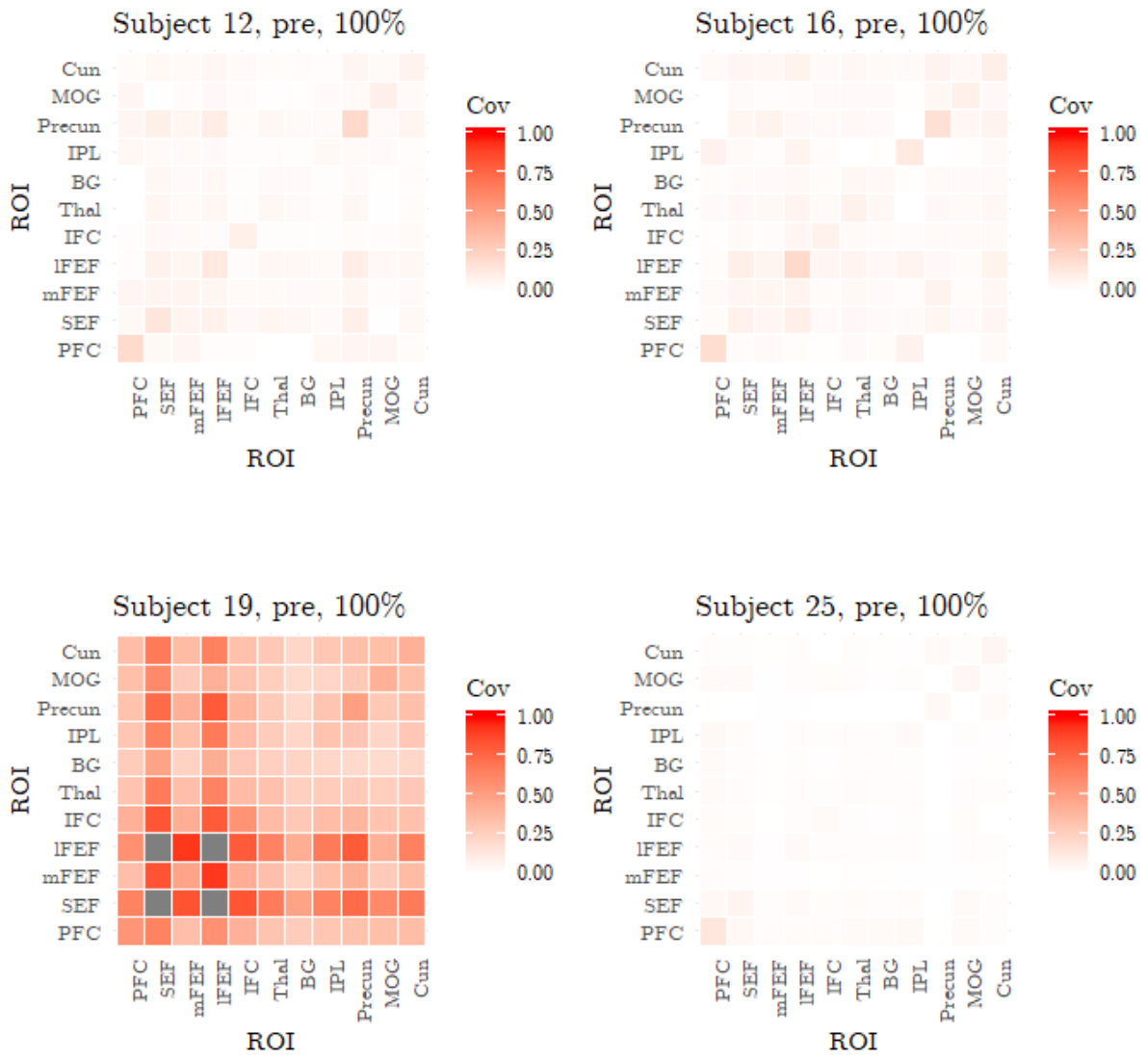


Figure 5.21: Covariance heat maps from the pre-test session in the 100% run for the event-related experiments. Covariance matrices from Subjects 12, 16, 19 and 25 are provided. The horizontal and vertical axes are ROI indices.

Chapter 6

Conclusion

Inference and visualization of brain circuitry using a graphical model is widely used in fMRI group analyses, and a state-of-the-art approach is to jointly estimate individual and group precision matrices. Aggregating subject-wise precision matrices is an essential step to yield generalized results about the entire group, and encounters a problem when not all subjects show a similar pattern in their brain networks. The proposed joint estimation of precision matrices with regularized aggregation method does not require the Gaussian distribution assumption, can identify outliers and provide a robust representative group graph. It assigns similar weights within a homogeneous cluster and detects outliers by allocating smaller weights to the possible outliers, as described in the simulation study.

We have also proposed the joint graphical lasso with regularized aggregation method to estimate and aggregate individual brain networks under the Gaussian distributional assumption. The proposed method can distinguish the effect of each subject on the group by assigning different weights. Likewise, the estimated weights are evenly distributed when the subjects are homogeneous, and the weights for the potential outliers are smaller than the majority of the group. This makes the proposed method valuable in practice, as it provides more robust representative group graphs by mitigating the effect induced by the outliers.

Both of the proposed JEMP with RA and JGL with RA methods focus on joint estimation; they estimate all subject's precision matrices simultaneously, so that they use all available information in data. The main objective of the two methods is robust estimation and aggregation which allows us to provide more generalized and interpretable results in the group-level analysis. JEMP with RA is particularly useful when the research goal is to estimate robust group precision matrix without the Gaussian assumption. It is free from the Gaussian distribution in the estimation procedure, and is robust to the outliers who provide sparse precision matrices. As we demonstrated in the simulation and real data studies, it captures more associations than JEMP. One limitation in the JEMP with RA comes from its graphical representation. We cannot directly infer undirected graphs from the estimated precision matrices, but we need the Gaussian assumption to create the graphs. On the other hand, the JGL with RA can directly provide undirected graphs from the estimated precision matrices. It can also detect outliers who have much deviated distances from roughly the median of the estimated inverse covariances in the group, by assigning lower weights. However, the method requires the Gaussian assumption in the estimation, which may be a strong assumption for high dimensional data, and it makes the estimation more sensitive to the standardization.

For the future work, we plan to apply different penalty functions in regularized aggregation (RA). The objective function in (2.13) in Section 2.5 in Chapter 2 can be generalized as Wang et al. (2017):

$$\begin{aligned} \min_{\boldsymbol{y},\boldsymbol{w}} L(\boldsymbol{y},\boldsymbol{w}) &= \min_{\boldsymbol{y},\boldsymbol{w}} \left\{ \|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{w}\|_2^2 + \lambda \|\boldsymbol{D}\boldsymbol{w}\|_\gamma^\gamma \right\}, \\ \text{s.t. } & \sum_{s=1}^{S} w^{(s)} = 1, \;\; w^{(s)} \geq 0; \;\; s = 1, \cdots, S, \end{aligned} \tag{6.1}$$

where we used $\boldsymbol{D} = \boldsymbol{I}$ and $\gamma = 2$. Instead, if we employ the first order pairwise difference matrix on $\boldsymbol{D}$ and set $\gamma = 1$, then (6.1) can be seen as the fused lasso aggregation (Wang

et al., 2017). We will examine how different penalty functions affect the performance of the proposed method.

We also plan to implement the estimation method that can control the temporal dependence structure of fMRI data in the estimation procedure. We employed a wavelet decomposition to reduce temporal correlation for the fMRI data that we used in the proposed JEMP with RA and JGL with RA methods, before the estimation process. However, if we can incorporate such temporal correlation in the estimation procedure, we might obtain more accurate outcome. We will explore these issues in future studies.

Bibliography

Afrasiabi, M. and Noroozian, N. (2015). Advantages and limitations of functional magnetic resonance imaging (fmri) of the human visual brain. *HORIZONS IN NEUROSCIENCE RESEARCH*, page 65.

Bandettini, P. A. and Cox, R. W. (2000). Event-related fmri contrast when using constant interstimulus interval: theory and experiment. *Magnetic Resonance in Medicine*, 43(4):540–548.

Banerjee, O., Ghaoui, L. E., and d'Aspremont, A. (2008). Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. *Journal of Machine learning research*, 9(Mar):485–516.

Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(Feb):281–305.

Biological Sciences Curriculum Study (2007). Information about mental illness and the brain. `https://www.ncbi.nlm.nih.gov/books/NBK20369/`.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. springer.

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122.

Cai, T., Liu, W., and Luo, X. (2011). A constrained $\ell 1$ minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607.

Cai, T. T., Li, H., Liu, W., and Xie, J. (2016). Joint estimation of multiple high-dimensional precision matrices. *Statistica Sinica*, 26(2):445.

Council, N. R. et al. (1992). *Combining information: Statistical issues and opportunities for research*, volume 1. National Academies.

Cribben, I., Haraldsdottir, R., Atlas, L. Y., Wager, T. D., and Lindquist, M. A. (2012). Dynamic connectivity regression: determining state-related changes in brain connectivity. *Neuroimage*, 61(4):907–920.

Dahl, J., Vandenberghe, L., and Roychowdhury, V. (2008). Covariance selection for non-chordal graphs via chordal embedding. *Optimization Methods & Software*, 23(4):501–520.

Danaher, P. (2018). *JGL: Performs the Joint Graphical Lasso for Sparse Inverse Covariance Estimation on Multiple Classes*. R package version 2.3.1.

Danaher, P., Wang, P., and Witten, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(2):373–397.

Fan, J., Feng, Y., and Wu, Y. (2009). Network exploration via the adaptive lasso and scad penalties. *The annals of applied statistics*, 3(2):521.

Fernandez-Casal, R. (2019). *npsp: Nonparametric Spatial Statistics*. R package version 0.7-5.

Filippi, M. (2016). *fMRI techniques and protocols*. Springer.

Flandrin, P. (1992). Wavelet analysis and synthesis of fractional brownian motion. *IEEE Transactions on information theory*, 38(2):910–917.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Friedman, J., Hastie, T., and Tibshirani, R. (2010a). Applications of the lasso and grouped lasso to the estimation of sparse graphical models. Technical report, Technical report, Stanford University.

Friedman, J., Hastie, T., and Tibshirani, R. (2010b). A note on the group lasso and a sparse group lasso. *arXiv preprint arXiv:1001.0736*.

Guo, J., Levina, E., Michailidis, G., and Zhu, J. (2011). Joint estimation of multiple graphical models. *Biometrika*, 98(1):1–15.

Hall, C. N., Howarth, C., Kurth-Nelson, Z., and Mishra, A. (2016). Interpreting bold: towards a dialogue between cognitive and cellular neuroscience.

Huettel, S. A., Song, A. W., McCarthy, G., et al. (2004). *Functional magnetic resonance imaging*, volume 1. Sinauer Associates Sunderland, MA.

Jordan, M. I. et al. (2004). Graphical models. *Statistical Science*, 19(1):140–155.

Kaplan, L. M. and Kuo, C.-C. (1993). Fractal estimation from noisy data via discrete fractional gaussian noise (dfgn) and the haar basis. *IEEE Transactions on Signal Processing*, 41(12):3554–3562.

Koller, D., Friedman, N., and Bach, F. (2009). *Probabilistic graphical models: principles and techniques*. MIT press.

Lazar, N. (2008). *The statistical analysis of functional MRI data*. Springer Science & Business Media.

Lazar, N. A., Eddy, W. F., Genovese, C. R., and Welling, J. (2001). Statistical issues in fmri for brain imaging. *International Statistical Review*, 69(1):105–127.

Lazar, N. A., Luna, B., Sweeney, J. A., and Eddy, W. F. (2002). Combining brains: a survey of methods for statistical pooling of information. *Neuroimage*, 16(2):538–550.

Lee, W. and Liu, Y. (2015). Joint estimation of multiple precision matrices with common structures. *The Journal of Machine Learning Research*, 16(1):1035–1062.

Li, B. and Solea, E. (2018). A nonparametric graphical model for functional data with application to brain networks based on fmri. *Journal of the American Statistical Association*, 113(524):1–19.

Li, S., Hsu, L., Peng, J., and Wang, P. (2013). Bootstrap inference for network construction with an application to a breast cancer microarray study. *The annals of applied statistics*, 7(1):391.

Li, T., Li, Q., Zhu, S., and Ogihara, M. (2002). A survey on wavelet applications in data mining. *ACM SIGKDD Explorations Newsletter*, 4(2):49–68.

Lindquist, M. A. et al. (2008). The statistical analysis of fmri data. *Statistical science*, 23(4):439–464.

McNamee, R. L. and Lazar, N. A. (2004). Assessing the sensitivity of fmri group maps. *Neuroimage*, 22(2):920–931.

Meinshausen, N. and Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4):417–473.

Meinshausen, N., Bühlmann, P., et al. (2006). High-dimensional graphs and variable selection with the lasso. *The annals of statistics*, 34(3):1436–1462.

Mitterschiffthaler, M. T., Ettinger, U., Mehta, M. A., Mataix-Cols, D., and Williams, S. C. (2006). Applications of functional magnetic resonance imaging in psychiatry. *Journal of Magnetic Resonance Imaging*, 23(6):851–861.

Munoz-Cespedes, J. M., Rios-Lago, M., Paul, N., and Maestu, F. (2005). Functional neuroimaging studies of cognitive recovery after acquired brain damage in adults. *Neuropsychology review*, 15(4):169–183.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

Nason, G. (2010). *Wavelet methods in statistics with R*. Springer Science & Business Media.

Nason, G. (2016). *wavethresh: Wavelets Statistics and Transforms*. R package version 4.6.8.

Ng, B., Varoquaux, G., Poline, J. B., and Thirion, B. (2013). A novel sparse group gaussian graphical model for functional connectivity estimation. In *International Conference on Information Processing in Medical Imaging*, pages 256–267. Springer.

Ogawa, S., Lee, T.-M., Kay, A. R., and Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, 87(24):9868–9872.

Pang, H., Liu, H., and Vanderbei, R. (2014). The fastclime package for linear programming and large-scale precision matrix estimation in r. *The Journal of Machine Learning Research*, 15(1):489–493.

Pauling, L. and Coryell, C. D. (1936). The magnetic properties and structure of hemoglobin, oxyhemoglobin and carbonmonoxyhemoglobin. *Proceedings of the National Academy of Sciences*, 22(4):210–216.

Pierce, J. E. and McDowell, J. E. (2017). Reduced cognitive control demands after practice of saccade tasks in a trial type probability manipulation. *Journal of cognitive neuroscience*, 29(2):368–381.

Shervashidze, N., Schweitzer, P., Leeuwen, E. J. v., Mehlhorn, K., and Borgwardt, K. M. (2011). Weisfeiler-lehman graph kernels. *Journal of Machine Learning Research*, 12(Sep):2539–2561.

Stoev, S., Taqqu, M. S., Park, C., and Marron, J. (2005). On the wavelet spectrum diagnostic for hurst parameter estimation in the analysis of internet traffic. *Computer Networks*, 48(3):423–445.

Sugiyama, M. (2018). *graphkernels: Graph Kernels*. R package version 1.6.

Sugiyama, M., Ghisu, M. E., Llinares-López, F., and Borgwardt, K. (2017). graphkernels: R and python packages for graph comparison. *Bioinformatics*, 34(3):530–532.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):91–108.

Turlach, B. A. and Weingessel, A. (2013). quadprog: Functions to solve quadratic programming problems. r package version 1.5-5.

Turner, R. (2016). Uses, misuses, new uses and fundamental limitations of magnetic resonance imaging in cognitive science. *Phil. Trans. R. Soc. B*, 371(1705):20150349.

Vishwanathan, S. V. N., Schraudolph, N. N., Kondor, R., and Borgwardt, K. M. (2010). Graph kernels. *Journal of Machine Learning Research*, 11(Apr):1201–1242.

Wainwright, M. J., Jordan, M. I., et al. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305.

Wand, M. (1994). Fast computation of multivariate kernel estimators. *Journal of Computational and Graphical Statistics*, 3(4):433–445.

Wang, L.-Y., Chung, J., Park, C., Choi, H., Rodrigue, A. L., Pierce, J. E., Clementz, B. A., and McDowell, J. E. (2019). Regularized aggregation of statistical parametric maps. *Human brain mapping*, 40(1):65–79.

Wang, L.-Y., Park, C., Yeon, K., and Choi, H. (2017). Tracking concept drift using a constrained penalized regression combiner. *Computational Statistics & Data Analysis*, 108(Apr):52–69.

Witten, D. M. and Tibshirani, R. (2009). Covariance-regularized regression and classification for high dimensional problems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3):615–636.

Yuan, M. and Lin, Y. (2007). Model selection and estimation in the gaussian graphical model. *Biometrika*, 94(1):19–35.

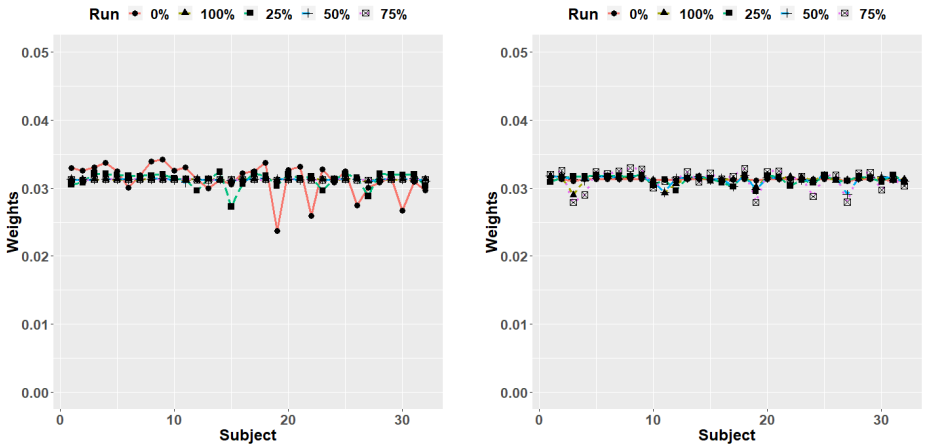
Zhu, Y. and Li, L. (2018). Multiple matrix gaussian graphs estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(5):927–950.

Appendix A

Results from the standardized covariances

A.1 JEMP with RA

A.1.1 Event-related Experiments

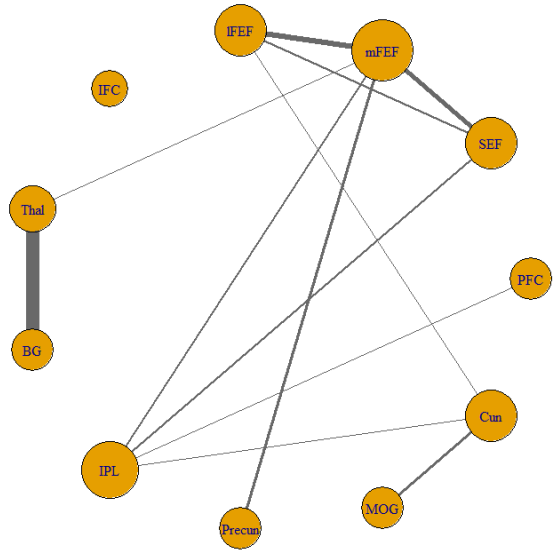


(a) Event experiments (Pre) (b) Event experiments (Post)

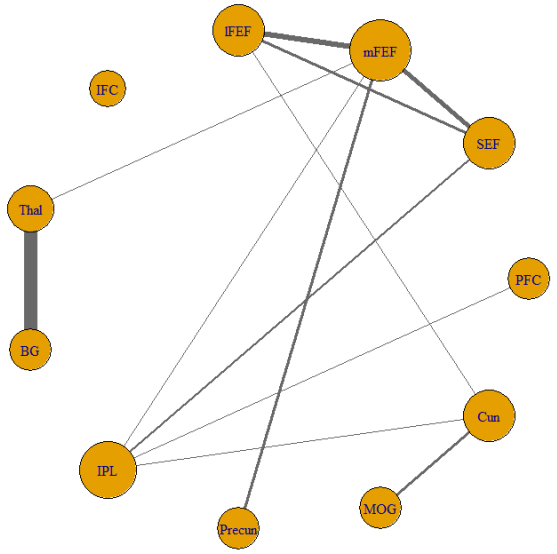
Figure A.1: Comparison of the estimated weights by JEMP with RA from two sessions for the 5 probability runs in each panel. The horizontal and vertical axes are subject index and the estimated weights, respectively.

Table A.1: Change in relationships after the practice by JEMP with RA for the 0% run. The only regions which have absolute differences in $\hat{\omega}_{ij}$ greater than 0.2 between the pre- and post-test sessions are included. Especially, the regions with an asterisk (*) have absolute differences in $\hat{\omega}_{ij}$ greater than 0.4.

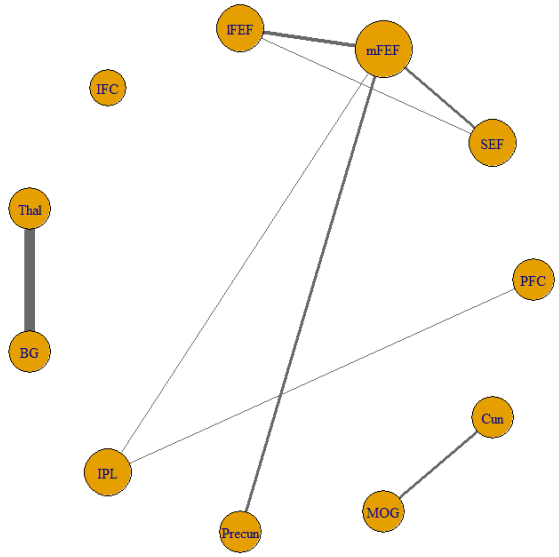
Stronger (Pre < Post)		Weaker (Pre > Post)	
PFC	-	PFC	-
SEF	-	SEF	mFEF, lFEF, IPL*
mFEF	-	mFEF	lFEF, Thal
lFEF	-	lFEF	Cun
IFC	-	IFC	-
Thal	-	Thal	BG*
BG	-	BG	-
IPL	-	IPL	Cun
Precun	-	Precun	-
MOG	-	MOG	-



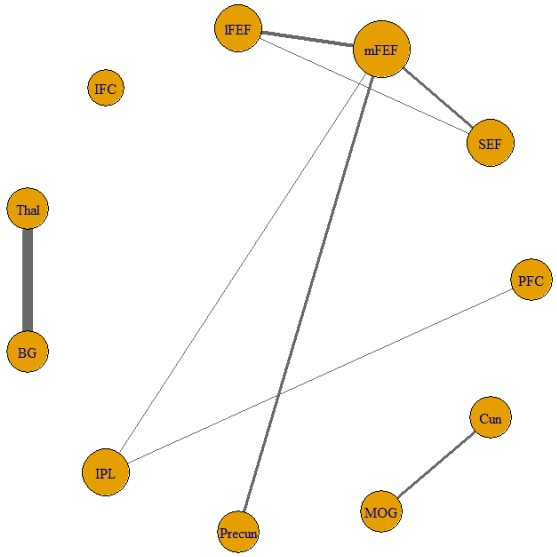
(a) JEMP with RA (pre)



(b) JEMP (pre)



(c) JEMP with RA (post)

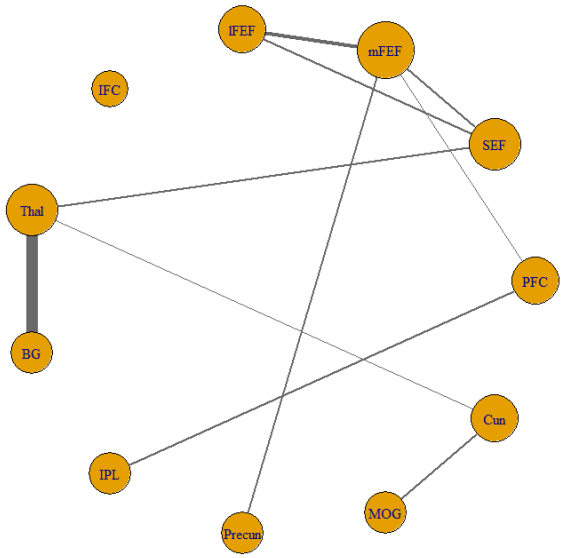


(d) JEMP (post)

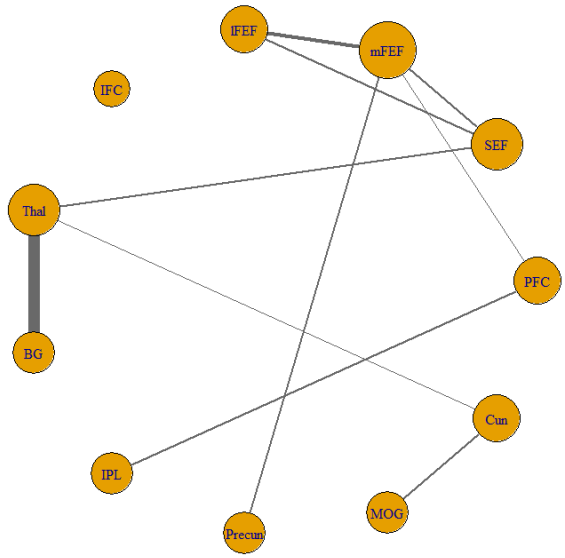
Figure A.2: Group graphs by JEMP with RA and JEMP for the 0% run. The edges weaker than 0.3 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

Table A.2: Change in relationships after the practice by JEMP with RA for the 100% run. The only regions which have absolute differences in $\hat{\omega}_{ij}$ greater than 0.2 between the pre- and post-test sessions are included. Especially, the regions with an asterisk (*) have absolute differences in $\hat{\omega}_{ij}$ greater than 0.4.

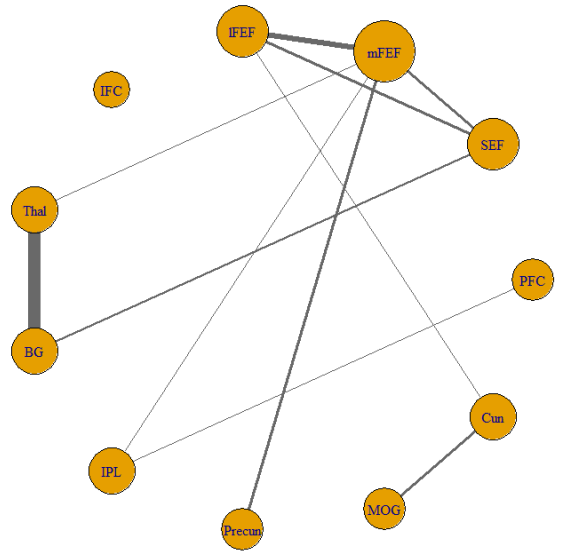
Stronger (Pre < Post)		Weaker (Pre > Post)	
PFC	-	PFC	mFEF
SEF	BG*	SEF	Thal*
mFEF	lFEF, Thal, IPL	mFEF	-
lFEF	Cun	lFEF	-
IFC	-	IFC	-
Thal	BG	Thal	Cun
BG	-	BG	-
IPL	-	IPL	-
Precun	-	Precun	-
MOG	Cun	MOG	-



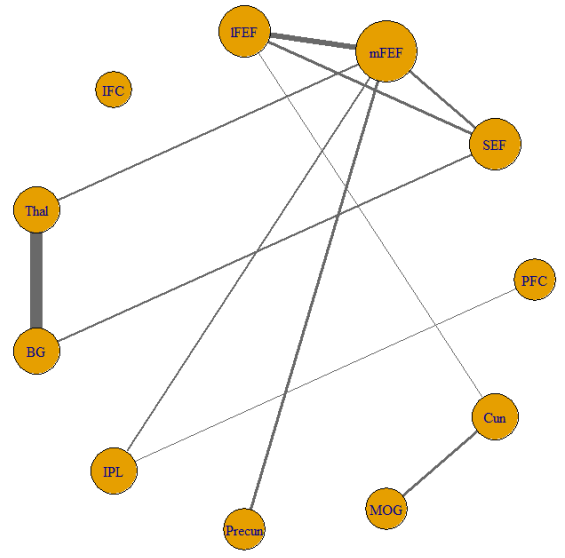
(a) JEMP with RA (pre)



(b) JEMP (pre)



(c) JEMP with RA (post)

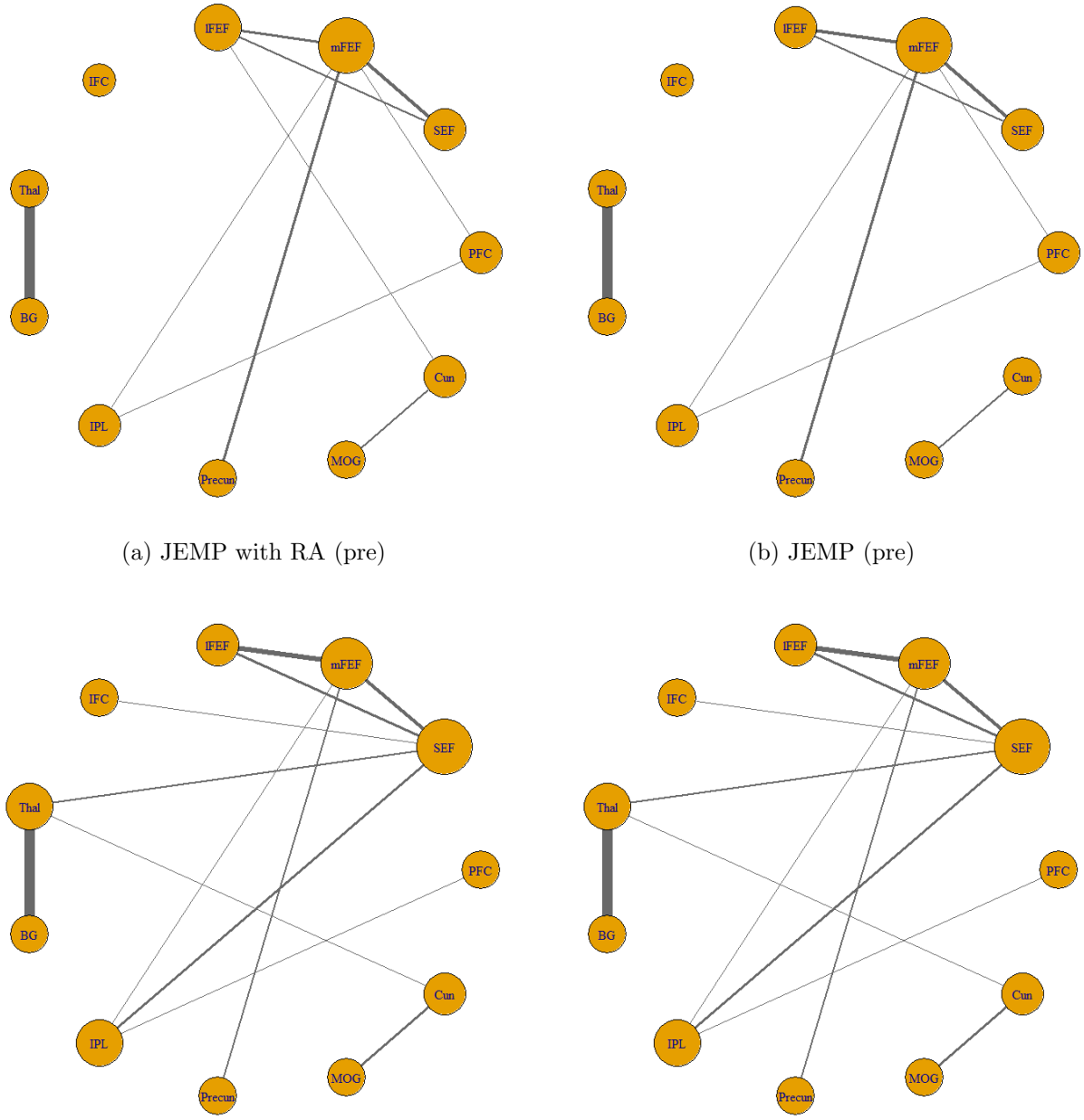


(d) JEMP (post)

Figure A.3: Group graphs by JEMP with RA and JEMP for the 100% run. The edges weaker than 0.3 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

Table A.3: Change in relationships after the practice by JEMP with RA for the 25% run. The only regions which have absolute differences in $\hat{\omega}_{ij}$ greater than 0.2 between the pre- and post-test sessions are included. Especially, the regions with an asterisk (*) have absolute differences in $\hat{\omega}_{ij}$ greater than 0.4.

Stronger (Pre < Post)		Weaker (Pre > Post)	
PFC	-	PFC	mFEF
SEF	IFC, Thal*, IPL*	SEF	-
mFEF	lFEF*	mFEF	-
lFEF		lFEF	Cun
IFC	-	IFC	-
Thal	Cun	Thal	-
BG	-	BG	-
IPL	-	IPL	-
Precun	-	Precun	-
MOG	Cun	MOG	-



(a) JEMP with RA (pre)

(b) JEMP (pre)

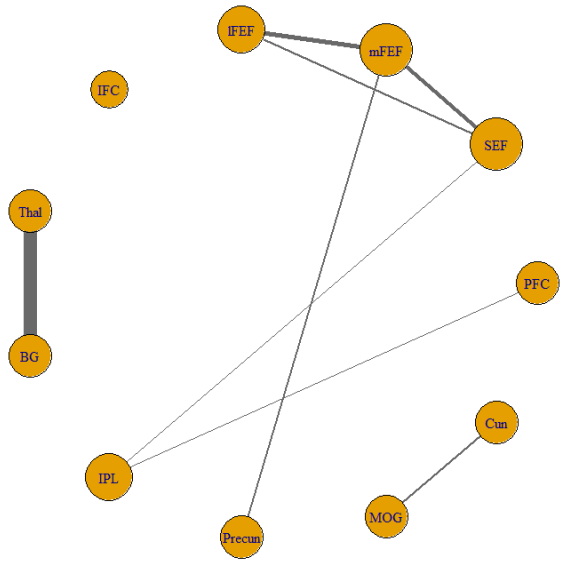
(c) JEMP with RA (post)

(d) JEMP (post)

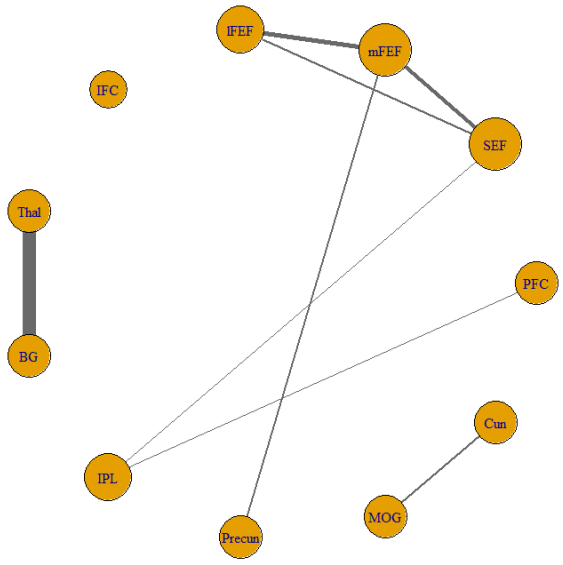
Figure A.4: Group graphs by JEMP with RA and JEMP for the 25% run. The edges weaker than 0.3 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

Table A.4: Change in relationships after the practice by JEMP with RA for the 50% run. The only regions which have absolute differences in $\hat{\omega}_{ij}$ greater than 0.2 between the pre- and post-test sessions are included. Especially, the regions with an asterisk (*) have absolute differences in $\hat{\omega}_{ij}$ greater than 0.4.

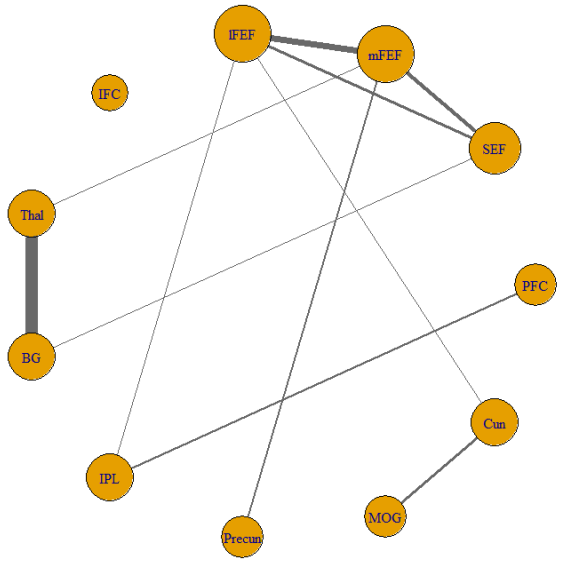
Stronger (Pre < Post)		Weaker (Pre > Post)	
PFC	-	PFC	-
SEF	lFEF, BG	SEF	IPL
mFEF	lFEF, Thal	mFEF	-
lFEF	IPL, Cun	lFEF	-
IFC	-	IFC	-
Thal	-	Thal	-
BG	-	BG	-
IPL	-	IPL	-
Precun	-	Precun	-
MOG	Cun	MOG	-



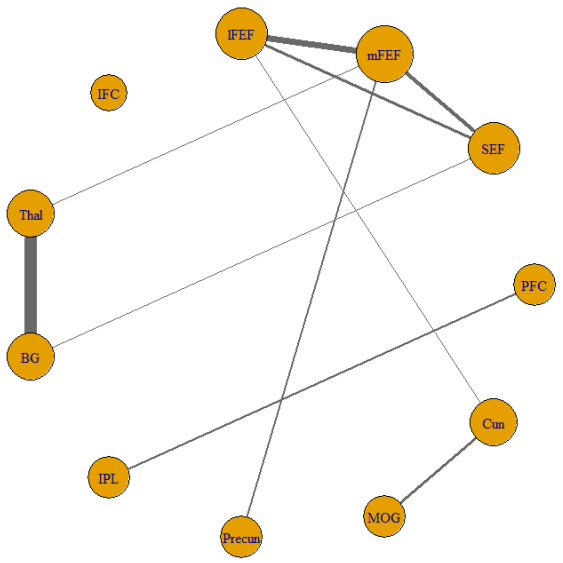
(a) JEMP with RA (pre)



(b) JEMP (pre)



(c) JEMP with RA (post)

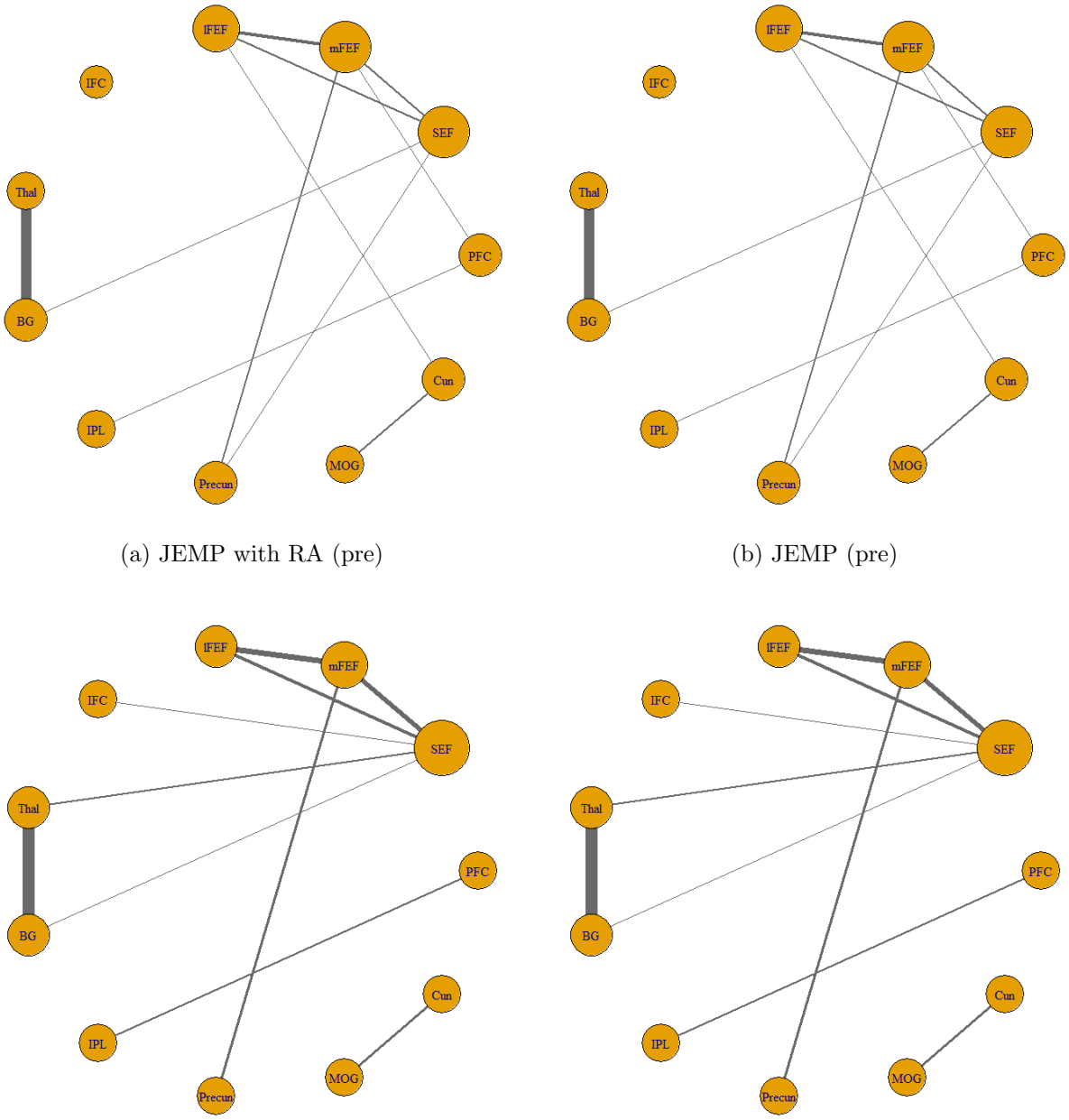


(d) JEMP (post)

Figure A.5: Group graphs by JEMP with RA and JEMP for the 50% run. The edges weaker than 0.3 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

Table A.5: Change in relationships after the practice by JEMP with RA for the 75% run. The only regions which have absolute differences in $\hat{\omega}_{ij}$ greater than 0.2 between the pre- and post-test sessions are included. Especially, the regions with an asterisk (*) have absolute differences in $\hat{\omega}_{ij}$ greater than 0.4.

Stronger (Pre < Post)		Weaker (Pre > Post)	
PFC	-	PFC	mFEF
SEF	mFEF*, lFEF*, IFC, Thal*	SEF	Precun
mFEF	lFEF*	mFEF	-
lFEF	-	lFEF	Cun
IFC	-	IFC	-
Thal	BG	Thal	-
BG	-	BG	-
IPL	-	IPL	-
Precun	-	Precun	-
MOG	Cun	MOG	-



(a) JEMP with RA (pre)

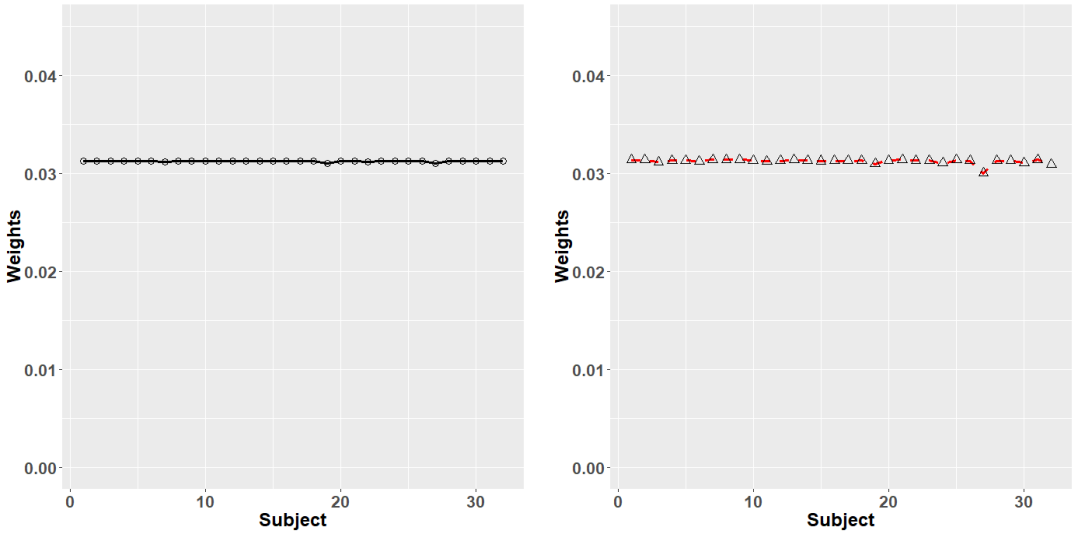
(b) JEMP (pre)

(c) JEMP with RA (post)

(d) JEMP (post)

Figure A.6: Group graphs by JEMP with RA and JEMP for the 75% run. The edges weaker than 0.3 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

A.1.2 Block Runs

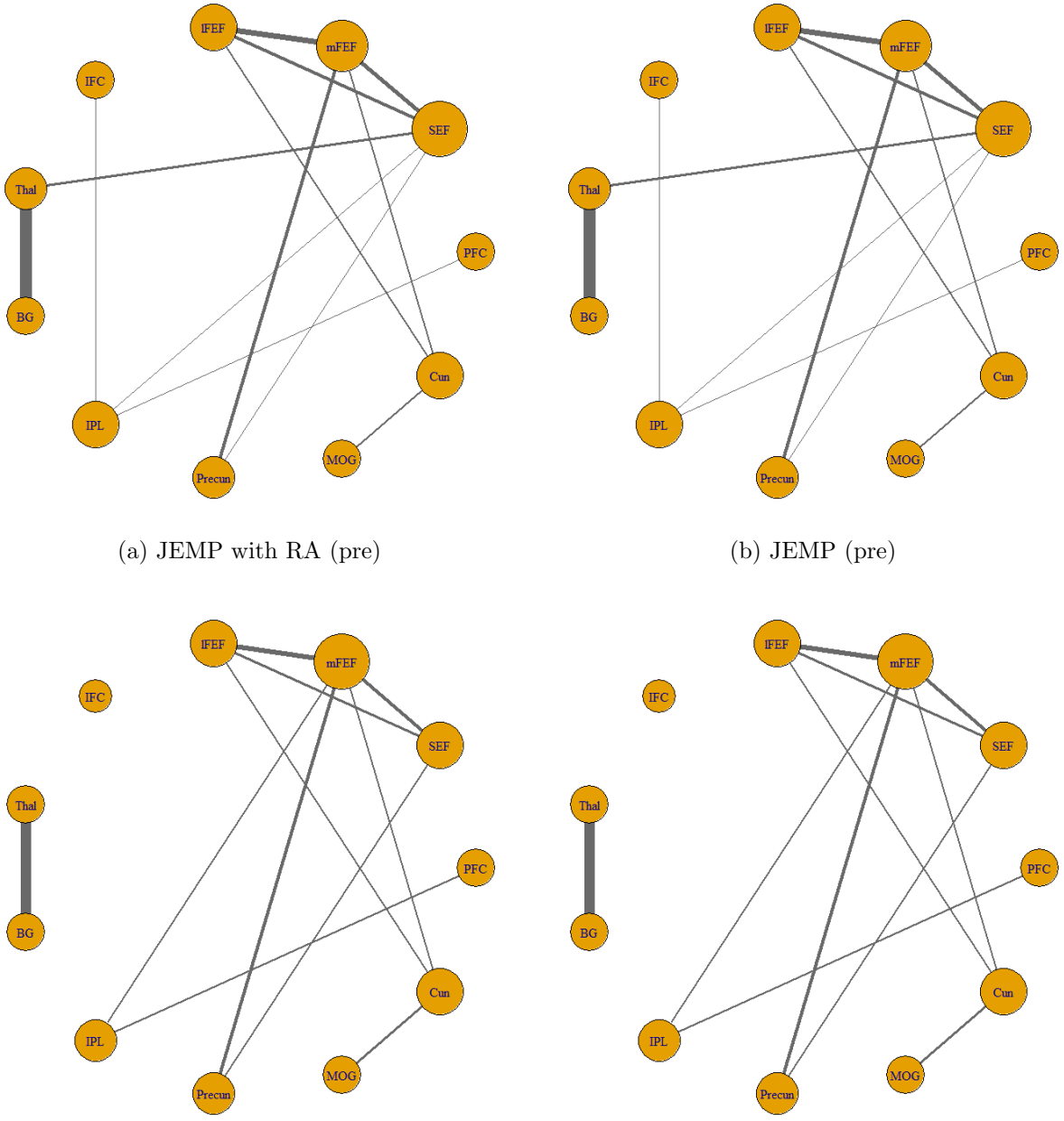


(a) Block runs (Pre) (b) Block runs (Post)

Figure A.7: Comparison of the estimated weights by JEMP with RA for the block runs in pre- and post-test sessions. In each panel, the horizontal and vertical axes are subject index and the estimated weights respectively.

Table A.6: Change in relationships after the practice by JEMP with RA for the block runs. The only regions which have absolute differences in $\hat{\omega}_{ij}$ greater than 0.2 between the pre- and post-test sessions are included. Especially, the regions with an asterisk (*) have absolute differences in $\hat{\omega}_{ij}$ greater than 0.4.

Stronger (Pre < Post)		Weaker (Pre > Post)	
PFC	-	PFC	-
SEF	-	SEF	Thal*, IPL
mFEF	IPL*	mFEF	lFEF
lFEF	-	lFEF	-
IFC	-	IFC	IPL
Thal	-	Thal	BG
BG	-	BG	-
IPL	-	IPL	-
Precun	-	Precun	-
MOG	-	MOG	-



(a) JEMP with RA (pre)

(b) JEMP (pre)

(c) JEMP with RA (post)

(d) JEMP (post)

Figure A.8: Group graphs by JEMP with RA and JEMP for the block runs. The edges weaker than 0.3 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

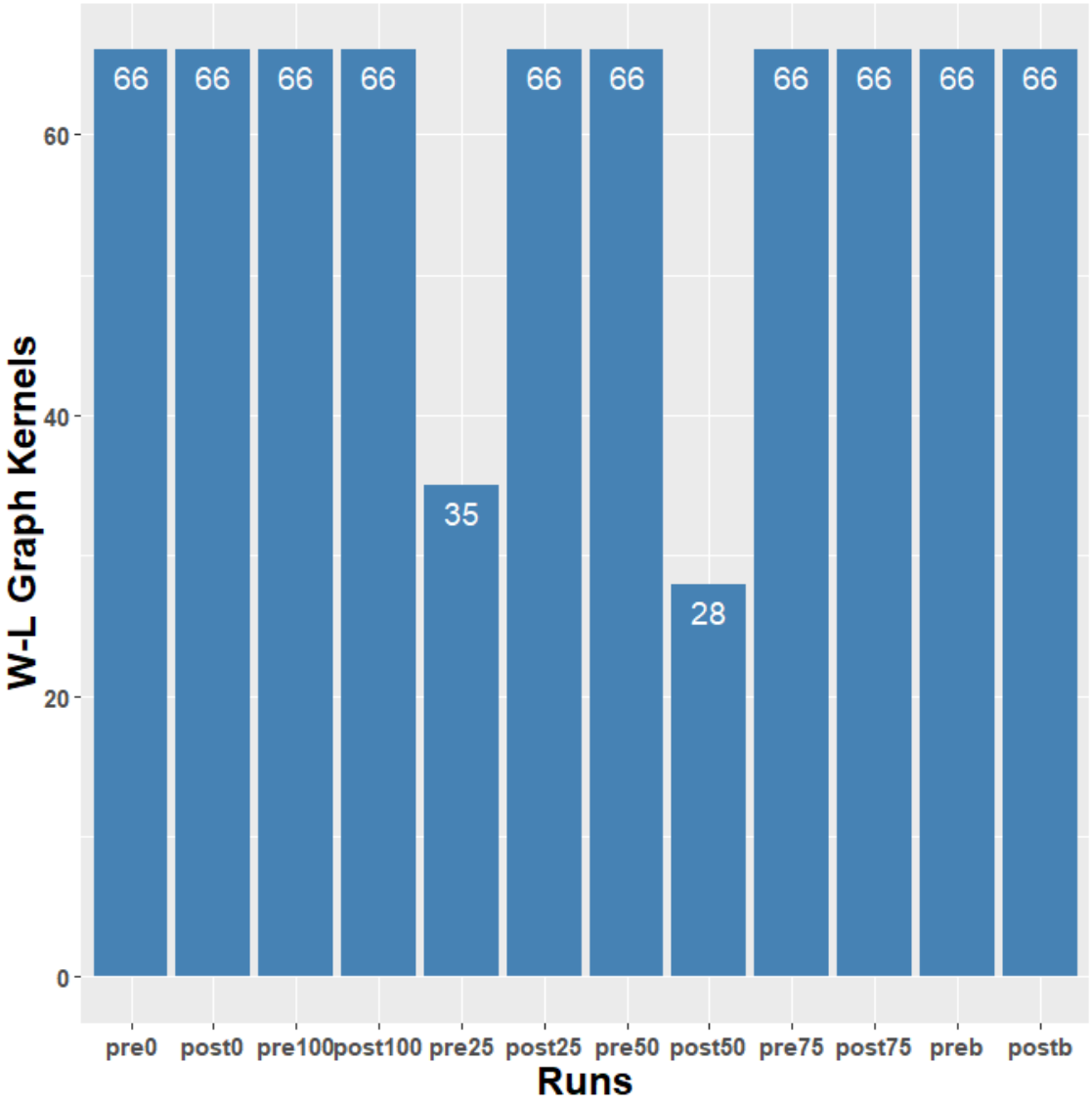


Figure A.9: Comparison of JEMP with RA and JEMP group graphs based on Weisfeiler-Lehman graph kernels for each session and run. The horizontal and vertical axes are index of runs (e.g. "pre0" is for the 0% run in the pre-test session of the event-related expriments, and "postb" represents the post-test session of the block runs) and W-L graph kernel values respectively.

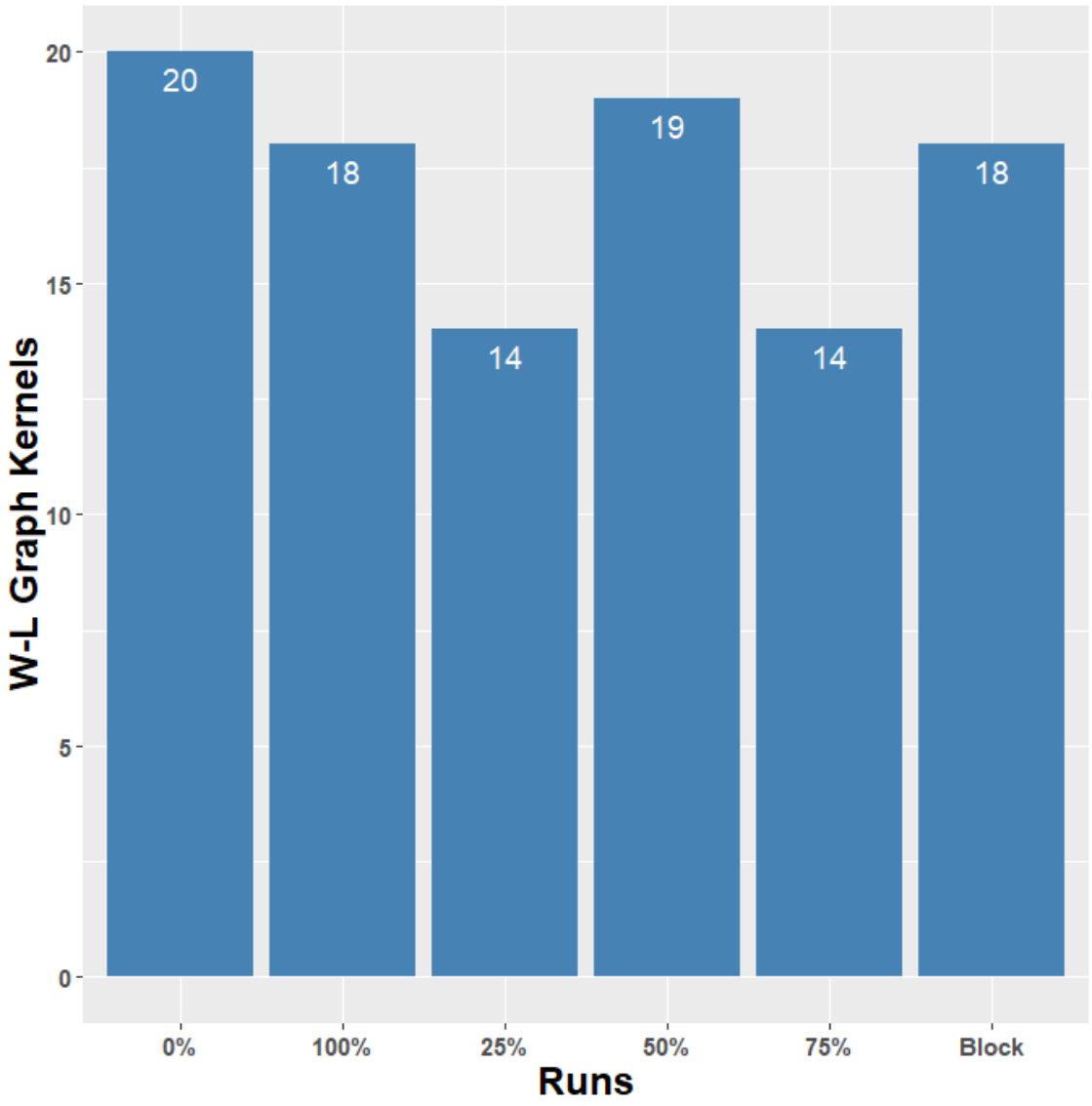
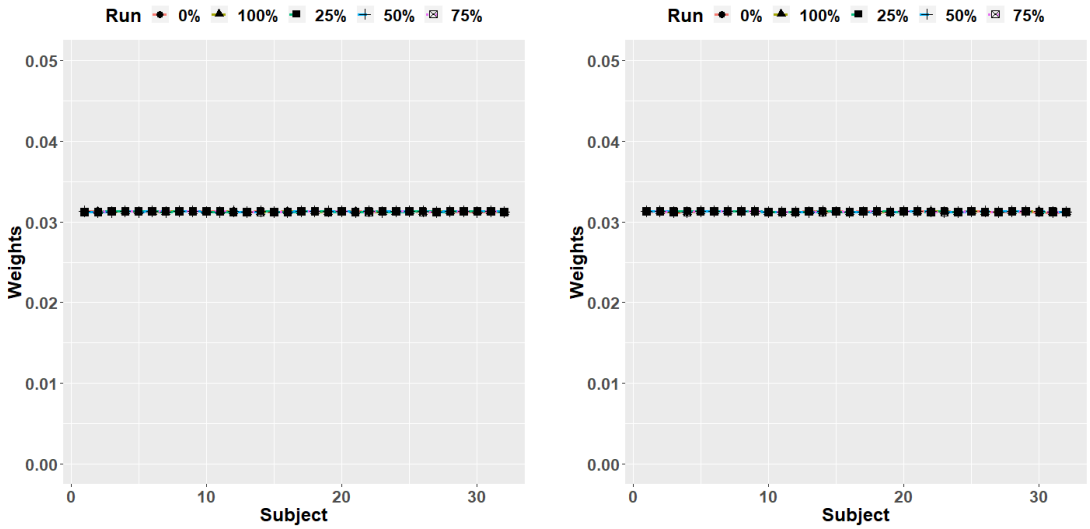


Figure A.10: Comparison of JEMP with RA group graphs for the pre- and post-test session based on Weisfeiler-Lehman graph kernels. The horizontal and vertical axes are index of runs and W-L graph kernel values respectively.

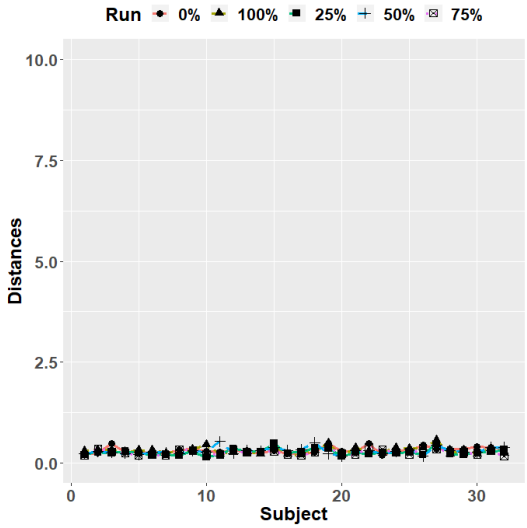
A.2 JGL with RA

A.2.1 Event-related Experiments

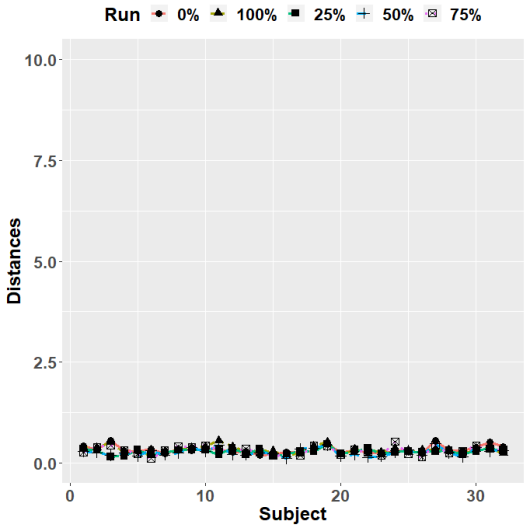


(a) Event experiments weights (Pre)

(b) Event experiments weights (Post)



(c) Event experiments distances (Pre)

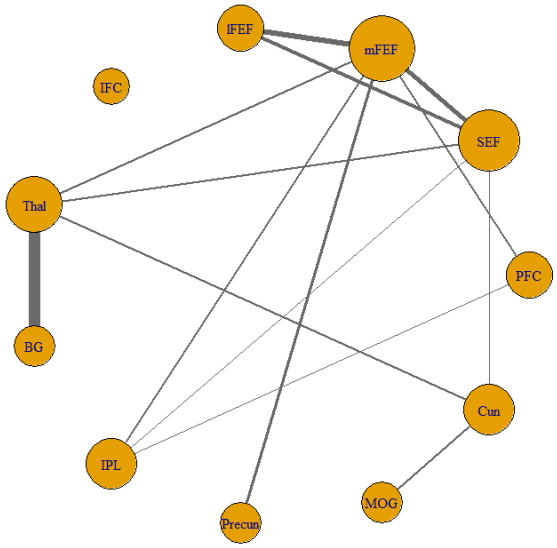


(d) Event experiments distances (Post)

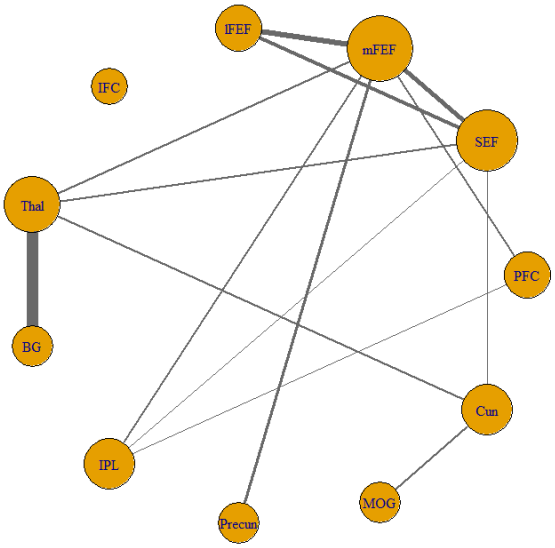
Figure A.11: Comparison of the estimated weights and distances by JGL with RA from two sessions for the 5 probability runs in each panel. The horizontal axe represents subject index, and the vertical axes are the estimated weights and the distances, respectively.

Table A.7: Change in relationships after the practice by JGL with RA for the 0% run. The only regions which have absolute differences in $\hat{\omega}_{ij}$ greater than 0.02 between the pre- and post-test sessions are included. Especially, the regions with an asterisk (*) have absolute differences in $\hat{\omega}_{ij}$ greater than 0.04.

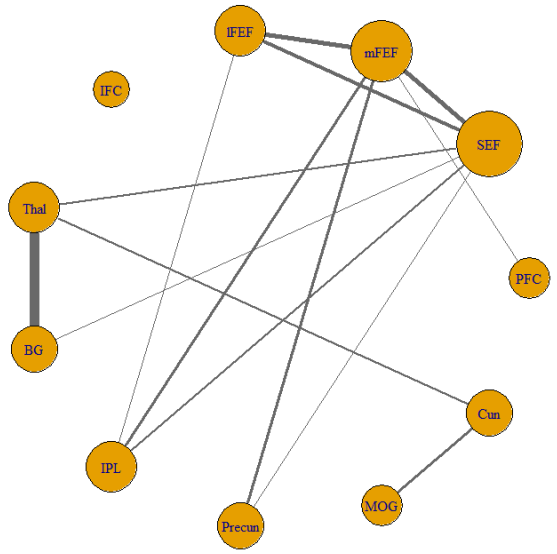
Stronger (Pre < Post)		Weaker (Pre > Post)	
PFC	-	PFC	IPL
SEF	BG, Precun	SEF	Cun
mFEF		mFEF	Thal*
lFEF	IPL	lFEF	-
IFC	-	IFC	-
Thal	-	Thal	BG
BG	-	BG	-
IPL	-	IPL	-
Precun	-	Precun	-
MOG	-	MOG	-



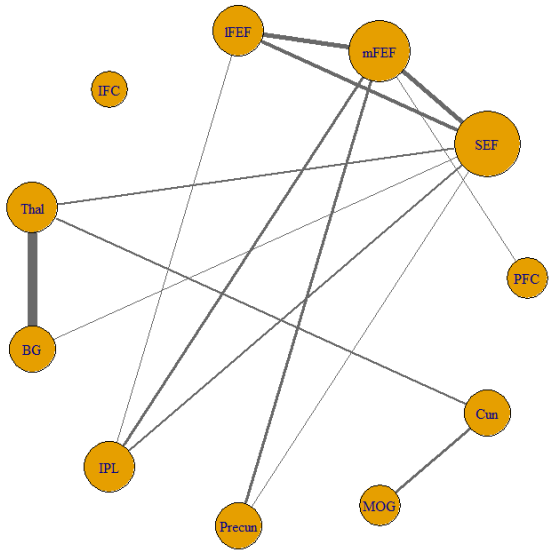
(a) JGL with RA (pre)



(b) JGL (pre)



(c) JGL with RA (post)

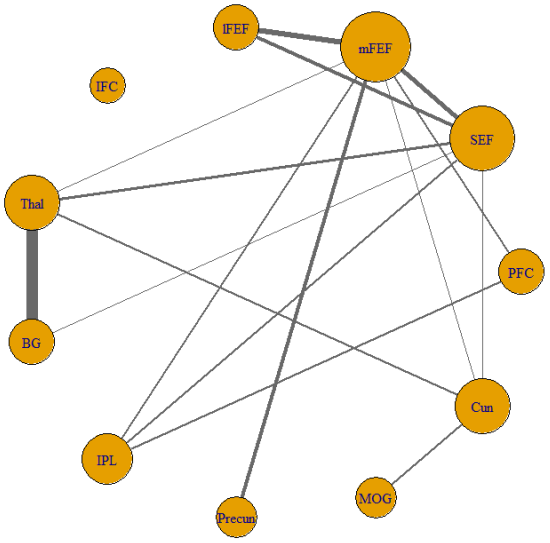


(d) JGL (post)

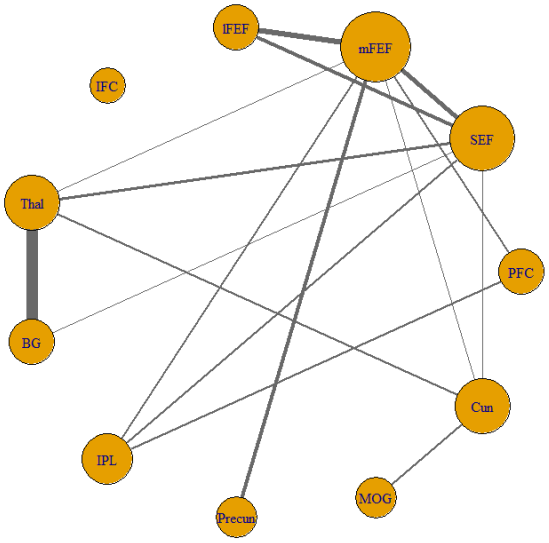
Figure A.12: Group graphs by JGL with RA and JGL for the 0% run. The edges weaker than 0.03 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

Table A.8: Change in relationships after the practice by JGL with RA for the 100% run. The only regions which have absolute differences in $\hat{\omega}_{ij}$ greater than 0.02 between the pre- and post-test sessions are included. Especially, the regions with an asterisk (*) have absolute differences in $\hat{\omega}_{ij}$ greater than 0.04.

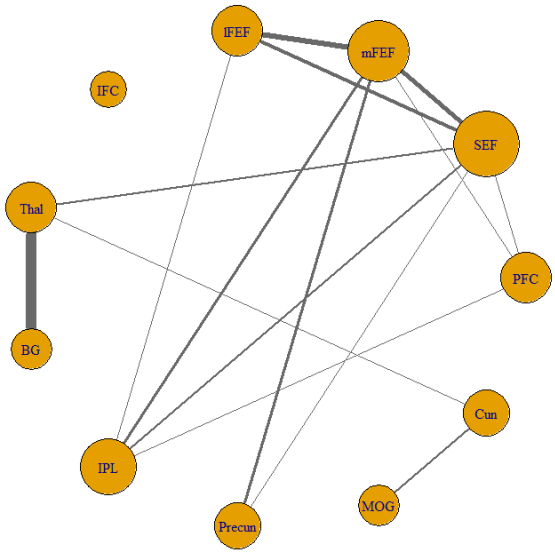
Stronger (Pre < Post)		Weaker (Pre > Post)	
PFC	SEF	PFC	IPL
SEF	Precun	SEF	BG, Cun
mFEF	-	mFEF	Thal, Cun
lFEF	IPL	lFEF	-
IFC	-	IFC	-
Thal	-	Thal	BG
BG	-	BG	-
IPL	-	IPL	-
Precun	-	Precun	-
MOG	-	MOG	-



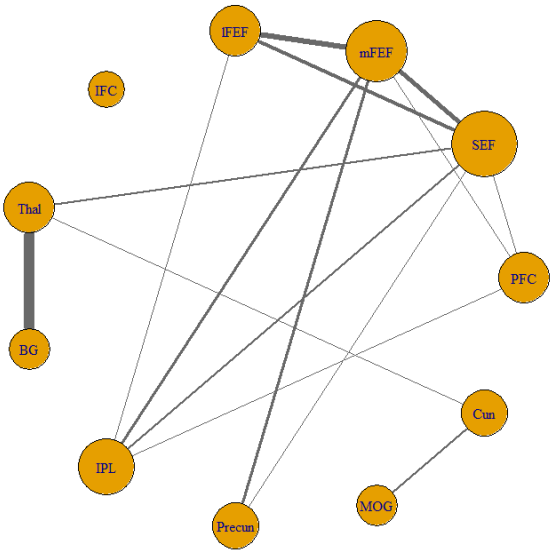
(a) JGL with RA (pre)



(b) JGL (pre)



(c) JGL with RA (post)

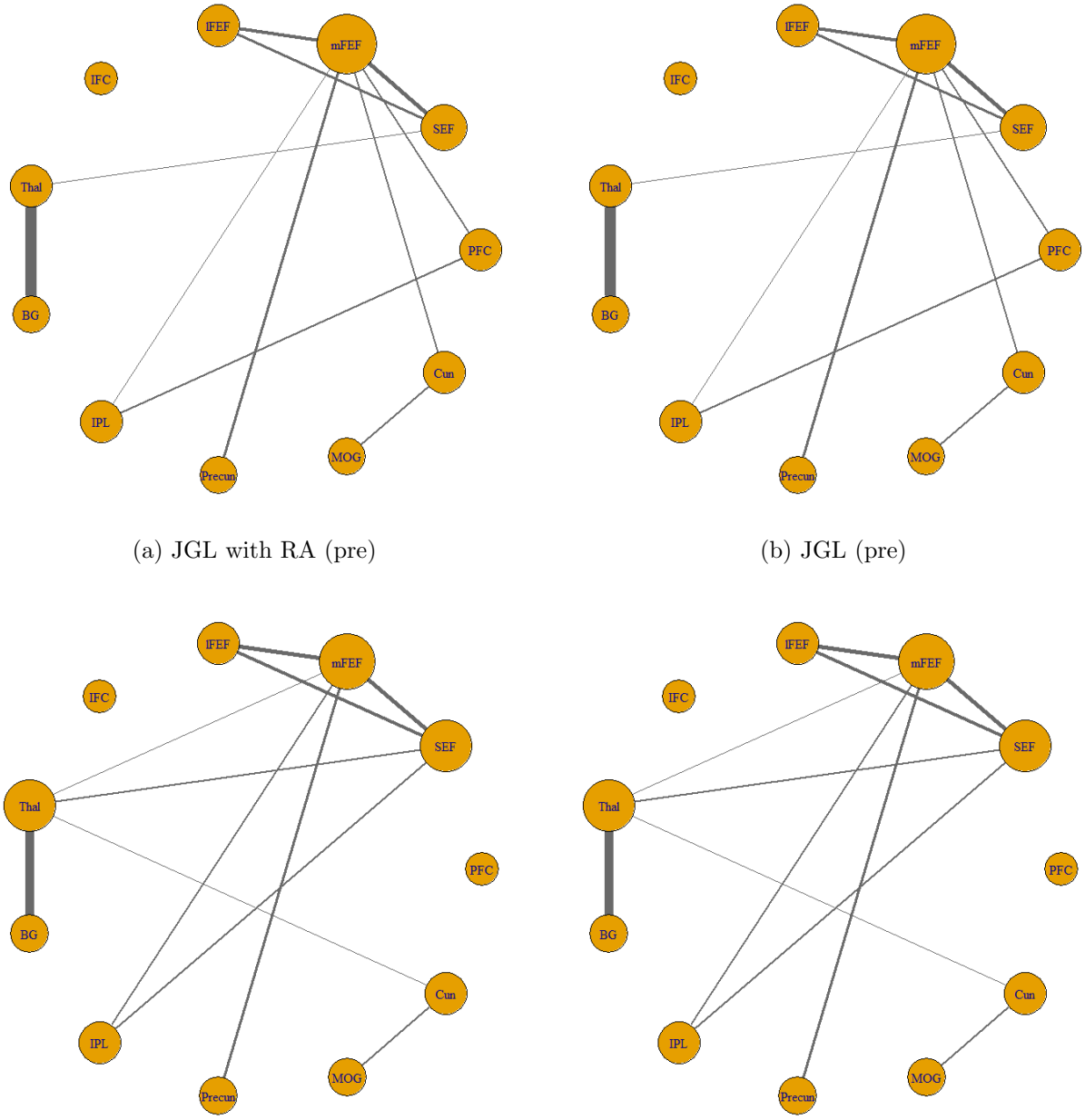


(d) JGL (post)

Figure A.13: Group graphs by JGL with RA and JGL for the 100% run. The edges weaker than 0.03 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

Table A.9: Change in relationships after the practice by JGL with RA for the 25% run. The only regions which have absolute differences in $\hat{\omega}_{ij}$ greater than 0.02 between the pre- and post-test sessions are included. Especially, the regions with an asterisk (*) have absolute differences in $\hat{\omega}_{ij}$ greater than 0.04.

Stronger (Pre < Post)		Weaker (Pre > Post)	
PFC	-	PFC	mFEF*, IPL*
SEF	IPL*	SEF	-
mFEF	Thal	mFEF	Cun*
lFEF	-	lFEF	-
IFC	-	IFC	-
Thal	Cun	Thal	BG*
BG	-	BG	-
IPL	-	IPL	-
Precun	-	Precun	-
MOG	-	MOG	-



(a) JGL with RA (pre)

(b) JGL (pre)

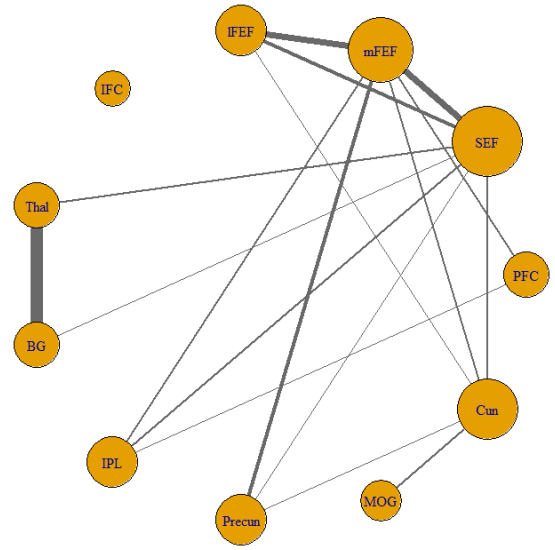
(c) JGL with RA (post)

(d) JGL (post)

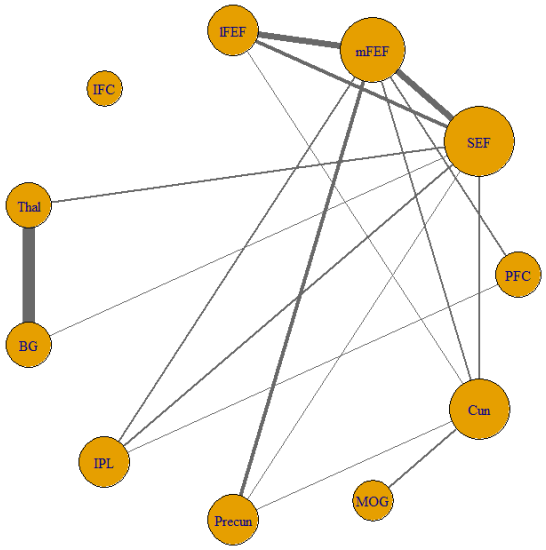
Figure A.14: Group graphs by JGL with RA and JGL for the 25% run. The edges weaker than 0.03 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

Table A.10: Change in relationships after the practice by JGL with RA for the 50% run. The only regions which have absolute differences in $\hat{\omega}_{ij}$ greater than 0.02 between the pre- and post-test sessions are included. Especially, the regions with an asterisk (*) have absolute differences in $\hat{\omega}_{ij}$ greater than 0.04.

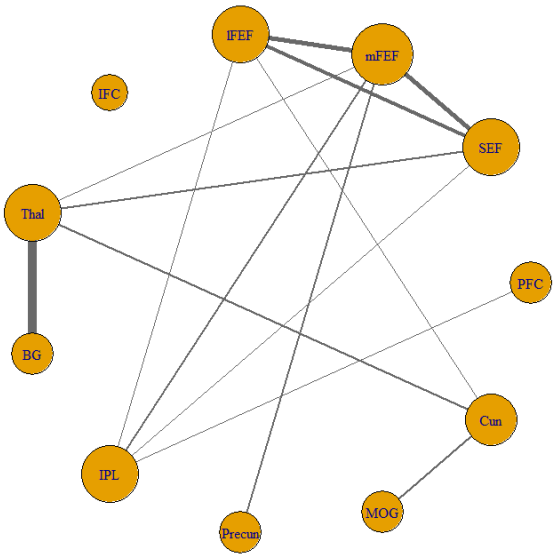
Stronger (Pre < Post)		Weaker (Pre > Post)	
PFC	-	PFC	mFEF*
SEF	-	SEF	mFEF, BG, Precun, Cun*
mFEF	Thal	mFEF	lFEF, Precun, Cun*
lFEF	IPL	lFEF	-
IFC	-	IFC	-
Thal	Cun*	Thal	BG*
BG	-	BG	-
IPL	-	IPL	-
Precun	-	Precun	-
MOG	-	MOG	-



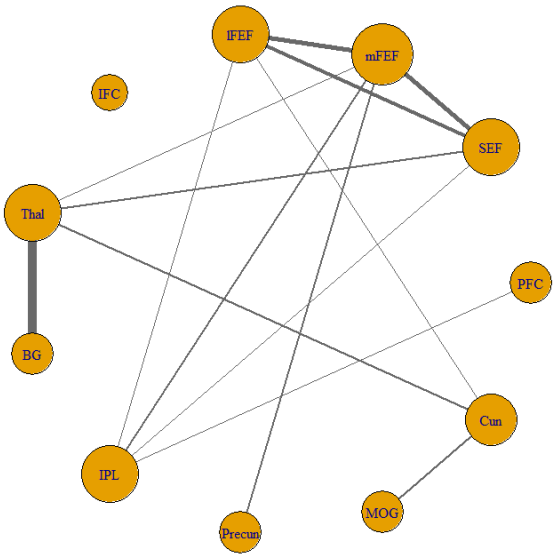
(a) JGL with RA (pre)



(b) JGL (pre)



(c) JGL with RA (post)

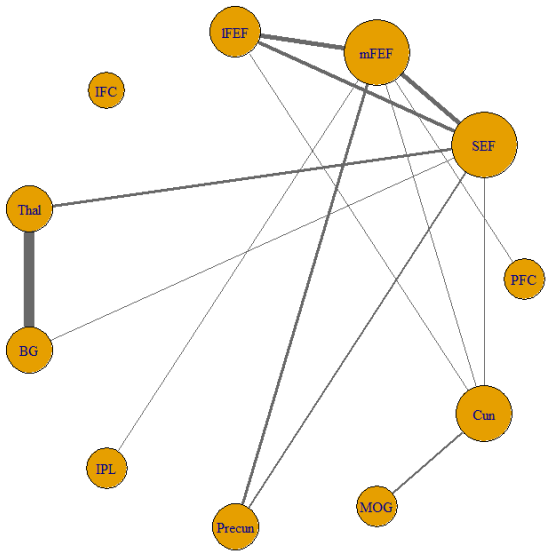


(d) JGL (post)

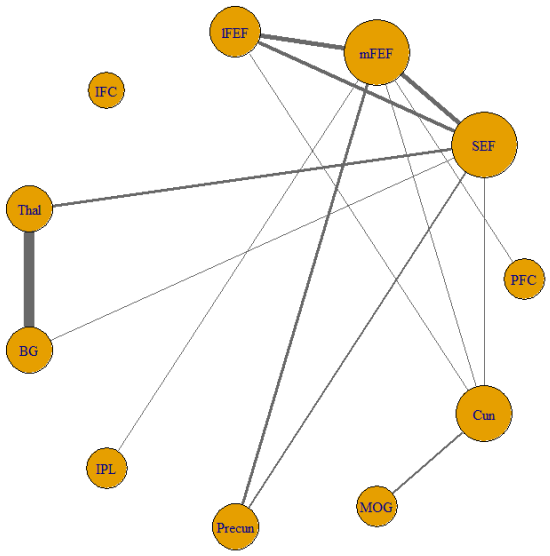
Figure A.15: Group graphs by JGL with RA and JGL for the 50% run. The edges weaker than 0.03 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

Table A.11: Change in relationships after the practice by JGL with RA for the 75% run. The only regions which have absolute differences in $\hat{\omega}_{ij}$ greater than 0.02 between the pre- and post-test sessions are included. Especially, the regions with an asterisk (*) have absolute differences in $\hat{\omega}_{ij}$ greater than 0.04.

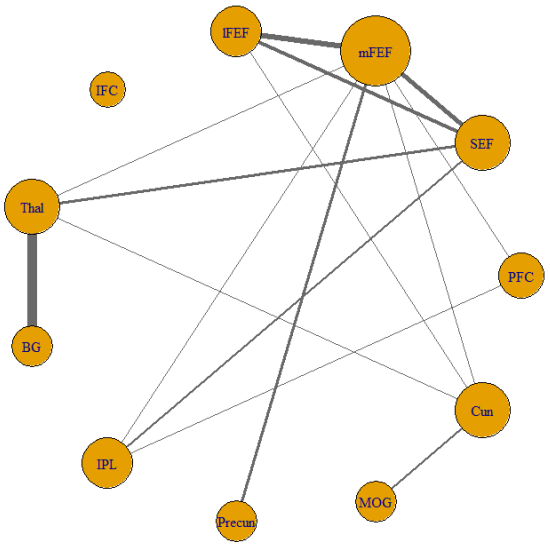
Stronger (Pre < Post)		Weaker (Pre > Post)	
PFC	IPL	PFC	-
SEF	IPL*	SEF	BG, Precun*, Cun
mFEF	Thal	mFEF	-
lFEF	-	lFEF	-
IFC	-	IFC	-
Thal	Cun	Thal	BG
BG	-	BG	-
IPL	-	IPL	-
Precun	-	Precun	-
MOG	-	MOG	-



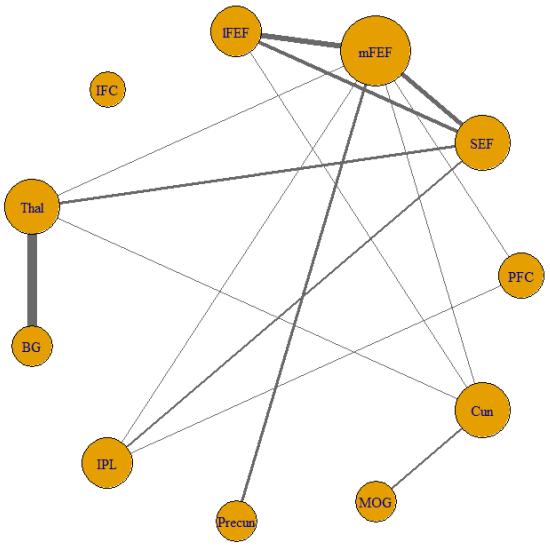
(a) JGL with RA (pre)



(b) JGL (pre)



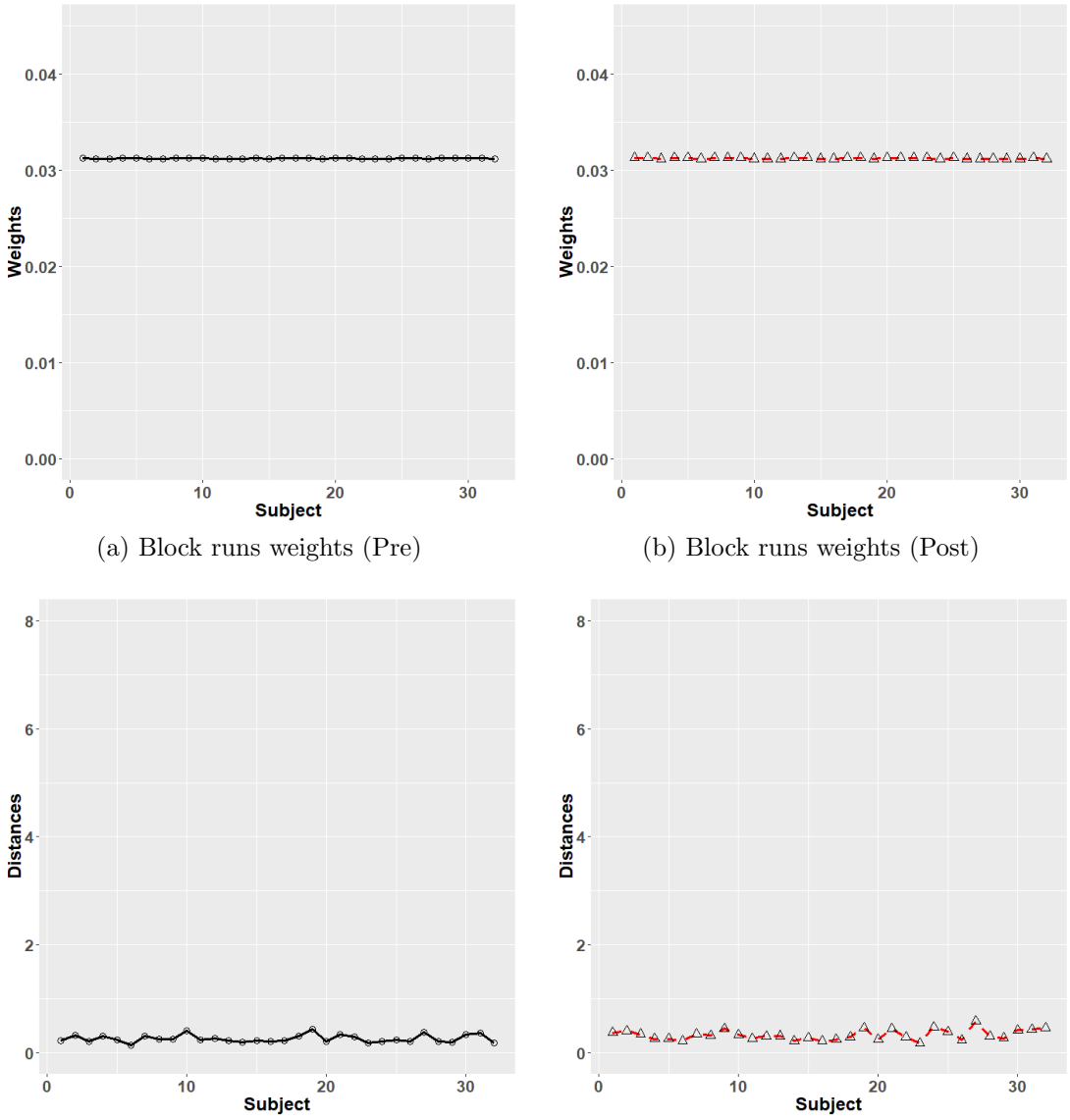
(c) JGL with RA (post)



(d) JGL (post)

Figure A.16: Group graphs by JGL with RA and JGL for the 75% run. The edges weaker than 0.03 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

A.2.2 Block Runs



(a) Block runs weights (Pre)

(b) Block runs weights (Post)

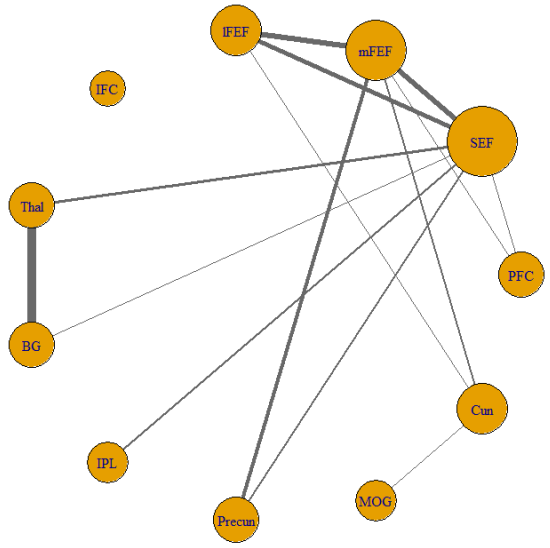
(c) Block runs distances (Pre)

(d) Block runs distances (Post)

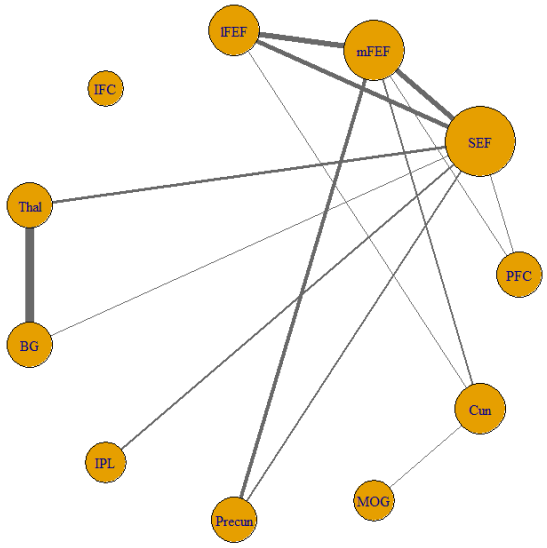
Figure A.17: Comparison of the estimated weights and distances by JGL with RA for the block runs in pre- and post-test sessions. The horizontal axe represents subject index, and the vertical axes are the estimated weights and the distances, respectively.

Table A.12: Change in relationships after the practice by JGL with RA for the block runs. The only regions which have absolute differences in $\hat{\omega}_{ij}$ greater than 0.02 between the pre- and post-test sessions are included. Especially, the regions with an asterisk (*) have absolute differences in $\hat{\omega}_{ij}$ greater than 0.04.

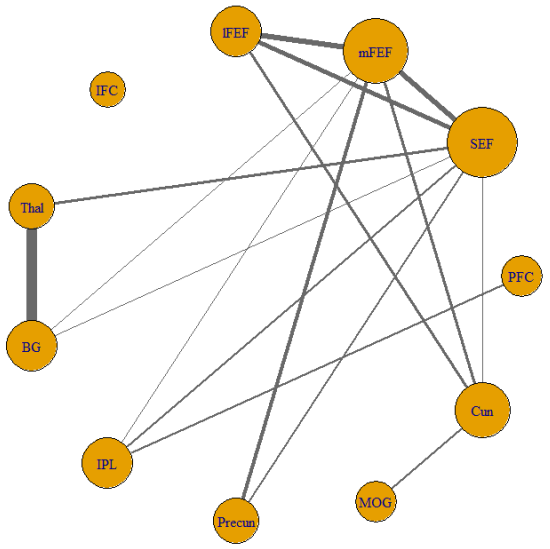
Stronger (Pre < Post)		Weaker (Pre > Post)	
PFC	IPL*	PFC	SEF, mFEF
SEF	Cun	SEF	-
mFEF	BG, IPL	mFEF	-
lFEF	Cun	lFEF	-
IFC	-	IFC	-
Thal	BG	Thal	-
BG	-	BG	-
IPL	-	IPL	-
Precun	-	Precun	-
MOG	-	MOG	-



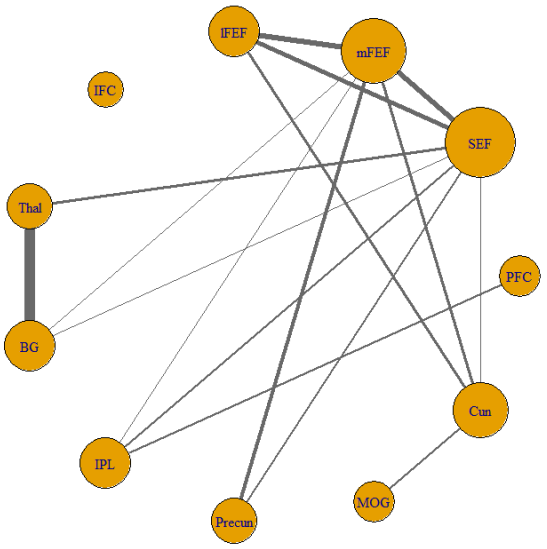
(a) JGL with RA (pre)



(b) JGL (pre)



(c) JGL with RA (post)



(d) JGL (post)

Figure A.18: Group graphs by JGL with RA and JGL for the block runs. The edges weaker than 0.03 are omitted. The size of nodes represents the number of connected edges, and the thickness of the edges shows the strength of associations between two ROIs.

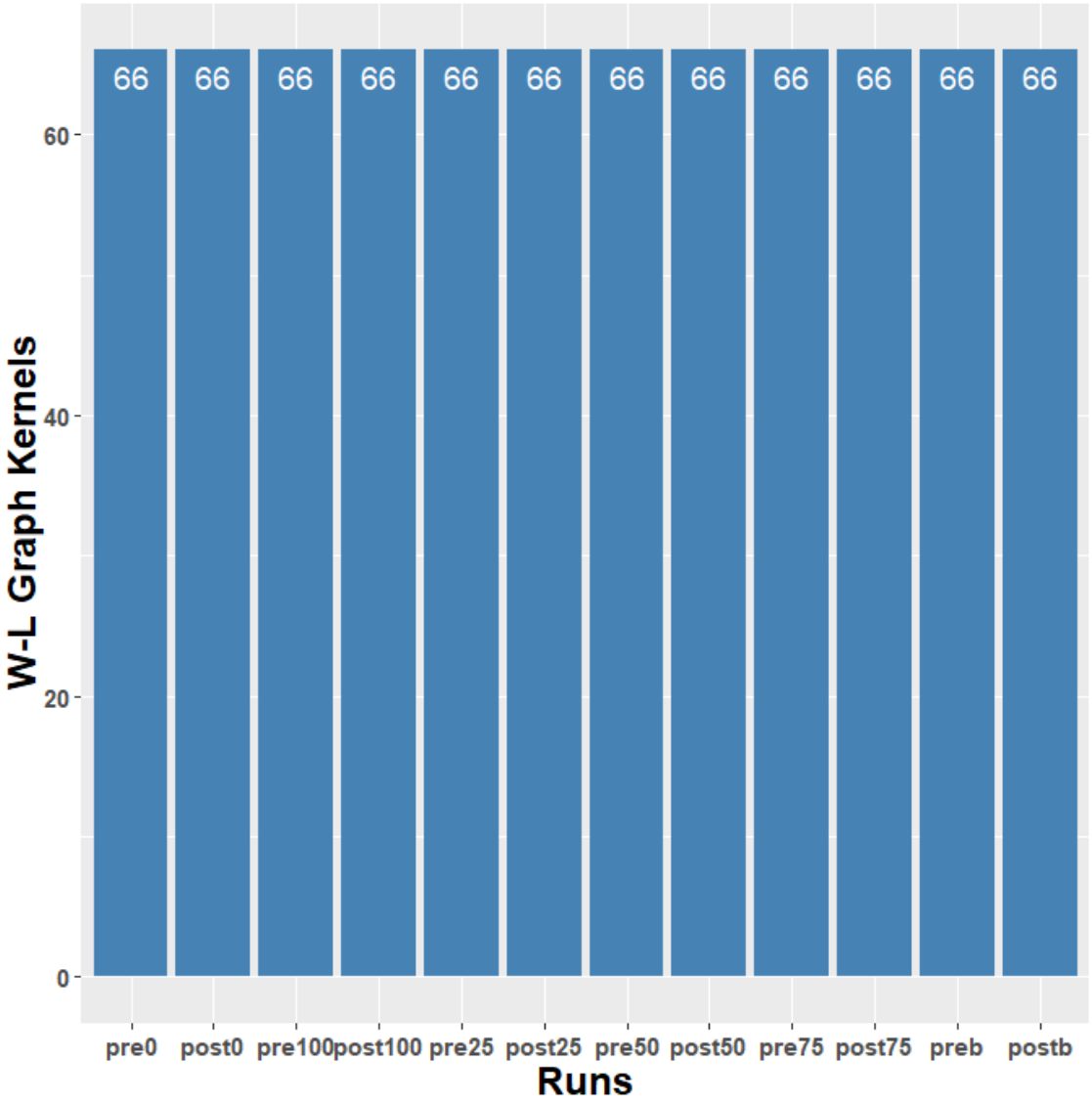


Figure A.19: Comparison of JGL with RA and JGL group graphs based on Weisfeiler-Lehman graph kernels for each session and run. The horizontal and vertical axes are index of runs (e.g. "pre0" is for the 0% run in the pre-test session of the event-related expriments, and "postb" represents the post-test session of the block runs) and W-L graph kernel values respectively.

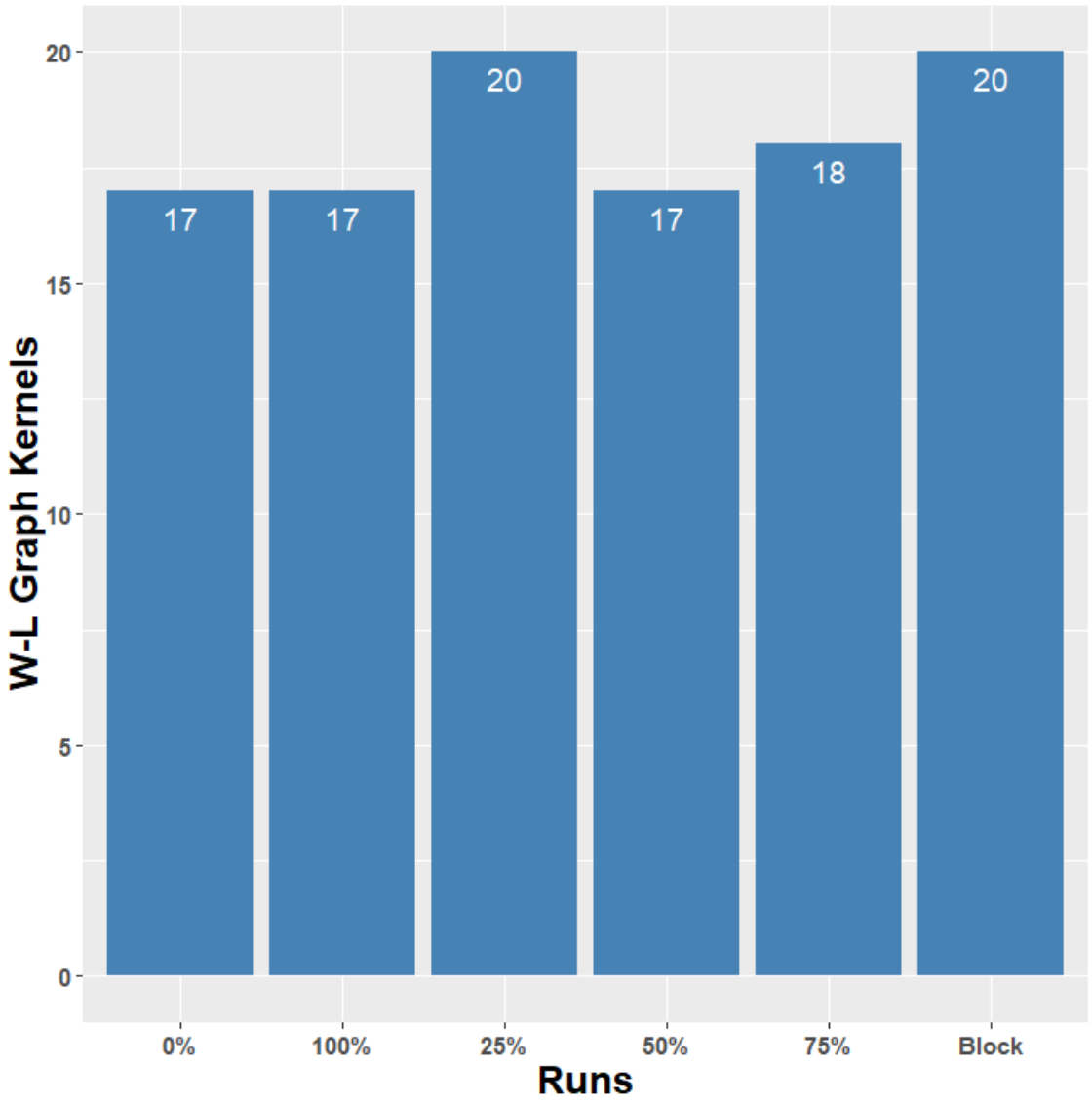


Figure A.20: Comparison of JGL with RA group graphs for the pre- and post-test session based on Weisfeiler-Lehman graph kernels. The horizontal and vertical axes are index of runs and W-L graph kernel values respectively.