Published in final edited form as:
Biometrics. 2014 December ; 70(4): 812–822. doi:10.1111/biom.12216.

Identifying Functional Co-activation Patterns in Neuroimaging Studies via Poisson Graphical Models

Wenqiong Xue[1,2,*], **Jian Kang**[1,3], **F. DuBois Bowman**[4], **Tor D. Wager**[5], and **Jian Guo**[6]

[1]Center for Biomedical Imaging Statistics, Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA

[2]Boehringer Ingelheim Pharmaceuticals, Inc., Ridgefield, CT

[3]Department of Radiology and Imaging Sciences, School of Medicine, Emory University, Atlanta, GA

[4]Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY

[5]Department of Psychology and Neuroscience, University of Colorado, Boulder, CO

[6]Department of Biostatistics, Harvard School of Public Health, Boston, MA

Summary

Studying the interactions between different brain regions is essential to achieve a more complete understanding of brain function. In this paper, we focus on identifying functional co-activation patterns and undirected functional networks in neuroimaging studies. We build a functional brain network, using a sparse covariance matrix, with elements representing associations between region-level peak activations. We adopt a penalized likelihood approach to impose sparsity on the covariance matrix based on an extended multivariate Poisson model. We obtain penalized maximum likelihood estimates via the expectation-maximization (EM) algorithm and optimize an associated tuning parameter by maximizing the predictive log-likelihood. Permutation tests on the brain co-activation patterns provide region pair and network-level inference. Simulations suggest that the proposed approach has minimal biases and provides a coverage rate close to 95% of covariance estimations. Conducting a meta-analysis of 162 functional neuroimaging studies on emotions, our model identifies a functional network that consists of connected regions within the basal ganglia, limbic system, and other emotion-related brain regions. We characterize this network through statistical inference on region-pair connections as well as by graph measures.

[*]wxue@emory.edu.

Supplementary Materials

The Web Supplementary Materials including Web Appendices A–E referenced in Sections 1, 2.1, 2.2, 2.3, and 3.1 in the article are available at the *Biometrics* website on Wiley Online Library. The Web Supplementary Materials also include a description of the software in the Web Appendix F. The Matlab source code along with README are available at the webpage: http://web1.sph.emory.edu/users/jkang30/software/PoissonGraph.html.

The code provides `Matlab` functions and an example script to make statistical inference on the Poisson graphical model.

Keywords

EM algorithm; Emotion; Functional brain networks; Functional co-activation pattern identification; Poisson Graphical Model

1. Introduction

Gaining insights into the human brain fundamentally relies on understanding interactions in neural activity between distinct brain regions. Functional neuroimaging technologies, such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET), enable investigations into such interactions. The concept of functional connectivity is applied widely in imaging studies to capture associations in brain function between different regions. For example, in a typical fMRI study, one can investigate functional connectivity by estimating the Pearson correlation, partial correlation, mutual information, spectral coherence, or joint activation between the spatially localized time series of blood oxygen level dependent (BOLD) signals, and a number of multivariate methods are available (Calhoun et al., 2001; McIntosh and Gonzalaz-Lima, 1994; Patel et al., 2006; Friston et al., 2003; Wager et al., 2009).

Determining functional connectivity using data from a single functional neuroimaging study suffers some limitations, such as low power and high false positive rates due to small sample sizes. This motivates the need to model neural activity interactions between brain regions across independent studies using ideas from meta-analyses. In this article, we perform co-activation meta-analysis to make inferences regarding functional connectivity, defined here by consistent co-activation between brain regions across different studies. Of note, our definition of FC is distinct from the most widely used one(s), which refer to Pearson (or partial) correlation between the time series for different voxels or brain regions (within-subject) (Friston et al., 2003). However, it still provides information on the extent of association by defining FC in terms of joint activation (within-subject) between two voxels (or regions), in accordance with definitions from previously published work (Patel et al. 2006; Friston, 2011). It is also important to note that our goal is different from that of common coordinate based meta-analyses of functional neuroimaging studies (Nielsen and Hansen, 2002; Turkeltaub et al., 2002; Wager et al., 2004; Kober et al., 2008; Eickhoff et al., 2009; Kang et al., 2011), which focus on identifying the consistent activation locations across studies.

The typical pipeline of analyzing fMRI BOLD time series in a single functional neuroimaging study is shown in Figure 1(A). After data acquisition, several preprocessing steps are taken prior to analysis to minimize the influence of acquisition and physiological artifacts. In a standard statistical analysis, a two-stage model is applied, wherein the first stage fits the time series separately at each voxel and for each subject, and the difference (or contrast) in signal magnitudes is quantified between different conditions. This procedure is repeated for all the voxels in the brain and the resulting statistics are summarized in a statistical parametric map (e.g. a *t* map). In the second stage, one then produces a group level statistic map and generates a threshold, which usually accounts for multiple testing, and obtains a thresholded *t* map that identifies the statistically significant brain locations.

This process is called statistical parametric mapping (SPM) in the neuroimaging literature. Single studies rarely report the full SPM. Instead, they only report peak activation coordinates, which are the voxels with the largest *t* statistic in significant regions (or among clustered sets of significant voxels) of an SPM. We refer to these peak activation coordinates as foci or a single focus. For any pair of regions, the number of foci located in both regions from all studies reflects the strength of the co-activations for that region pair.

Using foci that are collected from different studies, methods to date have focused on identifying the co-activation for region pairs (Cauda et al., 2011; Kober et al., 2008; Patel et al., 2013; Postuma and Dagher, 2006; Robinson et al., 2012; Torta and Cauda, 2011) or component-driven approaches (e.g., principal component analysis (PCA) and independent component analysis (ICA) combined with clustering methods; Kober et al., 2008). In particular, Nielsen and Hansen (2004) proposed a matrix factorization algorithm, in which a matrix that represents the activation associated with particular tasks is decomposed. Kober et al. (2008) proposed a functional grouping approach, which analyzes the spatial density of reported foci using multilevel kernel density analysis (MKDA) and then combines non-metric multidimensional scaling and cluster analysis to group regions based on their co-activation patterns. Neuman et al. (2010) develop a structural learning approach in a Bayesian framework to construct a directed functional network, which yields probabilistic dependency between brain regions. The above methods are intended to identify co-activated regions, but they do not permit likelihood-based statistical inference for co-activation patterns of multiple foci.

There is a need to make joint inferences on co-activation for multiple regions (e.g., graphical) structures by directly modeling a multivariate distribution of foci counts in different regions, which can characterize the covariance structure of foci linking to a brain co-activation pattern. This enables researchers to estimate the joint likelihood of activations for multiple regions. We propose a multivariate Poisson graphical model (Weiss and Freeman, 2001) to estimate the co-activation patterns and to make inferences about functional networks defined by consistent co-activation patterns across studies. We characterize region-level co-activation patterns using the covariance of the number of foci in different regions, where these region-specific activation counts are jointly modeled by a multivariate Poisson distribution (Kawamura 1979; Karlis 2003; Yang and Kang 2010) (see Web Appendix A for details). We impose sparsity on the covariance function by assuming that only a small number of the potentially massive number of region pairs are co-activated for a particular brain function. Such sparsity considerations are supported by previous research findings (Huang et al., 2010). We propose a penalized likelihood approach to efficiently estimate the sparse covariance function. Specifically, we introduce a set of latent variables to facilitate obtaining the penalized maximum likelihood estimates (PMLE) of the covariance via the expectation-maximization (EM) algorithm. The latent variables explicitly model the expected number of co-activation foci between regions. The undirected functional network is then determined by the region-level estimated covariance between the numbers of foci. The proposed shrinkage method is tuned to reproduce the sparsity found in typical brain networks, and we optimize the shrinkage parameter based on the predictive log-likelihood function.

We consider an analysis of emotion studies to motivate our proposed modeling framework. We collect findings from 162 functional neuroimaging studies (publications), including 57 PET and 105 fMRI studies. Collectively, these studies yield 2478 activation coordinates from 437 contrasts (e.g., happy vs. neutral) as in Kober et al. (2008). These studies were published in peer-reviewed journals from 1990 to 2005. We consider a total of seven different emotions, including sad (45 contrasts), happy (36 contrasts), anger (26 contrasts), fear (68 contrasts), disgust (44 contrasts), surprise (2 contrasts), mixed (41 contrasts) and affective (175 contrasts), a category for commonly used emotional stimuli that are not clearly associated with a single emotion category. For each study, the activation locations for these contrasts are included when they meet the criteria of statistical significance defined within each individual study. For each contrast, the coordinates of foci are assigned to different brain regions based on a widely used neuroanatomic parcellation (details provided later). The foci count over different regions represents the data in our analysis. An illustration of the emotion data is in Figure 1(B).

Our model makes several novel contributions. First, we are among the first to propose a framework for functional co-activation pattern identification to determine brain networks based on region-level activation counts, which are commonly used in neuroimaging meta-analyses targeting localization. Our framework is based on a graphical model to represent sparse brain networks, extending the multivariate Poisson model. Secondly, our approach provides more interpretable results than many existing methods by explicitly modeling the strength of functional co-activations. Third, we propose a fast computational algorithm for parameter estimation as well as a feasible permutation testing procedure to assess the statistical significance of the identified brain network(s).

2. Methods

2.1 The Bivariate Model

We start with a bivariate model for any two regions i and j in the brain. Let n denote the number of contrasts in all the studies included in the meta-analysis. For contrast k(k = 1, …, n), let $X_{i,k}$ and $X_{j,k}$ represent the number of foci in regions i and j. We assume that $(X_{i,k}, X_{j,k})'$ follows a bivariate Poisson distribution with parameter vector $\boldsymbol{\lambda}= (\lambda_{ii}, \lambda_{jj}, \lambda_{ij})$, where λ_{ii} is the variance of $X_{i,k}$, λ_{jj} represents the variance of $X_{j,k}$, and λ_{ij} is the covariance of $X_{i,k}$ and $X_{j,k}$ (Kocherlakota and Kocherlakota, 1992). The joint probability function is

$$P(X_{i,k}=x_{i,k}, X_{j,k}=x_{j,k})=e^{-(\lambda_{ii}+\lambda_{jj}+\lambda_{ij})}\frac{\lambda_{ii}^{x_{i,k}}}{x_{i,k}!}\frac{\lambda_{jj}^{x_{j,k}}}{x_{j,k}!}\sum_{s=0}^{\min(x_{i,k},x_{j,k})}\binom{x_{i,k}}{s}\binom{x_{j,k}}{s}s!\left(\frac{\lambda_{ij}}{\lambda_{ii}\lambda_{jj}}\right)^{s}, \quad (1)$$

where $x_{l,k}$ = 1, 2,…, for l = i, j. Marginally $X_{i,k}$ and $X_{j,k}$ both follow Poisson distributions with respective parameters $\lambda_{ii} + \lambda_{ij}$ and $\lambda_{jj} + \lambda_{ij}$. The covariance between $X_{i,k}$ and $X_{j,k}$, λ_{ij}, is interpreted as the strength of the co-activation between regions i and j, and $\lambda_{ij} = 0$ implies no statistical dependence between the regions.

We impose sparsity on the brain network by adding a penalty term to the likelihood based on (1), which shrinks λ_{ij} toward zero. Let $\mathbf{X}_i$ and $\mathbf{X}_j$ represent the vector containing foci

counts for all contrasts in region *i* and *j*, respectively. We minimize the following penalized negative log-likelihood function with respect to $\boldsymbol{\lambda}$ given θ:

$$-l_{\text{obs}}(\boldsymbol{\lambda};\mathbf{X}_i,\mathbf{X}_j)+\theta\lambda_{ij} = \sum_{k=1}^{n}\left\{\sum_{l=i,j}[\lambda_{ll}-X_{l,k}\log(\lambda_{ll})]+\lambda_{ij}-\log\left[\sum_{s=0}^{\min(x_i,x_j)}\begin{pmatrix}X_{i,k}\\ s\end{pmatrix}\begin{pmatrix}X_{j,k}\\ s\end{pmatrix}s!\left(\frac{\lambda_{ij}}{\lambda_{ii}\lambda_{jj}}\right)^{s}\right]\right\}+\theta\lambda_{ij}, \quad (2)$$

where the parameter θ controls the degree of sparsity. Larger values of θ will tend to shrink the covariance parameters toward zero, reflecting more sparsity in the brain network. For each network, the potential number of connections is $\begin{pmatrix}p\\ 2\end{pmatrix}$, where *p* is the number of regions in the network. Despite this large number, Huang et al. (2010) posit that the functional connectivity network can be estimated by a sparse matrix due to network theories suggesting efficient neural processing. By incorporating sparsity, our approach attempts to tease out the more prominent connections in the brain.

The joint probability function of ($\mathbf{X}_i$, $\mathbf{X}_j$) is complicated, especially when the number of dimensions is large. Kano and Kawamura (1991) derived a recursive scheme for constructing the probability function of a multivariate Poisson distribution. However, the computational demand and errors induced by recursion increase with the number of dimensions. Kalis (2003) alternatively proposed an EM algorithm (Dempster et al., 1977; Meng and Van Dyk, 1997; McLachlan and Krishnan, 1997) based on the multivariate reduction derivation of the multivariate Poisson distribution for estimation, which we adopt in our paper.

To simplify computations, we introduce a latent Poisson variable $Y_{ij,k}$ with mean λ_{ij} to represent the number of co-activations in the two regions. We define

$$\begin{cases} Y_{ii,k}=X_{i,k}-Y_{ij,k}, \\ Y_{jj,k}=X_{j,k}-Y_{ij,k}, \end{cases} \quad (3)$$

where $Y_{ii,k}$ and $Y_{jj,k}$ represent the number of localized foci in regions *i* and *j*, respectively (without joint activation). This implies a new representation of $\mathbf{X}_i$: $X_{i,k} = Y_{ii,k} + Y_{ij,k}$ and $X_{j,k} = Y_{jj,k} + Y_{ij,k}$. Also, the knowledge of latent variables $Y_{ij,k}$'s completely specify the data. By the distribution of $Y_{ij,k}$, $X_{i,k}$, and $X_{j,k}$, it is straightforward to show that

$$Y_{h,k} \overset{i.i.d}{\sim} \text{Poisson}(\lambda_h), \text{ for } h \in \{ii, jj\}. \quad (4)$$

The expectation of $Y_{ij,k}$, λ_{ij}, characterizes the covariance between $X_{i,k}$ and $X_{j,k}$. Denoting $\mathbf{Y}_{ij} = (Y_{ij,1}, \ldots, Y_{ij,n})'$ collectively for all contrasts, the penalized complete data negative log-likelihood functiona is given by

$$-l_{\text{comp}}(\boldsymbol{\lambda};\boldsymbol{Y}_{ij},\mathbf{X}_i,\mathbf{X}_j)+\theta\lambda_{ij} = \sum_{k=1}^{n}\left\{\sum_{l=i,j}[\lambda_{ll}-(X_{l,k}-Y_{ij,k})\log(\lambda_{ll})]+\lambda_{ij}-Y_{ij,k}\log(\lambda_{ij})\right\}+\theta\lambda_{ij}. \quad (5)$$

We use an EM-algorithm to minimize (5). The E-step calculates the conditional expectation of unobserved data $\mathbf{Y}_{ij}$ given the observed data $\mathbf{X} = (\mathbf{X}_i, \mathbf{X}_j)$ using the current estimates of the parameters, and the M-step minimizes the penalized complete data negative log-likelihood.

The EM algorithm is described as follows, with an initial value of $\boldsymbol{\lambda}^{(0)}$, for $t = 0, \ldots, T - 1$, in the $(t + 1)$th step,

- E-step, compute

$$Y_{ij,k}^{(t+1)} = E[Y_{ij,k} | X_{i,k}, X_{j,k}; \boldsymbol{\lambda}^{(t)}] = \sum_{y_{ij,k}=0}^{\min(x_{i,k}, x_{j,k})} \frac{y_{ij,k} P(Y_{ij,k}, X_{i,k}, X_{j,k}; \boldsymbol{\lambda}^{(t)})}{\sum_{y_{ij,k}=0}^{\min(x_{i,k}, x_{j,k})} P(Y_{ij,k}, X_{i,k}, X_{j,k}; \boldsymbol{\lambda}^{(t)})}, \quad (6)$$

- M-step, update the estimates by

$$\lambda_{ij}^{(t+1)} = \frac{\sum_{k=1}^{n} Y_{ij,k}^{(t+1)}}{\theta + n},$$
$$\lambda_{ll}^{(t+1)} = \frac{1}{n} \sum_{k=1}^{n} X_{l,k} - \frac{\theta + n}{n} \lambda_{ij}^{(t+1)}, \text{ for } l = i, j. \quad (7)$$

The iteration proceeds between the E-steps and M-steps and stops after attaining specified convergence criteria. Positive constraints are added in the algorithm to ensure the estimates are nonnegative. The joint probability function $P(Y_{ij,k}, X_{i,k}, X_{j,k})$ in (6) is derived in Web Appendix B. This algorithm is a special case of one addressed by Kalis (2003), which considers multivariate cases with the same covariance for all the pairs of random variables. Next, we consider a more general case that assumes an unstructured covariance matrix.

2.2 The Multivariate Model

A p-dimensional Poisson model might include p-way interactions. This potentially leads to a very complicated model with extremely high computational costs. In practice, an m-way interaction model should suffice for high dimensional neuroimaging applications, where $m << p$. We limit our attention to two-way interactions, e.g., the covariance between $\mathbf{X}_i$ and $\mathbf{X}_j$ for $1 \leq i, j \leq p$, which should be sufficient to construct brain networks of interest.

Suppose that we observe the number of foci in p regions, denoted $\mathbf{X}_k = (X_{1,k}, \ldots, X_{p,k})'$ for each contrast $k = 1, \ldots, n$. We assume that $\mathbf{X}_k$ follows a multivariate Poisson distribution with parameters $\boldsymbol{\lambda} = (\lambda_{ij})_{1 \leq i,j \leq p}$. Extending (3), we have

$$X_{i,k} = \sum_{j=1}^{p} Y_{ij,k}, \text{ for } i = 1, \cdots, p, \quad (8)$$

and $\mathbf{Y}_k = (Y_{ij,k})_{1 \leq i \leq j \leq p}$ is a collection of independent Poisson random variables, where each $Y_{ij,k}$ follows a Poisson distribution with parameter λ_{ij}. Also, each $X_{i,k}$ marginally follows a Poisson distribution with parameter $\sum_{j=1}^{p} \lambda_{ij}$. The observed number of foci in region i, $X_{i,k}$

can be decomposed into p parts: $Y_{ii,k}$, the number of (singular) localized activations in region i, and $\{Y_{ij,k}\}_{j\neq i}$, the number of co-activations in region i and each of the remaining regions. We have symmetry in covariance parameters, i.e., $\lambda_{ij} = \lambda_{ji}$, as well as in the number of co-activations between regions i and j, and between regions j and i, i.e., $Y_{ij,k} = Y_{ji,k}$.

To incorporate sparsity in the covariance structure of the brain network, we utilize the following penalized observed log-likelihood:

$$-l_{\text{obs}}(\boldsymbol{\lambda};\mathbf{X}_1,\mathbf{X}_2,\cdots,\mathbf{X}_n)+\theta\sum_{i=1}^{p}\sum_{j=i+1}^{p}\lambda_{ij}. \quad (9)$$

Similar to the bivariate model, we compute the complete data likelihood for ($\tilde{\boldsymbol{Y}}_k$, $\mathbf{X}_k$), where $\tilde{\boldsymbol{Y}}_k=\{Y_{ij,k}\}_{1\leq i<j\leq p}$ contains all the information of co-activations patterns. Specifically, we consider the following penalized complete data negative log-likelihood

$$\begin{aligned}&-l_{\text{comp}}(\boldsymbol{\lambda};\tilde{\mathbf{Y}}_1,\cdots,\tilde{\mathbf{Y}}_n,\mathbf{X}_1,\cdots,\mathbf{X}_n)+\theta\sum_{i=1}^{p}\sum_{j=i+1}^{p}\lambda_{ij}\\&=\sum_{k=1}^{n}\sum_{i=1}^{p}\sum_{j=i}^{p}[\lambda_{ij}-Y_{ij,k}\log(\lambda_{ij})]+\theta\sum_{i=1}^{p}\sum_{j=i+1}^{p}\lambda_{ij}.\end{aligned} \quad (10)$$

To minimize the observed negative log-likelihood (9) using the EM algorithm, we calculate the conditional expectation of each $Y_{ij,k}$ given the observed data and current estimates of the unknown parameters. Note that the conditional probability of $Y_{ij,k}$ given the observed data only depends on $X_{i,k}$ and $X_{j,k}$, i.e., $E(Y_{ij,k}|\mathbf{X}_k;\boldsymbol{\lambda}) = E(Y_{ij,k}|X_{i,k}, X_{j,k};\boldsymbol{\lambda})$. Given the initial value $\boldsymbol{\lambda}^{(0)}$, in the $(t+1)$th step, for $t = 0, \ldots, T-1$, the procedure proceeds as follows:

- E-step, for $i = 1, \ldots, p$, $j = i+1, \ldots, p$, compute

$$\begin{aligned}&Y_{ij,k}^{(t+1)}=E[Y_{ij,k}|X_{i,k},X_{j,k};\lambda^{(t)}]\\&=\sum_{y_{ij,k}=0}^{\min(x_{i,k},x_{j,k})}\frac{y_{ij,k}P(Y_{ij,k},X_{i,k},X_{j,k};\lambda^{(t)})}{\sum_{y_{ij,k}=0}^{\min(x_{i,k},x_{j,k})}P(Y_{ij,k},X_{i,k},X_{j,k};\lambda^{(t)})}\end{aligned} \quad (11)$$

- M-step, update the estimates by

$$\begin{aligned}\lambda_{ij}^{(t+1)}&=\frac{\sum_{k=1}^{n}Y_{ij,k}^{(t+1)}}{\theta+n},\ \text{for } 1\leq i<j\leq p\\\lambda_{ii}^{(t+1)}&=\frac{1}{n}\sum_{k=1}^{n}X_{i,k}-\frac{\theta+n}{n}\sum_{j\neq i}\lambda_{ij}^{(t+1)}\end{aligned} \quad (12)$$

The joint probability function $P(Y_{ij,k}, X_{i,k}, X_{j,k})$ in (11) is derived in Web Appendix C. Note that the penalty term shrinks some covariance estimates exactly to zero and others to very small (near zero) values so that co-activations are sparsely detected. We declare that two regions are not connected if the estimated λ_{ij} is near zero. In practice, we use 10^{-3} as the threshold, which on average, corresponds to fewer than one pair of co-activating foci reported on the two regions across 1000 independent studies.

2.3 Tuning Parameter

We consider the predictive log-likelihood as the criterion to determine the optimal value of the tuning parameter θ. In our upcoming simulation studies, we also examine the mean-squared error as a supplementary tool to verify our findings. Suppose we have two sets of data: the training data for parameter estimations and the testing data for validations, denoted as $\mathbf{X}_{\text{train}}$ and $\mathbf{X}_{\text{test}}$, respectively. The estimate of λ_{ij} derived from the EM algorithm, given θ, using the training dataset $\mathbf{X}_{\text{train}}$ is denoted by $\hat{\lambda}_{ij}(\theta)$. The predictive log-likelihood evaluated at $\hat{\boldsymbol{\lambda}}(\theta)=\left\{\hat{\lambda}_{ij}(\theta)\right\}_{1\leq i,j\leq p}$ is defined as:

$$l_{\text{obs}}(\hat{\boldsymbol{\lambda}}(\theta);\mathbf{X}_{\text{test}})=\sum_{k=1}^{n} l_{\text{obs}}(\hat{\boldsymbol{\lambda}}(\theta);\mathbf{X}_{\text{test},k}). \quad (13)$$

For a detailed expression, please refer to Web Appendix D.

We select the value of θ that yields the maximum predictive log-likelihood using two independent grid searches, including coarser grids and finer grids over a specified range. We use five-fold cross validation to optimize θ in the simulation studies and ten-fold cross validation in the data application to achieve the balance in training fitting and prediction error between smaller and larger dataset. In particular, five-fold was chosen in the simulation studies mainly due to the relatively low computational costs, while ten-fold was preferred in the analysis of our emotion dataset to ensure a sufficiently large training set to accommodate the fairly small sample. To verify the findings from the predictive log-likelihood, we also optimize θ in the simulation studies by minimizing the mean-squared error (MSE) over a specified grid, with the MSE defined as

$$\text{MSE}(\theta)=\sum_{i\leq j}(\lambda_{ij}-\hat{\lambda}_{ij}(\theta))^2. \quad (14)$$

In our data application, we only use the predictive log-likelihood for optimizing the tuning parameter, since the MSE depends on the true value of λ. In Section 3.2.4, we check the accuracy of the predictive log-likelihood method by demonstrating that it yields results that are very close to those based on the MSE criterion using simulated data.

2.4 Statistical Testing

In order to make valid inference on co-activation patterns between regions and the associated functional network, we construct two permutation tests: Test I focuses on detecting the existence of connections for a set of region pairs, and Test II focuses a particular network. We utilize the penalized likelihood approach developed in Section 2.2 to determine a set of region pairs for Test I and a particular brain network for Test II. However, the two tests are both built upon a standard likelihood, without penalization. Specifically, we first compute the MLE of λ_{ij} without shrinkage, denoted $\hat{\lambda}_{ij}^{\text{MLE}}$. We create a set of permutation samples by permuting the contrast labels for each region *i*. We compute the estimated λ_{ij} for each region pair *i* and *j* from the permuted datasets, denoted $\hat{\lambda}_{ij}^{\text{Perm}}$, to

obtain a simulated null distribution of $\hat{\lambda}_{ij}$, given that there is no connection between regions *i* and *j*. Then we calculate the uncorrected p-value for each region pair *i* and *j*, which equals the proportion of $\hat{\lambda}_{ij}^{\mathrm{Perm}}$'s that are greater than $\hat{\lambda}_{ij}^{\mathrm{MLE}}$. This p-value reflects evidence against the null hypothesis of Test I $H_0 : \lambda_{ij} = 0$ versus the alternative that $H_a : \lambda_{ij} > 0$. We apply the false discovery rate (FDR) approach by Benjamini and Hochberg (1995) to correct for multiple comparisons for inferences about λ_{ij}.

Based on $\hat{\lambda}_{ij}^{obs}$'s, we build a functional network, denoted by Φ, whose significance can be tested by Test II. Specifically, we consider the following hypothesis for network identification: H_0: $\lambda_{ij} = 0, \forall\{ij\} \in \Phi$ vs. H_a: $\exists\{ij\} \in \Phi$ s.t. $\lambda_{ij} > 0$. With a similar permutation procedure, the p-value for the network identification equals the proportion of $\hat{\lambda}_{ij}$'s, estimated from the permuted datasets, that are greater than the corresponding $\hat{\lambda}_{ij}^{obs}$ for each pair of *i* and *j*.

2.5 Graph Measures of the Network

To furgher describe properties of a discovered brain network, we assess topological properties by performing graph analyses using a set of measures, commonly referred to as graph metrics (not to be confused with our graphical model). Suppose each network is composed of *s* nodes and *t* edges, which represent *s* brain regions and *t* co-activation connections, respectively. Several metrics have been developed to describe the relationships between nodes. For example, the clustering coefficient *C* measures the average likelihood of connecting neighbors. For each node *i*, the clustering coefficient is defined as $C_i = 2E_i/k_i(k_i - 1)$, where k_i is the degree of node *i* and E_i is the number of direct links connecting neighbors of node *i*. The path length *L* is the average minimum number of connections to link two nodes. Network topology is described as a small-world network if compared to a similar random network, the small-world index $\sigma = (C/C_{\mathrm{random}})/(L/L_{\mathrm{random}}) > 1$ (Watts and Strogatz, 1998; Humphries et al., 2006). Here, a similar random network is defined as a network with the same number of nodes, the same number of edges, and the same degree distribution (Simpson et al., 2013). Examples of small-world networks include road maps, food chains, and social influence network, in which most nodes can be connected to others by a small number of connections. We conduct statistical testing on the small-worldness property of an identified functional network using permutation methods, specifically addressing the hypotheses: H_0: $\sigma \leq 1$ vs. $H_a : \sigma > 1$.

Hubs play a central role in a network since they serve as the common connections to other nodes. We define hubs as nodes with high degree (D). The degree of a node *i* is the number of times node *i* is connected to other nodes (Freeman, 1977). We examine the high-degree nodes, i.e., the nodes with a degree or centrality at least one standard deviation above the network mean, in the network (Sporns et al., 2007).

3. Results

3.1 Identifying Functional Co-activation Patterns in Neuroimaging Emotion Studies

We apply our proposed method to a dataset consisting of 437 contrasts from 162 studies (Kober et al., 2008). Each contrast is associated with one of seven emotion categories (sad, happy, anger, fear, disgust, surprise and affective). We use the reported coordinates of activation for each contrast to assign the contrast to a specific brain region. The number of reported points in each region is used as the data for our model. Approximately 6 activated coordinates are reported for each contrast, on average. We use the GlaxoSmithKline Clinical Imaging Centre (CIC) (Tziortzi et al., 2011) brain atlas based on the Harvard-Oxford atlas (Makris et al., 2006) and consider 19 regions of interest (ROIs) related to emotion processing, yielding a 19 × 437 data matrix and 171 region pairs. The objectives of this study include estimating the co-activation patterns and the corresponding functional network for emotion as a whole (as an illustration), performing statistical testing for the connections between regions, and characterizing the identified brain network.

The dorsolateral prefrontal cortex (DLFC) is the most frequently reported region across studies, with a 0.500 probability of reported activation (217 of 437 contrasts), and the right globus pallidus (GP_R) is reported the least across studies, with a 0.007 probability of reported activation (3 out of 437 contrasts). On average, each region is found to be associated with the neural processing of emotions with probability 0.140 (61 times out of 437 contrasts).

We perform cross validation to choose the value of θ that yields the largest predictive likelihoods, and we estimate the covariance parameters based on the selected θ. Following the steps described in section 2.4, we identify the functional network and perform statistical testing on the network and marginal distributions of co-activations between regions. Our method detects an emotional processing network including 17 ROIs with 79 connections.

We find strong functional co-activation patterns within the limbic system, the basal ganglia and other frequently reported emotion related brain regions, as shown in Figure 2. The anterior cingulate cortex (ACC) is thought to be involved in reward and other diverse affective/motivational processes. Our analysis reveals that it is functionally connected to 11 other regions in the identified network. Among the 6 region pairs with the highest covariance estimates, ACC appears 4 times, while the other two region pairs are bilateral homologues of the same structure. For example, ACC is functionally connected with orbitofrontal cortex (OFC), which is one of the major centers for affective processing ($\hat{\lambda}_{ij}=0.023$, $p < 0.005$). Strong co-activation is also identified between ACC and the striatum (Str) ($\hat{\lambda}_{ij}=0.018$, $p < 0.005$); ACC and the thalamus (Thal) ($\hat{\lambda}_{ij}=0.013$, $p < 0.005$), as well as ACC and the frontal operculum (frOP) ($\hat{\lambda}_{ij}=0.012$, $p < 0.005$). Almost all of the marginal connections shown in Figure 2 are significant after an FDR correction ($p < 0.005$), and the network is significant as well ($p < 0.005$). The heat map depicting the estimated covariance matrix is available in Web Appendix E.

We also examine graph properties of the emotion processing network. The clustering coefficient of the identified network is $C = 0.710$, and the path length is $L = 1.129$. The corresponding small-world index $\hat{\sigma}=1.027$ is significantly greater than 1 ($p < 0.005$), which indicates that the identified network has properties which are consistent with a small-world network compared to the average of 1000 random networks ($C_{\text{random}} = 0.693$, $L_{\text{random}} = 1.131$). We use degree and centrality measures to identify the network hubs. We find several regions that play important roles in the emotion processing network, e.g., the right insular (Ins_R) (D=14), Thalamus (Thal) (D=14), the left amygdala (Amy_L) (D=11) and the medial frontal cortex (MFC) (D=12). Ins, Thalamus and the amygdala are among the most reported emotion-related regions, and the medial frontal cortex is involved in cognitive control and related processes in a variety of settings and may reflect some of the cognitive "ingredients" of the emotion generation process or, alternatively, may play a more direct role in the generation of emotional feelings.

We further examine subnetworks separately for each emotion. We focus on negative emotions, due to the restricted number of studies and contrasts involving positive emotions. For all the region pairs within the identified network, we count the number of times that both regions have at least one peak activation coordinate reported for particular emotions. We expect to see more sparse networks for distinct emotions relative to the network containing all emotions (see Figure 3). The region pairs with top frequencies for anger, disgust, fear, and sadness appear in Table 1. Within each subnetwork, we focus on the pair that co-activates most and find that some of the emotions share co-activation patterns. For instance, DLFC and MFC, which are involved in cognitive control, are identified by anger and sadness emotions. OFC and ACC are found to show strong co-activations in anger and disgust, indicating that these two emotions may stimulate similar neural activity. In addition, a bilateral co-activation in the amygdala is detected in fear. The subnetwork analysis shows that although different types of emotions have their own contributions to the collective affective processing network, similar co-activation patterns may underlie distinct emotions.

3.2 Simulation Studies

We conduct simulation studies to evaluate estimation accuracy for our model, to assess two approaches for selecting the optimal tuning parameter θ, and to examine the impact of different values of θ on the resulting network. We generate two simulation datasets for our assessments.

3.2.1 Simulated Datasets

Dataset 1: The first simulation setting includes three regions and six non-zero parameters. Specifically, we let

$$\boldsymbol{\lambda}=\begin{pmatrix} 1 & 3 & 1 \\ 2 & 5 & \\ 3 & & \end{pmatrix}. \quad (15)$$

We generate a total of 300 datasets, and for each simulation, we draw 100 bootstrap samples to evaluate estimation accuracy.

Dataset 2: The second simulation setting builds a network consisting of eight regions. We assume the existence of co-activations for 8 of the 28 region pairs. Specifically, we set co-activations for the following region pairs: 1 and 2 ($\lambda_{12} = 3$), 1 and 5 ($\lambda_{15} = 4$), 1 and 6 ($\lambda_{16} = 2$), 2 and 7 ($\lambda_{27} = 2$), 3 and 6 ($\lambda_{36} = 3$), 4 and 8 ($\lambda_{48} = 4$), 5 and 7 ($\lambda_{57} = 5$), and 7 and 8 ($\lambda_{78} = 1$). We generate 500 datasets for this setting.

3.2.2 Estimation Accuracy—First, we evaluate accuracy when $\theta = 0$. Using simulated dataset 1, we estimate the six non-zero parameters λ_{ij} by applying our penalized multivariate Poisson model, we calculate the variance of each parameter via bootstrap resampling, and we examine the estimated coverage rates. Table 2 shows the average bias, with the percentage change in parentheses, and the coverage rate of the estimation from 300 simulations. The average bias is 0.008 with an average of 0.46% change over six parameters. Also the average coverage rate is 94.17%. These results indicate that our method accurately estimates the parameters of interest.

We compare our proposed model to the sample covariance between $X_{i,k}$ and $X_{j,k}$, which is a moment estimate of λ_{ij}. The results in Table 2 indicate that our penalized multivariate Poisson model substantially improves the performance by decreasing the biases (0.008 vs. 0.141) with percentage changes (0.46% vs. 7.9%) and increasing the coverage rates (94.17% vs. 92.67%).

3.2.3 Impact of θ on Networks—We use simulated dataset 2 to examine the impact of θ on the number of connections in the network. We consider different values of the penalty term θ on a natural log scale ranging from −1 to 6, which corresponds to θ ranging from 0.37 to 403. We set $\lambda_{ij} = 0$, if the estimated value is below 10^{-3}. Generally, as θ increases, the number of zeros also increases as shown in Figure 4(A), the biases increase, and the coverage rates decrease. When θ varies within a small range, the change on the number of zero connections is small. As θ is close to 200, all the connections shrink to zero, revealing the impact of the penalty term on network estimation.

3.2.4 The Choice of θ—We consider the same brain network described in the second simulation setting to evaluate the predictive log-likelihood from (13) and the mean-squared error approaches to choose the optimal tuning parameter θ. From coarse to finer parcellation of θ, we find that $\theta = 3$ yields the largest average predictive log-likelihood −12.039 from five-fold cross validation. The smallest MSE, which equals to 0.5497, is achieved when $\theta = 2.8$. We note that when θ varies from 2.0 to 3.6, the difference between the calculated MSE and smallest MSE is less than 0.01. Figure 4(B) shows the trend of the predictive log-likelihood and of the MSE within a small range of values for θ, which reveals that our proposed cross-validation generally succeeds in identifying θ near the MSE-minimizing value.

4. Discussion

We propose a Poisson graphical model to identify functional co-activation patterns and produce undirected brain networks in functional neuroimaging studies. Our method jointly models the region-level numbers of foci that are reported by different independently

performed studies. The estimated sparse covariance matrix between regions is used to construct an undirected brain network associated with a particular brain function. We also perform a permutation test to assess the significance of the functional connectivity between regions. We extend the original multivariate Poisson model by including a penalty term to account for sparsity of the brain network and perform estimation using the EM algorithm. The simulation studies show that our method achieves nearly the nominal coverage rate. We select the shrinkage parameter by optimizing the predictive log-likelihood and the MSE. The results show that the shrinkage method produces more accurate estimates of the covariance and reduces the computation time. We show that the predictive log-likelihood and MSE converge to a similar optimal value in our simulation studies.

Combining the results reported across different brain imaging studies increases the accuracy and power to detect co-activation patterns compared to single analyses. Using the regional foci counts, our method provides a systematic framework to estimate the co-activation patterns, which can be employed to test specific relationships between brain regions of interest or to establish groups of contiguous voxels that show similar functional characteristics and may be treated as prior information in the future studies. Of note, our method is quite different from the meta-analytic connectivity mapping (MACM, Robinson et al., 2010) in that the input data of the two methods are quite different. Our method directly models the regional foci count across studies to build up the co-activation brain network, while the MACM proceeds by creating an activation likelihood estimation map and depends on the choice of seed voxel/region.

One limitation of our model is that it is based on a pre-defined parcellation of the brain (Harvard-Oxford atlas, Makris et al., 2006). This parcellation is widely used in the neuroscience community and, specifically, in a range of neuroimaging analyses for functional networks. One extension would be considering higher order interactions (beyond two-way interactions), which maybe an important aspect of more fully understanding brain networks. Our analysis of 162 functional neuroimaging emotion studies uses foci data reported from different studies. The studies, however, may have different imaging modalities, sample sizes, and criteria and thresholds for testing statistical significance. To obtain a set of standardized count data, we can consider a Poisson covariance regression model estimating the co-activations adjusted for the significance levels and other covariates, which could increase the flexibility of our proposed method.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was partially supported by the National Center for Advancing Translational Sciences of the National Institutes of Health (NIH) under Award Number UL1TR000454 (Kang) and the National Institute of Neurological Disorders and Stroke of the NIH grant U18 NS082143-01 (Bowman).

References

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological). 1995; 57:289–300.

Calhoun VD, Adali T, Pearlson GD, Pekar JJ. A Method for Making Group Inferences from Functional MRI Data Using Independent Component Analysis. Human Brain Mapping. 2001; 14:140–151. [PubMed: 11559959]

Cauda F, Cavanna AE, D'Agata F, Sacco K, Duca S, Geminiani GC. Functional connectivity and coactivation of the nucleus accumbens: a combined functional connectivity and structure-based meta-analysis. Journal of cognitive neuroscience. 2011; 23:2864–2877. [PubMed: 21265603]

Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Methodological). 1977; 39:1–38.

Eickhoff SB, Laird AR, Grefkes C, Wang LE, Zilles K, Fox PT. Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. Human Brain Mapping. 2009; 30:2907–2926. [PubMed: 19172646]

Freeman LC. A set of measures of centrality based on betweenness. Sociometry. 1977; 40:35–41.

Friston KJ, Penny W. Dynamic causal modeling. NeuroImage. 2003; 19:1273–1302. [PubMed: 12948688]

Friston KJ. Functional and effective connectivity: a review. Brain Connectivity. 2011; 1(1):13–36. [PubMed: 22432952]

Huang S, Li J, Sun L, Ye J, Fleisher A, Wu T, Chen K, Reiman E. Learning Brain Connectivity of Alzheimer's Disease by Sparse Inverse Covariance Estimation. NeuroImage. 2010; 50:935–949. [PubMed: 20079441]

Humphries MD, Gurney K, Prescott TJ. The brainstem reticular formation is a small-world, not scale-free, network. Proc Biol Sci. 2006; 273:503–511. [PubMed: 16615219]

Johnson NL, Kotz S, Balakrishnan N. Discrete multivariate distributions. Recherche. 1997; 67:2.

Kang J, Johnson TD, Nichols TE, Wager TD. Meta analysis of functional neuroimaging data via Bayesian spatial point processes. Journal of the American Statistical Association. 2011; 106:124–134. [PubMed: 21706069]

Kano K, Kawamura K. On recurrence relations for the probability function of multivariate generalized Poisson distribution. Communications in statistics-theory and methods. 1991; 20:165–178.

Karlis D. An EM algorithm for multivariate Poisson distribution and related models. Journal of Applied Statistics. 2003; 30:63–77.

Kawamura K. The structure of multivariate Poisson distribution. Kodai Mathematical Journal. 1979; 2:337–345.

Kober H, Barrett LF, Joseph J, Bliss-Moreau E, Lindquist K, Wager TD. Functional grouping and cortical-subcortical interactions in emotion: a meta-analysis of neuroimaging studies. NeuroImage. 2008; 42:998–1031. [PubMed: 18579414]

Kocherlakota, S.; Kocherlakota, K. Bivariate discrete distributions. New York: Marcel Dekker; 1992.

Krummenauer F. Limit theorems for multivariate discrete distributions. Metrika. 1998; 47:47–69.

Loukas S, Kemp CD. On computer sampling from trivariate and multivariate discrete distributions. Journal of Statistical Computation and Simulation. 1983; 17:113–123.

Mahamunulu DM. A note on regression in the multivariate Poisson distribution. Journal of the American Statistical Association. 1967; 61:251–258.

Makris N, Goldstein JM, Kennedy D, Hodge SM, Caviness VS, Faraone SV, et al. Decreased volume of left and total anterior insular lobule in schizophrenia. Schizophr Res. 2006; 83:155–71. [PubMed: 16448806]

McIntosh AR, Gonzalez-Lima F. Structural equation modeling and its application to network analysis in functional brain imaging. Human Brain Mapping. 1994; 2:2–22.

McLachlan, GJ.; Krishnan, T. The EM algorithm and extensions. Wiley; New York: 1997.

Meng XL, Van Dyk D. The EM Algorithm – an Old Folk-song Sung to a Fast New Tune. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 1997; 59:511–567.

Neumann J, Fox PT, Turner R, Lohmann G. Learning partially directed functional networks from meta-analysis imaging data. NeuroImage. 2010; 49:1372–1384. [PubMed: 19815079]

Nielsen FÅ, Hansen LK. Modeling of activation data in the BrainMap database: Detection of outliers. Human Brain Mapping. 2002; 15:146–156. [PubMed: 11835605]

Nielsen FÅ, Hansen LK, Balslev D. Mining for associations between text and brain activation in a functional neuroimaging database. Neuroinformatics. 2004; 2:369–379. [PubMed: 15800369]

Patel R, Bowman FD, Rilling JK. A Bayesian approach to determining connectivity of the human. Human Brain Mapping. 2006; 27:267–276. [PubMed: 16092131]

Patel R, Spreng RN, Turner GR. Functional brain changes following cognitive and motor skills training: a quantitative meta-analysis. Neurorehabilitation and neural repair. 2013; 27:187–199. [PubMed: 23093519]

Postuma RB, Dagher A. Basal ganglia functional connectivity based on a meta-analysis of 126 positron emission tomography and functional magnetic resonance imaging publications. Cerebral cortex. 2006; 16:1508–1521. [PubMed: 16373457]

Robinson JL, Laird AR, Glahn DC, Blangero J, Sanghera MK, Pessoa L, et al. The functional connectivity of the human caudate: an application of meta-analytic connectivity modeling with behavioral filtering. NeuroImage. 2012; 60:117–129. [PubMed: 22197743]

Simpson S, Laurienti P, Bowman FD. Analyzing Complex Functional Brain Networks: Fusing Statistics and Network Science to Understand the Brain. Statistics Surveys. 2013 (in press).

Sporns O, Honey CJ, Kötter R. Identification and classification of hubs in brain networks. PloS one. 2007; 10:e1049. [PubMed: 17940613]

Torta DM, Cauda F. Different functions in the cingulate cortex, a meta-analytic connectivity modeling study. NeuroImage. 2011; 56:2157–2172. [PubMed: 21459151]

Turkeltaub PE, Eden GF, Jones KM, Zeffiro TA. Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. NeuroImage. 2002; 16:765–780. [PubMed: 12169260]

Tziortzi AC, Searle GE, Tzimopoulou S, Salinas C, Beaver JD, Jenkinson M, et al. Imaging dopamine receptors in humans with [(11)C]-(+)-PHNO: Dissection of D3 signal and anatomy. NeuroImage. 2011; 54:264–277. [PubMed: 20600980]

Wager, TD.; Hernandez, L.; Jonides, J.; Lindquist, M. Elements of functional neuroimaging. In: Cacioppo, JT.; Tassinary, LG.; Berntson, GG., editors. Handbook of Psychophysiology. 4. Cambridge University Press; Cambridge: 2007a. p. 19-55.

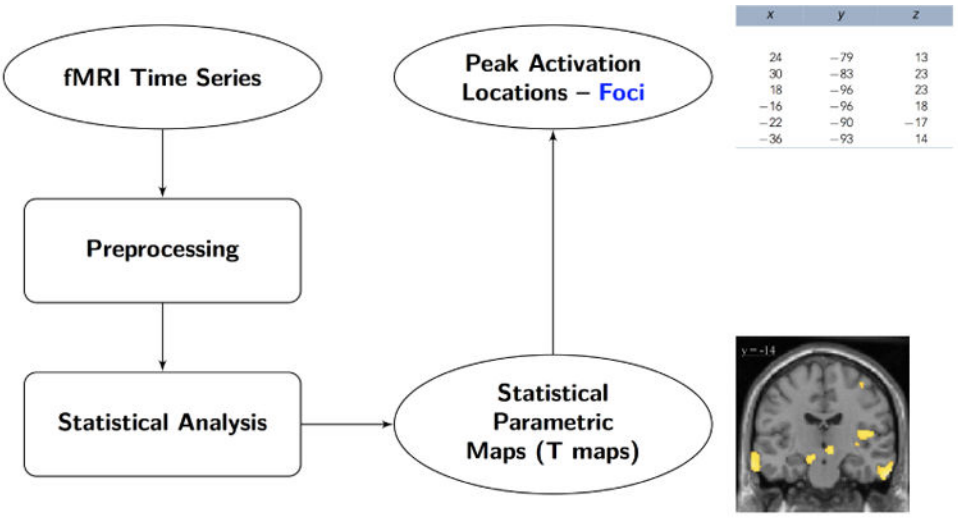
Wager TD, Jonides J, Reading S. Neuroimaging studies of shifting attention: a meta-analysis. NeuroImage. 2004; 22:1679–1693. [PubMed: 15275924]

Wager TD, Lindquist MA, Nichols TE, Kober H, Van Snellenberg J. Evaluating the consistency and specificity of neuroimaging data using meta-analysis. NeuroImage. 2009; 45:S210–S221. [PubMed: 19063980]

Watts DJ, Strogatz SH. Collective dynamics of ‘small-world’ networks. Nature. 1998; 393:440–442. [PubMed: 9623998]

Weiss Y, Freeman WT. Correctness of belief propagation in gaussian graphical models of arbitrary topology. Neural computation. 2001; 13:2173–2200. [PubMed: 11570995]

Yang Y, Kang J. Joint analysis of mixed Poisson and continuous longitudinal data with non ignorable missing values. Computational Statistics & Data Analysis. 2010; 30:63–77.



(A)

Foci from Different Contrasts

Study	Contrast	Emotion	x	y	z
1	1	sad	-59	-14	-1
			...	...	...
			12	-74	-35
	2	happy	48	-78	-1
			...	...	...
			-57	-16	-2
2	3	anger	48	-74	-3
			...	....	...
			-24	0	-12
	4	sad	-57	-13	-2
			...	...	...
			14	-72	-32
	5	disgust	35	-80	15
			...	...	...
			-24	0	-12
...	....	...	...	...	...
164	437	fear	37	26	26
			...	...	...
			-37	12	-17

Regional Foci Count

Contrast	Region 1	Region 2	....	Region 19
1	0	1	....	1
2	1	2	...	1
3	1	0	...	1
4	3	2	...	5
5	2	0	...	3
...	....	....	...	....
437	2	1	...	2

(B)

Figure 1.

(A): Pipeline of analyzing the fMRI time series in a single functional neuroimaging study.

(B): Foci collected from different studies. They are summarized into the regional foci count which are the observed data in our model.

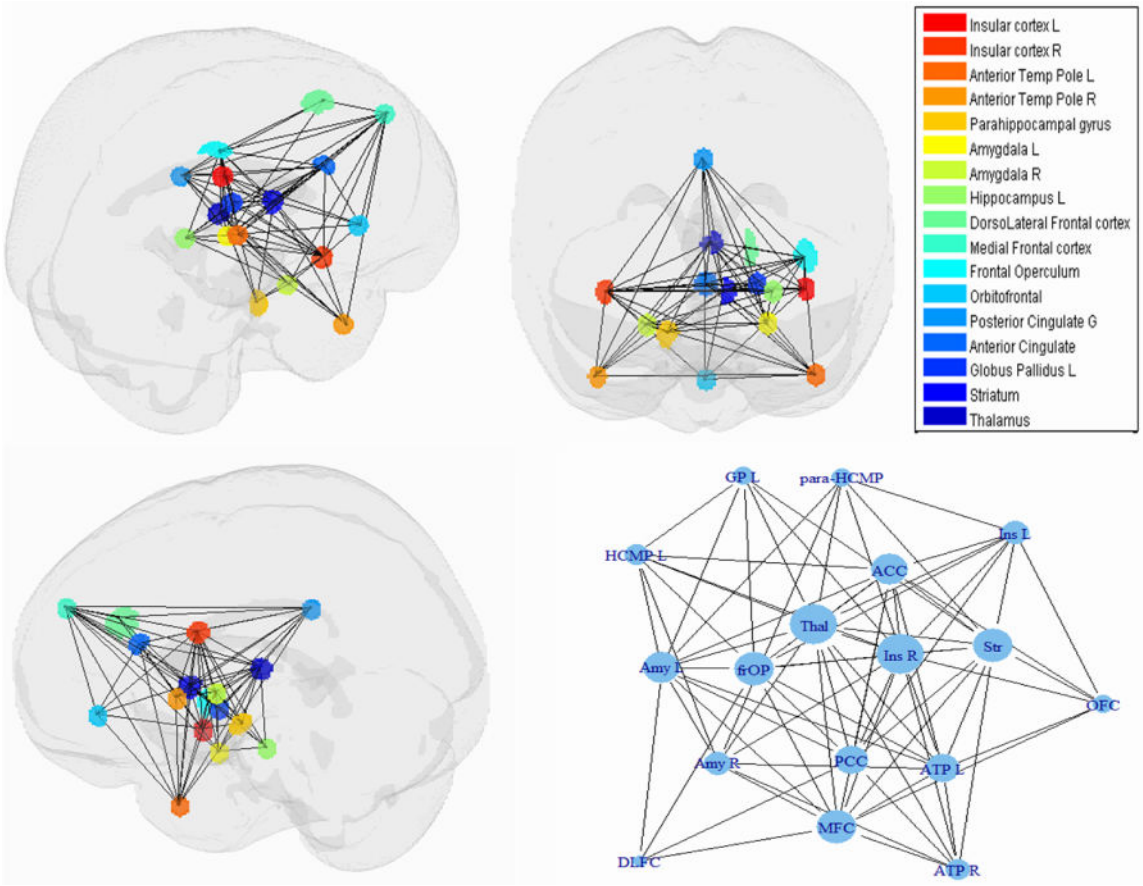


Figure 2.

Three different views of the functional network identified from 162 functional neuroimaging studies with 437 contrasts. 17 ROIs are included in the network. The size of each node in the graphic display represents the degree of the node. ATP – Anterior Temporal Pole, para-HCMP – Parahippocampal, HCMP – Hippocampus, DLFC – DorsoLateral Frontal Cortex, PCC – Posterior Cingulate Cortex.

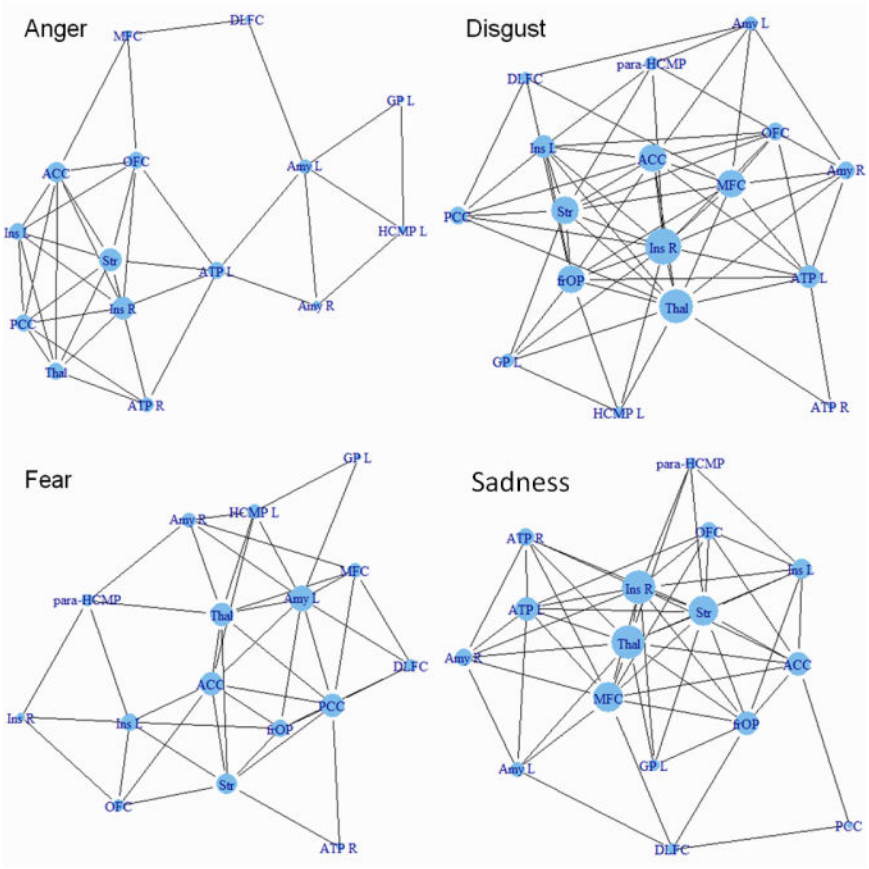


Figure 3.

Functional networks identified for particular emotions from 162 functional neuroimaging studies with 437 contrasts. The negative emotions considered include anger, disgust, fear, and sadness.

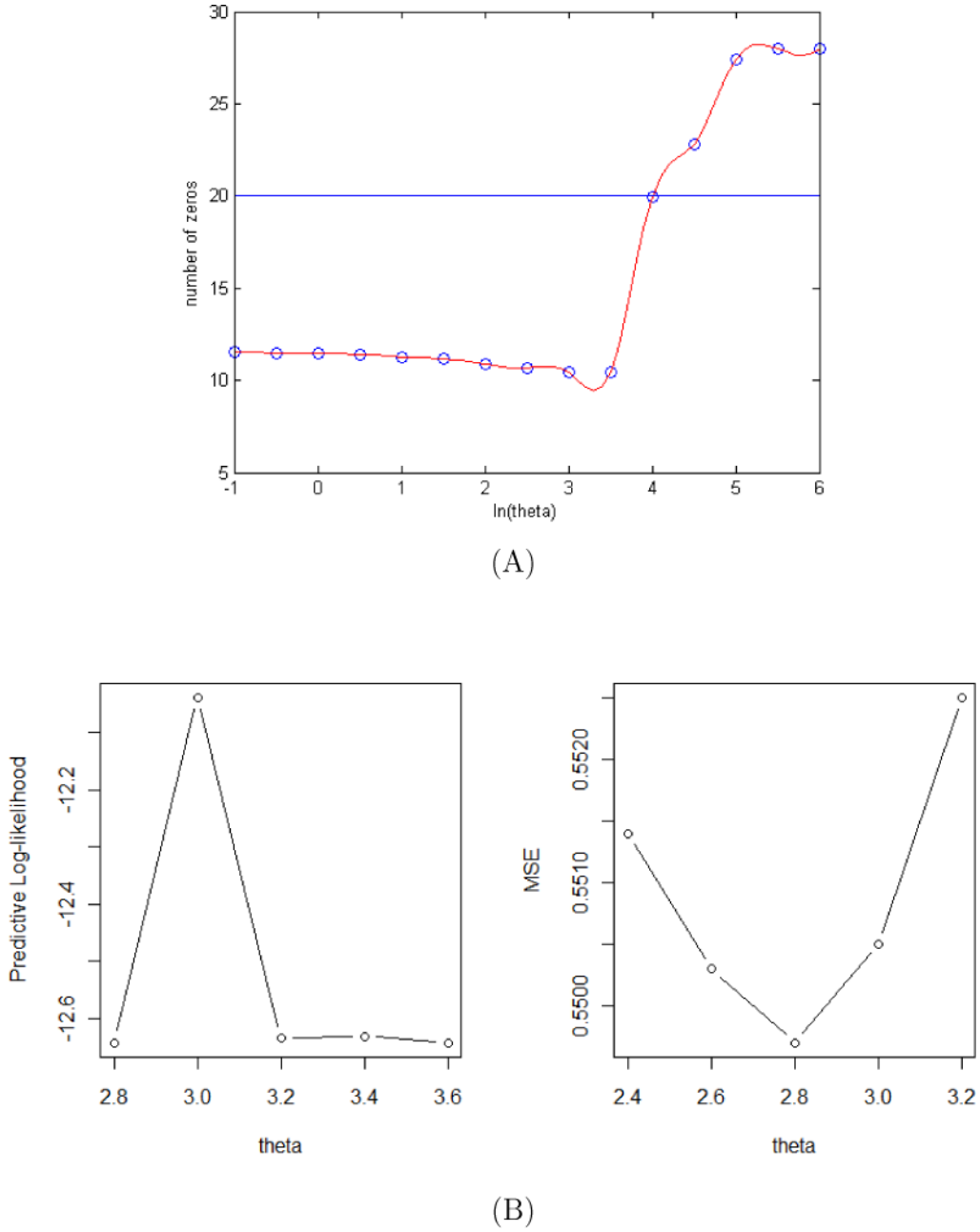


Figure 4.

(A) Change of the average number of zeros detected from 100 simulations vs. ln(θ). The horizontal line indicates the true number of zeros. (B) Relationship between the predictive log-likelihood and θ (left); the predictive log-likelihood achieves the largest value when θ = 3. The relationship between the mean-squared error and θ (right); the MSE achieves smallest value when θ = 2.8.

Table 1

Region pairs with strong co-activations for anger, disgust, fear and sadness. The frequency represents the number of contrasts that have at least one focus in both regions. The percentage refers to the proportion of the contrasts in the corresponding emotion group

Emotion	Region	Region	Frequency (%)
Anger (26 contrasts)	ATP_L	Amy_L	3 (11.5%)
	DLFC	MFC	3 (11.5%)
	OFC	ACC	3 (11.5%)
Disgust (44 contrasts)	OFC	ACC	8 (18.2%)
	Ins_R	OFC	5 (11.4%)
Fear (68 contrasts)	Amy_L	Amy_R	6 (8.8%)
	ACC	Str	5 (7.4%)
Sadness (45 contrasts)	DLFC	MFC	10 (22.2%)
	MFC	Thal	6 (13.3%)

Table 2

Comparison of the bias, with the percentage changes in parenthesis and the coverage rates of $\hat{\boldsymbol{\lambda}}$ for networks with three regions as specified in (15) from 300 simulations between proposed method and covariance approach.

Penalized Multivariate Poisson Model					
Bias (%)			Coverage Rate		
0.0020 (0.20%)	0.0019 (0.06%)	0.0115 (1.15%)	93.33%	95.00%	90.67%
	0.0142 (0.71%)	0.0024 (0.05%)		96.00%	95.00%
		0.0169 (0.56%)			95.00%

Covariance Method					
Bias (%)			Coverage Rate		
0.1657 (16.57%)	0.0624 (2.08%)	0.1381 (13.81%)	92.00%	92.33%	91.67%
	0.0576 (2.88%)	0.1457 (2.91%)		96.33%	91.67%
		0.2747 (9.16%)			92.00%