

Supplementary Materials for Developing Multimodal Dynamic Functional Connectivity as a Neuroimaging Biomarker

Suprateek Kundu¹, Jin Ming¹, and Jennifer Stevens²

¹: Department of Biostatistics and Bioinformatics, Emory University, Atlanta, Ga 30322, USA.

²: Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, Ga 30322, USA.

Imaging Data Acquisition and Pre-processing

Resting-state fMRI collected via a Siemens 3.0 Tesla Magnetom Trio TIM whole-body MR scanner (Siemens, Malvern, PA) included 146 brain volumes for each participant. The voxel size for the resting state fMRI acquisition was $3 \times 3 \times 3 \text{ mm}^3$ with full brain coverage. The other related parameters were: TR/TE=2000/30ms, flip=90°, FOV=200x200 mm, slice thickness/gap = 4mm/mm. We applied several standard preprocessing steps for resting-state fMRI data. It included despiking, slice timing correction, motion correction, MNI(2mm) standard space registration, percent signal change normalization, linear trend removal, movement removal, bandpass filtering from 0.009 to 0.08. and spatial smoothing with a 6mm FWHM Gaussian Kernel. Finally, Power Atlas System was applied to aggregate the data into 264 ROIs. The average measurement for all voxels in each ROI was used as a representative measure for that ROI.

DTI data was acquired on a Siemens 3-T TIM-Trio scanner at Emory University Hospital with the following acquisition parameters: 39x2.5mm thick axial slices, matrix=128x128, FOV=220x220 mm with voxel size = $1.72 \times 1.72 \times 2.5 \text{ mm}^3$, with 60 directions, and a series of 4 b0 images. Standard pre-processing procedures, such as eddy current correction and bias-field correction were applied to the diffusion weighted data. Subsequently, we use the FSL functions `bedpostx` and `probtrackx2` to estimate the distribution of fiber tracts at each voxel and the count of white matter fibers tracts connecting all pairs of brain regions, respectively. In order to obtain the SC scores, we followed the procedure in [11]. In particular, we computed the average of the number of tracts reaching from the first to the second region, and from the second to the first region, divided by the total number of tracts sent out. Fiber tracks passing through gray matter or cerebrospinal fluid were discarded. These SC scores/strengths can be interpreted as the probability of structural connectivity between ROIs, and are made symmetric across edges. The average SC strength across all participants is illustrated as a Figure in Supplementary Materials.

Calculating SC strengths using DTI data

The pre-processed DTI data was used to derive structural connectivity strengths between all pairs of brain regions used for our functional connectivity analysis. In particular, the FSL functions `bedpostx` and `probtrackx2` were used on the pre-processed DTI data to estimate the distribution of fiber tensors at each voxel and the count of white matter fibers tracts connecting all pairs of brain regions, respectively. In order to obtain the SC scores, we compute $p_{jk} = \text{average}\{N_{jk}/N_{j*}, N_{kj}/N_{k*}\} \in (0, 1)$, where N_{jk} equals the number of permissible tracts initiated at region j that pass through region k and N_{j*} refers to the total number of permissible tracts initiated at region j . Fiber tracks passing through gray matter or cerebrospinal fluid are discarded. These SC scores can be interpreted as the probability of structural connectivity between regions j and k , which we often refer to as the strength of SC, and were used to inform the estimation of dynamic functional connectivity under the proposed approach.

Description of Clinical Measures

The presence and severity of PTSD was determined via the PTSD symptom scale (PSS; [8]), is an 18-item self-reported questionnaire that provides a measure of re-experiencing, avoidance, and arousal symptoms that have occurred in the 2 weeks before test administration. Participants were classified as either typical controls or participants with PTSD (PTSD+) based on DSM-IV criteria, consistent with earlier studies ([9],[14]). Participants were classified as PTSD+ if they endorsed the presence of one or more symptom in the re-experiencing cluster (items 1–4); three or more symptoms in the avoidance and numbing symptom cluster (items 5–11); two or more symptoms in the hyperarousal cluster (items 12–17); and symptom duration of 3 months or longer (as measured by question 18), in keeping with DSM-IV criterion for PTSD.

Resilience denotes the ability of an organism to adapt following adversity, such that resilient functioning refers to better mental well-being compared to others with a similar degree of adverse experiences (childhood trauma in our case). We followed the conceptual approach of [13], in which the level of resilient functioning is inferred from the residuals of the relation between childhood trauma severity and psychological functioning. Figure 1 below shows those individuals as functioning better than expected given their childhood trauma experiences using green lines that imply resilient functioning, or individuals who are less resilient (implying vulnerable functioning) via red lines. Because of our focus on trauma and stress-related disorders,

we calculated a resilience score as the variance in PTSD symptoms beyond what would be expected based on childhood trauma exposure, a primary environmental risk factor for PTSD. We estimated the linear association between childhood trauma exposure (CTQ) and PTSD symptom severity (PSS), and extracted standardized residuals from each participant. Positive residual scores indicate higher-than-expected PTSD symptoms, representing psychiatric risk. Negative residual scores indicate lower-than-expected PTSD symptoms, representing psychiatric resilience. We show a graphical representation below, in an expanded GTP sample (n=100 AA women), for further context.

The Beck Depression Inventory (BDI) is a psychometrically validated and widely used self-report inventory of current depressive symptoms [2]. A total of 21 items assess the presence and severity of depressive symptoms over the past 2 weeks rated on a scale of 0 (not at all/never) to 3 (extremely/every day). A total BDI score was calculated by summing all individual items.

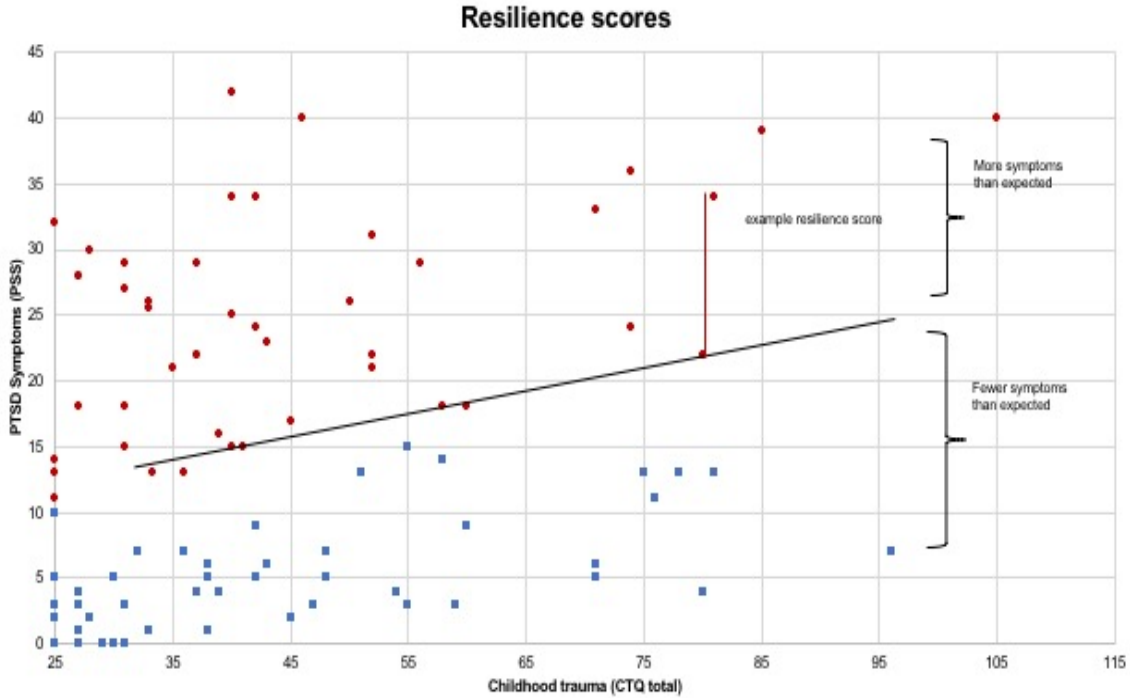


Figure 1. Resilience scores calculated as residuals from a linear regression model that uses PSS scores as the outcome and childhood trauma as the covariate, using the data from GTP study.

Supplementary Method: Multimodal Dynamic Functional Connectivity

Statistical Model

Denote \mathbf{y}_t as the $V \times 1$ vector of spatially distributed fMRI measurements over V voxels or regions of interest (ROI), at the t -th time point ($t = 1, \dots, T$). Denote the SC probability corresponding to the edge (j, l) as p_{jl} where $j \neq l, j, l = 1, \dots, V$, and denote the corresponding SC probability matrix as \mathcal{P} . These SC probabilities are obtained from DTI data, and are made symmetric (i.e. $p_{jl} = p_{lj}$) as in [11]. We specify the following dynamic GGM

$$\mathbf{y}_t \sim N(\boldsymbol{\mu}_t, \boldsymbol{\Omega}_{\mathcal{P}, G_t}^{-1}), t = 1, \dots, T, \quad (1)$$

where $N(\boldsymbol{\mu}, \boldsymbol{\Omega}^{-1})$ refers to a multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Omega}^{-1}$, $\boldsymbol{\Omega}_{\mathcal{P}, G_t}$ denotes the inverse covariance or precision matrix at time point t that depends on the time-varying network G_t characterized by the vertex set \mathcal{V} and edge set E_t , as well as brain SC information \mathcal{P} . The vertex set $\mathcal{V} = \{1, \dots, V\}$ consists of a set of pre-defined

voxels/ROIs or nodes, the edge set E_t contains the set of all edges present in G_t , and $\{\Omega_{\mathcal{P},G_1}, \dots, \Omega_{\mathcal{P},G_T}\}$ encodes the strength of the time varying FC. The goal of change point modeling is to develop an algorithm that is able to learn the best fitting partition of the time course defined by the change points $0 = a_1 < a_2 < \dots < a_K < a_{K+1} = T$ so that G_t remains constant for consecutive time points except for discrete jumps at the change points. These change points are unknown for our problems of interest, and estimated in an unsupervised and data-adaptive manner. For conciseness, we denote $\Omega_{\mathcal{P},k}$ as the constant precision matrix for the k -th state phase corresponding to the interval $(a_{k-1}, a_k]$. throughout the article. A diagrammatic illustration for our approach is provided in Figure 1 in the manuscript.

Structurally Informed Precision Matrix Estimation

Suppose the time course is partitioned into pre-specified non-overlapping intervals or state phases. The precision matrix estimate for the k th state phase is obtained as a MAP estimator that is guided by given SC information via the edge-specific shrinkage parameters in a Bayesian Gaussian graphical model as in [11]. This approach models the edge-specific shrinkage parameters as random variables based on the given SC information and a baseline parameter that is independent of the brain anatomy and controls variations in FC for a given SC level. The prior formulation encourages functional connections for edges corresponding to non-zero structural connectivity as is often observed in literature [11], but allows one to learn the dynamic FC in a data adaptive manner. Incorporating SC knowledge to compute FC is designed to improve the power to detect edges that may otherwise not be identified due to noise in fMRI data, and provides greater accuracy in detecting temporal fluctuations in the network. The final output involves a distinct sparse precision matrix for each state phase, with zeros implying the absence of a connection.

Suppose the time course is partitioned into pre-specified non-overlapping intervals $\mathcal{X}_1, \dots, \mathcal{X}_{K+1}$, where $\mathcal{X}_k = (a_{k-1}, a_k]$ contains n_k time scans such that $\sum_{k=1}^{K+1} n_k = T$. Conditional on these intervals, the sample mean and the covariance matrix for the partition \mathcal{X}_k are given by $\hat{\mu}_k = n_k^{-1} \sum_{t \in \mathcal{X}_k} \mathbf{y}_t$ and $S_k = n_k^{-1} \sum_{t \in \mathcal{X}_k} (\mathbf{y}_t - \hat{\mu}_k)(\mathbf{y}_t - \hat{\mu}_k)^T$ respectively. The precision matrix estimate for the k th state phase is obtained as a MAP estimator under a Bayesian version of the GGM in (1) involving appropriate prior distributions as follows

$$\begin{aligned} \mathbf{y}_t &\sim N(\hat{\mu}_k, \Omega_{\mathcal{P},k}^{-1}) \quad , \quad \pi(\Omega_{\mathcal{P},k} \mid \lambda_k) = C_{\lambda_k, \mathbf{v}}^{-1} \prod_{l=1}^V E(\omega_{k,il}^{\mathcal{P}}; \frac{\mathbf{v}}{2}) \prod_{j<l} DE(\omega_{k,jl}^{\mathcal{P}}; \mathbf{v} \lambda_{k,jl}) 1(\Omega_{\mathcal{P},k} \in M_V^+), \\ \pi(\lambda_k \mid \theta_k, \eta_k, \mathcal{P}) &= C_{\lambda_k, \mathbf{v}} \prod_{j<l} LN(\theta_{k,jl} - \eta_k p_{jl}, \sigma_\lambda^2), \quad t \in \mathcal{X}_k, \quad k = 1, \dots, K+1, \end{aligned} \quad (2)$$

where $\pi(\cdot)$ represents the prior distribution, $C_{\lambda_k, \mathbf{v}}$ is the intractable normalizing constant for $\pi(\Omega_{\mathcal{P},k})$ [11], $\lambda_k = \{\lambda_{k,jl}, j < l, j, l = 1, \dots, V\}$ denotes vector of edge-specific shrinkage parameters in $\pi(\Omega_{\mathcal{P},k})$, \mathbf{v} is the overall penalty parameter (higher values imply greater network sparsity and vice-versa), $E(\cdot)$, $DE(\cdot)$, $LN(\cdot)$, denotes the exponential distribution, double exponential distribution, and log-normal distribution respectively, $1(\cdot)$ is the indicator function, and M_V^+ denotes the collection of all $V \times V$ symmetric and positive definite matrices. We assign a log-normal type prior on λ_k , which restricts the shrinkage parameters to non-negative values and enables us to model the edge-specific shrinkage parameters in terms of the SC strengths. This prior $\pi(\lambda_k \mid \theta_k, \eta_k, \mathcal{P})$ includes unknown hyper-parameters $\theta_k = \{\theta_{k,jl}, j < l, j, l = 1, \dots, V\}$ which represent edge-specific baseline effects that are independent of the given SC knowledge, and η_k that are positive random variables controlling the average effect of SC on FC across the different state-phases. These hyperparameters are unknown and are modeled using priors $\theta_{k,jl} \sim N(\theta_0, \sigma_\theta^2)$, $j < l, j, l = 1, \dots, V$ and $\eta_k \sim Ga(a_\eta, b_\eta)$ respectively under a fully Bayesian specification ($k = 1, \dots, K+1$).

The anatomically informed prior on the shrinkage parameters in (2) specifies a probabilistic relationship between the edge specific shrinkage parameters and the given SC knowledge via $\eta_1, \dots, \eta_{K+1}$. In particular, increasing positive values of η implies an increasing dependence of FC on the given SC, while small values of η implies a negligible relationship. Moreover, the shrinkage parameters are stochastically monotonically decreasing with respect to the SC strength, under the restriction $\eta > 0$. This implies that as the SC strength for the edge (j, l) is increased, the corresponding shrinkage parameter $\lambda_{k,jl}$ will take smaller values in probability, resulting in $\omega_{k,jl}$ values which are away from zero. In other words, the presence (absence) of FC is encouraged for large (small) values of the corresponding SC strength, via the shrinkage parameters $\lambda_1, \dots, \lambda_{K+1}$. The above model specifications are designed to respect the relationship between FC and SC commonly observed in literature [11]. Additionally, the baseline effects $(\theta_1, \dots, \theta_{K+1})$ corresponds to variations in neuronal activity that are independent of the brain anatomy. Overall, increasing (decreasing) absolute values of θ discourages (encourages) the presence of the corresponding edge, independent of the anatomical information. The proposed model enables (a) more flexibility in the FC-SC relationship by allowing the possibility of strong FC corresponding to poor SC strengths, and vice-versa; and (b) heterogeneity in FC across edges which possesses similar SC strength. We note that (2) adapts the approach in [11] to the case of dynamic FC involving state phase specific networks.

The posterior distribution for parameters corresponding to the state phase k is given by

$$P(\Omega_{\mathcal{D},k}, \lambda_k, \theta_k, \eta_k \mid \mathbf{y}_t, t \in \mathcal{X}_k) = Ga(\eta_k; a_\eta, b_\eta) \prod_{t \in \mathcal{X}_k} N(\mathbf{y}_t; \hat{\boldsymbol{\mu}}_k; \Omega_{\mathcal{D},k}) \pi(\Omega_{\mathcal{D},k} \mid \lambda_k) \pi(\lambda_k \mid \theta_k, \eta_k, \mathcal{D}) \pi(\theta_k) \pi(\eta_k), \quad (3)$$

where $\pi(A \mid B)$ denotes the conditional distribution of A given B . One can optimize the log-posterior distribution (obtained by taking the log-transform of the above posterior) to calculate the MAP estimate for the parameters $\Theta_k = (\Omega_k, \lambda_k, \theta_k, \eta_k)$ as $\hat{\Theta}_k = \arg \max_{\Theta_k} l(\hat{\Theta}_k)$, where

$$\begin{aligned} l(\hat{\Theta}_k) = & -\frac{n_k}{2} \log(\det(\Omega_{\mathcal{D},k})) + \frac{1}{2} \text{tr}(S_k \Omega_{\mathcal{D},k}) + \nu \sum_{j < l} \lambda_{k,jl} |\omega_{k,jl}^{\mathcal{D}}| + \sum_{j < l} \frac{(\log(\lambda_{k,jl}) - (\theta_{k,jl} - \eta_k p_{jl}))^2}{2\sigma_\lambda^2} \\ & - (a_\eta - 1) \log(\eta_k) + b_\eta \eta_k + \sum_{j < l} \frac{(\theta_{k,jl} - \theta_0)^2}{2\sigma_\theta^2} - V \log\left(\frac{\nu}{2}\right) + \frac{\nu}{2} \sum_{l=1}^V \omega_{k,l}. \end{aligned} \quad (4)$$

All the parameters in the objective function are updated iteratively until convergence for a fixed value of the sparsity parameter ν . The precision matrix is updated given other parameters using the existing graphical lasso algorithm ([10]), whereas θ_k, η_k , are updated via a closed form expressions and λ_k are updated via a Newton-Raphson step since a closed form solution does not exist. The objective function (4) is optimized over a range of penalty parameter values ν and we choose the value of the penalty parameter that optimizes some goodness of fit score such as the Bayesian Information Criteria (BIC) ([16]) that guard against overfitting. For a given state phase, the update steps are described in Appendix A.

Structurally Informed Change point estimation

The above dynamic precision matrix estimation was conditional on pre-specified change points that are unknown in practice. We now describe a data-adaptive algorithm for change point estimation, which is motivated by the dynamic connectivity regression approach in [4]. The approach involves a greedy partitioning scheme which begins by obtaining estimates for the precision matrix based on the entire time series using the structurally informed precision matrix estimation described above, subject to no change points. The model is fit over varying sparsity levels, and the optimal network corresponding to the minimum Bayesian Information Criteria (BIC) score is chosen. See Supplementary Materials for more details.

Upon completion of this step, the time course is split into two partitions $\{1 : \Delta\}$ and $\{(\Delta + 1) : T\}$, with the understanding that any split of the time course that results in a BIC reduction is acceptable. For each partition, the structurally informed precision matrix estimation is performed independently to obtain the optimal network. This procedure is repeated along the entire time path, with the data partitioned into two subsets corresponding to split points ranging from $\Delta + 1$ to $T - \Delta + 1$. In order to ensure reliable estimation of the network in each state phase we fix the minimum number of time scans (usually 5-10) between consecutive change points. The partition with the smallest combined BIC score is chosen and, if this score is less than the BIC score for the entire data set, the corresponding split point is identified as the first change point. The procedure continues by recursively applying the same method to each individual partition element until they can no longer be split any further, i.e. no additional reduction in BIC is seen. As the final output, the algorithm will have split the entire time course into non-overlapping partitions. The number and location of the change points are thus determined in a data adaptive manner and guided by SC knowledge, due to the structurally informed precision matrix estimation component. Although motivated by [5], the proposed approach is distinct in terms of incorporating the brain SC information to compute dynamic FC, as well as being scalable to high-dimensional networks via a novel sub-network sampling scheme described below.

A sub-network sampling scheme

In our experience, the change point estimation approach used described above may not yield accurate change point detection results under rapid transitions of the network, and unfortunately it is not be scalable to a large number of nodes, as per our experience with extensive simulations. To overcome such difficulties, we propose a novel heuristic sub-network sampling strategy, which is based on the key observation that alterations in the network represented by change points may be detected using only a subset of the nodes in the network, as long as the edges connected to at least one of these subset of nodes undergoes temporal connectivity changes. One potential pitfall of this strategy is that the changes corresponding to the subset of nodes may not be strong enough compared to the overall changes in the network at a given time point, in which case it may not be detected. However, by repeatedly sampling random subset of nodes and computing the connectivity change points under each of these subsets, the hope is that the true underlying change points will be detected by a large proportion of these sub-networks. By drawing an adequate number of sub-networks and computing change points for each sub-network, one can ensure that all important connectivity changes between subsets of nodes in the network are accounted for.

Our strategy is to apply the proposed multimodal dynamic FC approach to a randomly sampled subset of nodes $\mathcal{V}^* \subset \mathcal{V}$, which yields a set of change points associated with the corresponding dynamic sub-network. We then repeat this process over multiple randomly sampled sub-networks (usually having 10 nodes each), thereby generating a collection of sets of change

points. Then, the frequency of each time point being identified as a change point over the collection of sampled sub-networks is computed. Finally, a systematic thresholding approach (described in Supplementary Materials) is proposed to detect the important change points as those which show up most frequently across the sub-networks. A useful feature of this thresholding mechanism is that one can obtain both discrete jumps in FC that is the hallmark of existing change point models, as well as estimate transition periods comprising a collection of successive time scans, which is an added novelty of the approach. We note that by increasing the number of sub-network samples, the accuracy for change point detection is expected to increase although it comes at a cost of increased computation time. However, the approach can be parallelized over sub-networks, thereby resulting in computational speed-ups, when needed.

Supplementary Algorithm: Computation for Structured Precision Matrix Estimation

We use the following update steps corresponding to the k th state phase ($k = 1, \dots, K + 1$) iteratively till we achieve convergence in the objective function. We denote $\alpha = \log(\lambda)$ in what follows.

1. Update the precision matrices: We update the precision matrix for the k th stage phase in the $(m + 1)$ -th step as

$$\hat{\Omega}_{\mathcal{P},k}^{(m+1)} = \arg \min_{\Omega} \left\{ -\log \det(\Omega) + \text{tr}(S_k \Omega) + \frac{V}{2} \sum_{j < l} \exp^{\alpha_{k,jl}^{(m+1)}} |\omega_{jl}| + \frac{V}{2} \sum_l \omega_{ll} \right\}, \quad (5)$$

where S_k is the empirical covariance matrix for the k -th bin, that is defined at the start of Section 2.1. The above can be solved using a approximation solver, QUIC ([12]), available in R.

2. Update the baseline effects via the closed form expression

$$\hat{\theta}_{k,jl}^{(m+1)} = \frac{\sigma_{\theta}^2 \left(\alpha_{k,jl}^{(m)} + \eta_k^{(m)} p_{jl} \right) + \sigma_{\lambda}^2 \theta_0}{\sigma_{\theta}^2 + \sigma_{\lambda}^2}.$$

3. Update η_k via the closed form expression below where $\gamma = \frac{\sum_{l < m} p_{lm}^2}{\sigma_{\lambda}^2}$, $\rho = -\frac{a\eta - 1}{\sigma_{\lambda}^2}$

$$\hat{\eta}_k^{(m+1)} = \frac{-\beta_k^{(m)} + \sqrt{(\beta_k^{(m)})^2 - 4\gamma\rho}}{2\gamma}, \text{ where } \beta_k^{(m)} = b_{\eta} + \frac{\sum_{l < m} \alpha_{k,jl}^{(m)} p_{jl}}{\sigma_{\lambda}^2} - \frac{1}{\sigma_{\lambda}^2} \sum_{j < l} \theta_{k,jl}^{(m)} p_{jl}.$$

4. Update α_k when we got $\mathbf{Y}, \hat{\Omega}_{\mathcal{P},k}^{(m)}, \boldsymbol{\mu}_k^{(m+1)}$ and $\eta_k^{(m+1)}$ with the following formula:

$$\hat{\alpha}_k^{(m+1)} = \arg \min_{\alpha} \nu \sum_{j < l} e^{\alpha_{k,jl}} |\omega_{k,jl}^{(m)}| + \sum_{j < l} \frac{(\alpha_{k,jl} - (\hat{\theta}_{k,jl}^{(m+1)} - \eta_k^{(m+1)}) p_{jl})^2}{2\sigma_{\lambda}^2}$$

As there is no closed form solution, we implement a Newton Raphson solver to find the optimal value. The above formula could to re-expressed as:

$$\arg \min_{\alpha} \exp(\boldsymbol{\alpha}_k)' |\tilde{\omega}_k^{(m+1)}| - \frac{1}{2\sigma_{\lambda}^2} (\boldsymbol{\alpha}_k - (\tilde{\theta}_k^{(m+1)} - \eta_k^{(m+1)}) \tilde{P})' (\boldsymbol{\alpha}_k - (\tilde{\theta}_k^{(m+1)} - \eta_k^{(m+1)}) \tilde{P})$$

where $\boldsymbol{\alpha}_k = \{\alpha_{k,12}, \alpha_{k,13} \dots \alpha_{k,(p-1)p}\}$, \tilde{P} denotes the upper diagonal elements of the structural connectivity matrix P . $\tilde{\omega}_k$ and $\tilde{\theta}_k$ denotes the upper diagonal elements of Ω_k and $\boldsymbol{\theta}_k$ respectively. $\exp(\boldsymbol{\alpha}_k)$ is the element wise exponential for each element of $\boldsymbol{\alpha}_k$. We could only focus upon the upper diagonal elements of $\boldsymbol{\alpha}_k$ because Ω_k is symmetric and we do not shrink diagonal elements.

Based on step size Δ , the Newton Raphson updating equation is:

$$\boldsymbol{\alpha}_k^{(m+1)} = \boldsymbol{\alpha}_k^{(m)} - \Delta g(\boldsymbol{\alpha}_k^{(m)}) H(\boldsymbol{\alpha}_k^{(m)})^{-1}$$

where $g(\boldsymbol{\alpha}_k^{(m)}) = v\sigma_\lambda^2 D_{|\omega_k^{(m)}|} e^{\boldsymbol{\alpha}_k^{(m)}} + [\boldsymbol{\alpha}_k^{(m)} - (\boldsymbol{\theta}_k^{(m+1)} - \eta_k^{(m+1)\bar{P}})]$,
and $H(\boldsymbol{\alpha}_k^{(m)}) = v\sigma_\lambda^2 D_{|\omega_k^{(m)}|} D_{|e^{\boldsymbol{\alpha}_k^{(m)}}|} + \mathbf{I}$. $D_{|\omega_k^{(m)}|}$ is a $\frac{V(V-1)}{2} \times \frac{V(V-1)}{2}$ diagonal matrix with elements as the upper triangular elements of $\boldsymbol{\Omega}_k^{(m)}$, and similarly for $D_{|e^{\boldsymbol{\alpha}_k^{(m)}}|}$. Based on simple calculation, H is diagonal matrix and could be easily inverted. The step size (Δ) is searched using a back tracking line search for each update of $\boldsymbol{\alpha}_k$ as in [3].

Supplementary Algorithm: Thresholding Strategy to Derive Change Points and Transition Periods

Suppose we sample J sub-networks containing v^* nodes each, where $v^* < V$ ROIs are selected randomly from \mathcal{V} for each sub-network (the number of nodes may also be made different across sub-networks in principle). We then run the proposed mDFC method for each sub-network to obtain a set of change points for the j -th sub-network as τ_j^* ($j = 1, \dots, J$), with the understanding that the number and location of change points may vary across sub-networks. The set of estimated change points across all the sub-networks is aggregated to obtain the set of all identified change points $\tau^* = \cup_{j=1}^J \tau_j^*$. Moreover, we also calculate the corresponding frequencies with which each time point was identified as a change point across the sub-networks. This is denoted by $\mathbf{w} = (w_1, \dots, w_T)$, with $w_t = 0$ for those time points that are not identified as change points under any sub-networks. The frequencies \mathbf{w} can be interpreted as an approximate probability for each time point to be a change point, and any time point t for which $w_t > 0$ is treated as a potential change point. Given τ^* and \mathbf{w} , we then apply a grouping approach to find out representative change points within this set, by identifying clusters of change points. In particular, two estimated change points are grouped into one cluster if they are consecutive time points or spaced one time point apart. This grouping method results in $L < T$ distinct groups or clusters of change points (see Figure 2 for a visualization pertaining to our simulation study).

Each cluster of change points is representative of a transition period for the network, that is more consistent with the slower timescale of the haemodynamic activity. In order to eliminate false clusters of change points, we adopt the following thresholding mechanism. First, the overall frequency of a particular cluster of change points is taken as the sum of the frequencies for all the change points within that cluster. Subsequently, all clusters having a combined frequency below a certain threshold are identified as false positives and eliminated. Based on extensive empirical experiments, we propose a threshold of $0.3J$ for the combined frequency for clusters, which proved to be a good choice in terms of identifying true change points and eliminating false positives. Alternatively, the number of clusters may also be determined in a data adaptive manner using some goodness of fit measure such as BIC. Once we get the final groups, one can designate a representative single change point for each cluster as the median time point for that cluster. Thus, one is able to obtain both discrete jumps in FC that is the hallmark of existing change point models, as well as estimate transition periods, which is an added novelty of the approach.

Description of Network Metrics

A description of the network metrics used for the analysis is provided below.

Global Efficiency A measure of information transmission across the entire brain calculated by averaging the inverse shortest path lengths across all brain nodes. Higher values of global efficiency indicate more efficient information transmission.

Local Efficiency A node-specific version of global efficiency. We examined the average local efficiency over all nodes in the brain.

Mean Clustering Coefficient (MCC) A measure of the interconnectedness of the brain network calculated by counting how many of a brain node's neighbors are also neighbors of each other, averaged over all nodes.

Small-Worldedness A measure of whether or not the brain exhibits small world properties calculated by examining the ratio of normalized mean clustering coefficient to normalized characteristic path length, $SW = \frac{MCC/MCC_0}{CPL/CPL_0}$. Here MCC_0, CPL_0 , refer to the metrics corresponding to a baseline distribution for each cohort and visit by generating 1000 surrogate random networks with the same connection density as the estimated graph for the cohort/visit and then calculating the average CPL and MCC for these surrogate graphs. The ratio of each metric to the average over the surrogate graphs is then taken as the normalized metric. A ratio $SW > 1$ indicates that the estimated brain network exhibits more small-worldedness than a random network.

Validation Studies

Validation Studies Using Simulated Data

We conducted extensive simulation studies to evaluate the performance of the proposed approach, under different network structures. In the first set of simulations (Scenario I), we generated data from an underlying change point model with three change points (i.e. four state phases). A network and the corresponding precision matrix was constructed at each time point (described below), and these were constant within each state phase. The measurements were generated under a Gaussian distribution characterized by the time-varying precision matrix. We generated data for $V = 20, 50$ regions and with $T = 300, 500$ time points. In a second set of simulations (Scenario II), we allow the network to change more slowly over time, that is more consistent with the timescale of the haemodynamic activity. In particular, instead of three change points as in the first scenario, we now have three transition periods, each comprising seven consecutive time points. A certain percentage of the edges are flipped from the network at the previous time point to obtain the modified network for the next time point within each transition period. The network is assumed to be constant between two consecutive transition periods. This scenario is more challenging since the network changes multiple times over the course of the experiment. The goal of this experiment is to investigate if the proposed approach can detect the transition periods and whether it can approximate the true dynamic network sufficiently well when the underlying assumptions of the proposed model may not hold. We generate data using three different types of networks (details provided in Supplementary Materials). In addition to Scenarios I and II, we also reported results under Erdos-Renyi networks for 100 nodes with three true change points, to test the performance in higher dimensions, and investigated the scenario involving a large number of change points (10) with 50 nodes. These challenging settings help us evaluate the performance of the methods for high dimensions and large number of fluctuations in the dynamic network. For each simulation setting, 25 replicates were used. Using the simulated data, we investigated the ability of the proposed approach in terms of its' ability to recover the true dynamic network in comparison with competing approaches such as siGGM and DCR. We note that while both the change point detection and the network estimation performance were reported under the DCR approach and the SC naive version of the proposed method, the siGGM only reports the network estimation performance, since it does not account for dynamic changes.

Networks Generated for Validation Experiments

We generated the functional network of the first bin using three different network structures: (a) Erdos-Renyi random graph [7] that has the same probability for all connections; (b) scale-free graph that uses the preferential attachment model of [1]; and (c) small-world random graph that was obtained using the [15] model, and which is motivated by the characteristics of real life brain networks derived from fMRI data. Two different average densities for the network were used, 0.15 and 0.3, which indicate sparse connectivity patterns in brain organization [6]. Once the functional network for the first state phase was generated, the networks of the other three bins are generated by flipping a certain proportion of the edges of the first bin independently. This implies a fixed proportion of edges in the functional network for the first bin were changed to be absent in the second functional network, whereas a fixed proportion of pairs of nodes that were not connected in the first bin became connected in the second bin. This procedure is repeated independently for all the bins to generate distinct networks for different state phases that have similar densities and also share common patterns, while also exhibiting inherent differences.

Performance Assessment for Validation Studies

We evaluate the performance of the proposed method with respect to two aspects: the ability to estimate true change points and accurately estimate the network. For data from Scenario I involving discrete connectivity jumps, the estimated change point is determined to be a true positive when the temporal distance between the true and estimated change point is less or equal than 2. For the case of data generated using Scenario II involving transition periods of seven consecutive change points, we denote an estimated change point to be true positive if it lies within the ± 2 from the center of the transition period. We report the proportion of true change points detected, and the number of false change points reported, averaged over all replicates. The performance of the graph estimation was assessed by comparing the estimated network with the true network at each time point, using the area under the curve (AUC) for the receiver operating characteristic (ROC) over different network sparsity levels. We also look at sensitivity and specificity for the optimal network obtained by minimizing the BIC score. Sensitivity measures the power to detect true connections and is equal to the proportion of true edges that are successfully detected in the estimated network. Specificity represents the proportion of the absent edges that were successfully detected as absent, and is an indicator of how well the false discoveries are controlled for. Higher values of AUC, sensitivity and specificity imply a more accurate network estimation. All metrics are reported after averaging over all time points and replicates.

Results from Validation Studies

Figures 2, 3 and 4 report the results for the validation studies corresponding to Scenarios I and II. We note that we could not report results under DCR for the 100 node scenario since the DCR is not scalable to high dimensional networks. Multimodal

Dynamic FC Provides Near Perfect Estimation for Network Changes Figure 2(a)-(b) provides a visual illustration of the change point detection performance under the first scenario, which plots the frequency with which each time point is detected as a change point under the sub-network sampling scheme. In this Figure, the peaks in frequency are seen to concentrate around the true jumps which illustrate the ability to accurately detect change points. The results in Figures 8 and 9 suggest that the proposed approach is able to detect essentially almost all the true change points under both scenarios, as computed via the proportion of change points detected. Moreover, the proportion of false change points detected (computed as the number of falsely detected change points over replicates) is close to 0 or negligible. In contrast, the detection of false change points is much higher under DCR, and it has poor performance in terms of detecting the true change points. In fact, the multimodal dynamic FC approach has a significantly higher proportion of change points detected, and significantly lower number of false positives. Additional experiments (not presented here) reveal that the performance of DCR improves when the total number of time points in the experiment, along with the distance between consecutive change points is increased. However, for practical experiments with a few hundred time points, the DCR approach seems to fail in terms of change point estimation. The proposed approach, which incorporates SC knowledge and espouses a novel subnetwork sampling scheme, performs considerably better in terms of detecting the true change points while incurring minimal false positives. Figure 2(b) presents the change point detection results for the high-dimensional case of $V = 100$ for data simulated under the Erdos-Renyi network with three jump points. The Figure clearly depicts high frequencies for change point detection around the true change points, thereby suggesting that the proposed approach could successfully detect true change points based on the sub-network sampling mechanism. Finally, the results for the case involving a higher number of true change points (10) with $V = 50$ regions and $T = 500$ time points, is presented in Figure 2(c), which clearly shows peaks under the sub-network sampling scheme around all of the 10 true change points, thereby indicating the power of the proposed approach in detecting discrete jumps. In order to accurately detect all the true change points, we increased the number of sub-network samples to 100 for this case. For this case, the DCR approach fails to detect an overwhelming majority of the jumps (results not reported). In addition, we assessed the performance of the proposed approach when both the number of change points and ROIs increases. Simulation results (not presented here) shows that as long as the number of sub-networks is large enough and true change points are not exceedingly close together, our method successfully detects peaks around the true change points, under the sub-network sampling scheme. The computation time for each sub-network is reasonable as long as the number of nodes in the sub-network is moderate - see Figure 2(d) for more details. We conjecture that an increasing number of sub-networks will be required for a good performance as the number of nodes in the network, as well as the number of true change points is increased. In summary, using a combination of a powerful sub-sampling scheme and incorporating prior SC knowledge, the proposed method is shown to provide vast improvements over existing approaches in literature in terms of detecting underlying true network changes.

Multimodal Dynamic FC Results in Higher Network Estimation Accuracy The results of graph estimation are reported in Figures 8 and 9. Compared to the competing methods, the proposed method consistently has a significantly higher area under the Receiver Operating Characteristic Curve (AUC) value for network estimation. In addition, it has a higher or comparable sensitivity and consistently has a significantly higher specificity, implying lower false positives in graph estimation and suitable power to detect true positives. In several cases, both the sensitivity and specificity under the multimodal dynamic FC were higher compared to the other two methods. A higher AUC, along with higher specificity and comparable sensitivity, illustrates the ability of the multimodal dynamic FC approach to better control for false positives while having a similar or higher power to detect true signals. We note that the accurate estimation of the dynamic network under the multimodal FC approach illustrates the importance of using brain SC to guide the estimation of dynamic FC.

Once the networks are generated, the corresponding precision matrix at each time point was constructed by randomly generating off-diagonals from a uniform(-1,1) distribution corresponding to the edges in the network at that time point, while fixing those off-diagonals corresponding to absent edges to be zero. The diagonal elements for the matrix are then recalibrated by summing the absolute values of the off-diagonals in the corresponding row/column, and adding a positive constant to this sum, so as to obtain a diagonally dominant matrix that is positive definite. Under a GGM, the fMRI observation at a particular time point was generated from a Gaussian distribution having the time-dependent precision matrix constructed above. Moreover, we generated the SC information based on the true FC, as in [11]. We specify that most edges that had strong FC (≥ 0.5) in first bin also have a strong SC (≥ 0.5), while edges with weak SC (≤ 0.5) could have either strong or weak FC. We note that since the networks in the latter bins were obtained by flipping edges in the network corresponding to the first bin, it is very likely to have a non-trivial number of edges with small SC but large FC and large SC with weak FC, which violates the modeling assumptions and presents a challenging scenario.

References

1. Barabási, A.-L., Albert, R. & Jeong, H. Mean-field theory for scale-free random networks. *Physica A: Statistical Mechanics and its Applications* **272**, 173–187 (1999).
2. Beck A., T., Steer R., A. & Brown G., K. *Manual for the Beck depression inventory-II* (Psychological Corporation, 1996).

3. Chang, C., Kundu, S. & Long, Q. Scalable Bayesian variable selection for structured high-dimensional data. *Biometrics* **74**, 1372–1382 (2018).
4. Cribben, I., Haraldsdottir, R., Atlas, L. Y., Wager, T. D. & Lindquist, M. A. Dynamic connectivity regression: determining state-related changes in brain connectivity. *Neuroimage* **61**, 907–920 (2012).
5. Cribben, I., Wager, T. & Lindquist, M. Detecting functional connectivity change points for single-subject fMRI data. *Frontiers in computational neuroscience* **7**, 143 (2013).
6. Eavani, H. *et al.* Identifying sparse connectivity patterns in the brain using resting-state fMRI. *Neuroimage* **105**, 286–299 (2015).
7. Erdős, P. & Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci* **5**, 17–60 (1960).
8. Falsetti, S. A., Resnick, H. S., Resick, P. A. & Kilpatrick, D. G. The Modified PTSD Symptom Scale: A brief self-report measure of posttraumatic stress disorder. *The Behavior Therapist* (1993).
9. Fani, N. *et al.* Attention bias toward threat is associated with exaggerated fear expression and impaired extinction in PTSD. *Psychological medicine* **42**, 533–543 (2012).
10. Friedman, J., Hastie, T. & Tibshirani, R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9**, 432–441 (2008).
11. Higgins, I. A., Kundu, S. & Guo, Y. Integrative Bayesian analysis of brain functional networks incorporating anatomical knowledge. *Neuroimage* **181**, 263–278 (2018).
12. Hsieh, C.-J., Dhillon, I., Ravikumar, P. & Sustik, M. Sparse inverse covariance matrix estimation using quadratic approximation. *Advances in neural information processing systems* **24**, 2330–2338 (2011).
13. Ioannidis, K., Askelund, A. D., Kievit, R. A. & Van Harmelen, A.-L. The complex neurobiology of resilient functioning after childhood maltreatment. *BMC medicine* **18**, 1–16 (2020).
14. Jovanovic, T. *et al.* Impaired fear inhibition is a biomarker of PTSD but not depression. *Depression and anxiety* **27**, 244–251 (2010).
15. Watts, D. J. & Strogatz, S. H. Collective dynamics of ‘small-world’ networks. *nature* **393**, 440 (1998).
16. Yuan, M. & Lin, Y. Model selection and estimation in the Gaussian graphical model. *Biometrika* **94**, 19–35 (2007).

Supplementary Table: Coordinates for ROIs in Power Atlas

The coordinates for the remaining ROIs are listed in the Table below.

ROI	X	Y	Z	Modules	ROI	X	Y	Z	Modules
1	-25	-98	-12	Unspecified	46	66	-8	25	Sensory/somatomotor
2	27	-97	-13	Unspecified	47	-3	2	53	Cingulo-opercular
3	24	32	-18	Unspecified	48	54	-28	34	Cingulo-opercular
4	-56	-45	-24	Unspecified	49	19	-8	64	Cingulo-opercular
5	8	41	-24	Unspecified	50	-16	-5	71	Cingulo-opercular
6	-21	-22	-20	Unspecified	51	-10	-2	42	Cingulo-opercular
7	17	-28	-17	Unspecified	52	37	1	-4	Cingulo-opercular
8	-37	-29	-26	Unspecified	53	13	-1	70	Cingulo-opercular
9	65	-24	-19	Unspecified	54	7	8	51	Cingulo-opercular
10	52	-34	-27	Unspecified	55	-45	0	9	Cingulo-opercular
11	55	-31	-17	Unspecified	56	49	8	-1	Cingulo-opercular
12	34	38	-12	Unspecified	57	-34	3	4	Cingulo-opercular
13	-7	-52	61	Sensory/somatomotor	58	-51	8	-2	Cingulo-opercular
14	-14	-18	40	Sensory/somatomotor	59	-5	18	34	Cingulo-opercular
15	0	-15	47	Sensory/somatomotor	60	36	10	1	Cingulo-opercular
16	10	-2	45	Sensory/somatomotor	61	32	-26	13	Auditory
17	-7	-21	65	Sensory/somatomotor	62	65	-33	20	Auditory
18	-7	-33	72	Sensory/somatomotor	63	58	-16	7	Auditory
19	13	-33	75	Sensory/somatomotor	64	-38	-33	17	Auditory
20	-54	-23	43	Sensory/somatomotor	65	-60	-25	14	Auditory
21	29	-17	71	Sensory/somatomotor	66	-49	-26	5	Auditory
22	10	-46	73	Sensory/somatomotor	67	43	-23	20	Auditory
23	-23	-30	72	Sensory/somatomotor	68	-50	-34	26	Auditory
24	-40	-19	54	Sensory/somatomotor	69	-53	-22	23	Auditory
25	29	-39	59	Sensory/somatomotor	70	-55	-9	12	Auditory
26	50	-20	42	Sensory/somatomotor	71	56	-5	13	Auditory
27	-38	-27	69	Sensory/somatomotor	72	59	-17	29	Auditory
28	20	-29	60	Sensory/somatomotor	73	-30	-27	12	Auditory
29	44	-8	57	Sensory/somatomotor	74	-41	-75	26	Default-mode
30	-29	-43	61	Sensory/somatomotor	75	6	67	-4	Default-mode
31	10	-17	74	Sensory/somatomotor	76	8	48	-15	Default-mode
32	22	-42	69	Sensory/somatomotor	77	-13	-40	1	Default-mode
33	-45	-32	47	Sensory/somatomotor	78	-18	63	-9	Default-mode
34	-21	-31	61	Sensory/somatomotor	79	-46	-61	21	Default-mode
35	-13	-17	75	Sensory/somatomotor	80	43	-72	28	Default-mode
36	42	-20	55	Sensory/somatomotor	81	-44	12	-34	Default-mode
37	-38	-15	69	Sensory/somatomotor	82	46	16	-30	Default-mode
38	-16	-46	73	Sensory/somatomotor	83	-68	-23	-16	Default-mode
39	2	-28	60	Sensory/somatomotor	84	-58	-26	-15	Unspecified
40	3	-17	58	Sensory/somatomotor	85	27	16	-17	Unspecified
41	38	-17	45	Sensory/somatomotor	86	-44	-65	35	Default-mode
42	-49	-11	35	Sensory/somatomotor	87	-39	-75	44	Default-mode
43	36	-9	14	Sensory/somatomotor	88	-7	-55	27	Default-mode
44	51	-6	32	Sensory/somatomotor	89	6	-59	35	Default-mode
45	-53	-10	24	Sensory/somatomotor	90	-11	-56	16	Default-mode

ROI	X	Y	Z	Modules	ROI	X	Y	Z	Modules
91	-3	-49	13	Default-mode	136	4	-48	51	Unspecified
92	8	-48	31	Default-mode	137	-46	31	-13	Default-mode
93	15	-63	26	Default-mode	138	-10	11	67	Ventral-attention
94	-2	-37	44	Default-mode	139	49	35	-12	Default-mode
95	11	-54	17	Default-mode	140	8	-91	-7	Unspecified
96	52	-59	36	Default-mode	141	17	-91	-14	Unspecified
97	23	33	48	Default-mode	142	-12	-95	-13	Unspecified
98	-10	39	52	Default-mode	143	18	-47	-10	Visual
99	-16	29	53	Default-mode	144	40	-72	14	Visual
100	-35	20	51	Default-mode	145	8	-72	11	Visual
101	22	39	39	Default-mode	146	-8	-81	7	Visual
102	13	55	38	Default-mode	147	-28	-79	19	Visual
103	-10	55	39	Default-mode	148	20	-66	2	Visual
104	-20	45	39	Default-mode	149	-24	-91	19	Visual
105	6	54	16	Default-mode	150	27	-59	-9	Visual
106	6	64	22	Default-mode	151	-15	-72	-8	Visual
107	-7	51	-1	Default-mode	152	-18	-68	5	Visual
108	9	54	3	Default-mode	153	43	-78	-12	Visual
109	-3	44	-9	Default-mode	154	-47	-76	-10	Visual
110	8	42	-5	Default-mode	155	-14	-91	31	Visual
111	-11	45	8	Default-mode	156	15	-87	37	Visual
112	-2	38	36	Default-mode	157	29	-77	25	Visual
113	-3	42	16	Default-mode	158	20	-86	-2	Visual
114	-20	64	19	Default-mode	159	15	-77	31	Visual
115	-8	48	23	Default-mode	160	-16	-52	-1	Visual
116	65	-12	-19	Default-mode	161	42	-66	-8	Visual
117	-56	-13	-10	Default-mode	162	24	-87	24	Visual
118	-58	-30	-4	Default-mode	163	6	-72	24	Visual
119	65	-31	-9	Default-mode	164	-42	-74	0	Visual
120	-68	-41	-5	Default-mode	165	26	-79	-16	Visual
121	13	30	59	Default-mode	166	-16	-77	34	Visual
122	12	36	20	Default-mode	167	-3	-81	21	Visual
123	52	-2	-16	Default-mode	168	-40	-88	-6	Visual
124	-26	-40	-8	Default-mode	169	37	-84	13	Visual
125	27	-37	-13	Default-mode	170	6	-81	6	Visual
126	-34	-38	-16	Default-mode	171	-26	-90	3	Visual
127	28	-77	-32	Default-mode	172	-33	-79	-13	Visual
128	52	7	-30	Default-mode	173	37	-81	1	Visual
129	-53	3	-27	Default-mode	174	-44	2	46	Fronto-parietal
130	47	-50	29	Default-mode	175	48	25	27	Fronto-parietal
131	-49	-42	1	Default-mode	176	-47	11	23	Fronto-parietal
132	-31	19	-19	Unspecified	177	-53	-49	43	Fronto-parietal
133	-2	-35	31	Unspecified	178	-23	11	64	Fronto-parietal
134	-7	-71	42	Unspecified	179	58	-53	-14	Fronto-parietal
135	11	-66	42	Unspecified	180	24	45	-15	Fronto-parietal