Network Community Structure Detection for Directional Neural Networks Inferred From Multichannel Multisubject EEG Data

Ying Liu*, Jason Moser, and Selin Aviyente, *Member, IEEE*

***Abstract*—In many neuroscience applications, one is interested in identifying the functional brain modules from multichannel, multiple subject neuroimaging data. However, most of the existing network community structure detection algorithms are limited to single undirected networks and cannot reveal the common community structure for a collection of directed networks. In this paper, we propose a community detection algorithm for weighted asymmetric (directed) networks representing the effective connectivity in the brain. Moreover, the issue of finding a common community structure across subjects is addressed by maximizing the total modularity of the group. Finally, the proposed community detection algorithm is applied to multichannel multisubject electroencephalogram data.**

***Index Terms*—Community detection, directed information, effective connectivity, electroencephalogram(EEG), group analysis.**

I. Introduction

With the advances in functional neuroimaging modalities such as electroencephalography (EEG) and functional magnetic resonance imaging (fMRI), it is now possible to measure human brain activity with increasing temporal and spatial resolution [1]–[3]. This vast amount of spatiotemporal data requires the development of computational methods, e.g., connectivity analysis, capable of quantifying the functional integration in the brain for a better understanding of the healthy and diseased brain [4]. Currently, three types of brain connectivity are considered: anatomic (description of the physical connections), functional (temporal correlation between distant neurophysiological events), and effective (causal influence that a neural system may exert over another). Anatomic connectivity has been assessed using both *in vitro* and *in vivo* methods such as diffusion tensor imaging [5] and molecular probes. Functional and effective connectivity, on the other hand, are based on the functional properties of the various cortical regions rather than the analysis of specific physical connections between them. Functional connectivity is defined as statistical dependences among remote neurophysiological events and is usually inferred on the basis of correlations between measurements of neuronal activity [6]–[8]. However, correlations can arise from stimulus-locked transients evoked by a common input or can reflect stimulus-induced oscillations mediated by synaptic connections [9]. Integration within a distributed system is usually better understood in terms of effective connectivity, which is defined as the influence that one neural system exerts over another, either at a synaptic or population level.

The relationships between multiple regions quantified by effective connectivity can be better characterized using complex network methods. In recent years, complex network tools have been applied to neuroscience and neuroimaging studies and have resulted in a better understanding of the brain at a system level [10]–[12]. Most of the work in the area of complex network analysis of the human brain has focused on quantifying brain networks with a small number of neurobiologically meaningful and easily computable graph theoretic measures [11]–[14]. These measures include the small-world parameter, efficiency [15], hierarchy [16], [17], centrality [18], and the distribution of network hubs [19]. Although these measures summarize the key aspects of complex networks at both the global (the whole network) and the local (each node together with its most immediate neighbors) levels, they do not provide any information about the intermediate scale of network organization which is more accurately described by the community structure or modularity of the network [10], [20]–[22]. The modules of a complex network are subsets of vertices such that connections within a module are denser than connections with the rest of the network. Module detection also allows one to obtain simplified reduced representations of complex networks in terms of subgraphs or communities [23].

In recent years, a lot of work has been done on applying community detection algorithms from graph theory to the study of functional brain networks [10], [21], [24]–[26]. Functional brain networks are usually described by undirected graphs with corresponding symmetric association matrices, where each entry indicates the pairwise functional connectivity between two regions. Therefore, most of the work on community detection for the study of brain networks has focused on undirected

Manuscript received September 15, 2012; revised January 10, 2013 and November 13, 2013; accepted December 21, 2013. Date of publication January 2, 2014; date of current version June 14, 2014. This work was in part supported by the National Science Foundation under Grant CCF-0728984, Grant CAREER CCF-0746971. This paper was presented in part at the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) [1]. *Asterisk indicates corresponding author.*

*Y. Liu is with the Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48824 USA (e-mail: liuying5@egr.msu.edu).

J. Moser is with the Department of Psychology, Michigan State University, East Lansing, MI 48824 USA (e-mail: jmoser@msu.edu).

S. Aviyente is with the Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48824 USA (e-mail: aviyente@egr.msu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2013.2296778

networks [24]. However, as Friston points out the functional integration of the brain can be better understood through effective connectivity since it reflects the dynamic (activity dependent and time dependent) characteristics of a system [6]. In this sense, the brain network can be better described by an effective connectivity network, where the edges of the graph have direction and the corresponding association matrix is no longer symmetric. As Leicht *et al.* claims, approaches which ignore the direction of links may fail to understand the dynamics of the system, and similarly any community detection method developed for undirected networks may fail to reveal the actual community structure [27]. Therefore, we expect that using effective connectivity would reveal new topological characteristics of the brain network that were not available by functional connectivity networks.

In this paper, we propose a multisubject hierarchical community detection algorithm for describing the modular structure of effective brain networks involved in cognitive control. The approach outlined in this paper advances the current study of brain networks in several key ways. First, the directed information (DI) measure is employed to quantify the effective connectivity in the brain. DI is a model free and directional measure that can quantify both the linear and nonlinear relationships between remote neurophysiological events [28]–[33]. Second, we employ recent work in the area of community detection in directed networks to infer functional modules. In this paper, we extend the state-of-the-art technique for network modularity optimization, Louvain method, to weighted directed networks to find the functional modules [34]. Third, we extend community detection algorithms developed for a single network to a group of networks in order to find a modular structure that best describes all of the subjects in a group. Finally, the algorithm proposed in this paper is applied to multichannel EEG recordings which provide a higher temporal resolution representation of the community structures in the brain compared to fMRI.

II. Background

A. Overview of Community Detection Methods

Many networks of interest in the sciences are found to divide naturally into communities or modules. The problem of detecting and characterizing this community structure is a key step for understanding complex networks [35]–[37]. The idea of community detection is closely related to data clustering, graph partitioning, and hierarchical clustering. Therefore, traditional approaches in these areas can be employed for community detection [38]. Two key approaches that have been widely investigated in community detection are: 1) spectral clustering-based techniques and 2) network modularity optimization strategies. Spectral clustering-based approaches rely on the optimization of the process of cutting the graph representing the given network. Since this problem is NP-hard, different approximate techniques such as the normalized cuts algorithm and ratio cuts algorithm have been proposed [39]. The main problem with spectral clustering-based techniques is that one has to know in advance the number and the size of communities in the network. Network modularity-based methods, on the other hand, rely on the modularity function Q to determine the optimal number of clusters in the network. A good partitioning of a network is expected to have high modularity Q with Q=(fraction of edges within communities)-(expected fraction of such edges) [20], where the expected fraction of edges is evaluated for a random graph. For a directed weighted network represented by a graph $G = (V, E)$ with N nodes and an association matrix $\mathbf{A}$, the modularity function is given as [40]:

$$Q = \frac{1}{W}\sum_{i,j=1}^{N}\left[A_{ij} - \frac{s_i^{\text{out}} s_j^{\text{in}}}{W}\right]\delta_{c_i,c_j} \tag{1}$$

where $A_{i,j}$ is the weight of edge $e_{i\to j}, s_i^{\text{in}} = \sum_j A_{ji}$ ($s_i^{\text{out}} = \sum_j A_{ij}$) is the inflow (outflow) of node $i, W = \sum_{i,j} A_{i,j}, c_i(c_j)$ is the community that node $i(j)$ belongs to; and δ_{c_i,c_j} is equal to 1 when i and j are in the same community and is equal to 0 otherwise. The problem of maximizing the network modularity has been proven to be NP complete. For this reason, several heuristic strategies to maximize the network modularity such as Girvan–Newman algorithm [20], the fast clustering algorithm [41], the extremal optimization method [37] and the Newman–Leicht mixture model-based approach [42] have been proposed.

Although most of the modularity-based community detection algorithms have focused on binary and undirected networks, in recent years there have been some extensions to weighted and directed networks [38], [43]. However, these approaches are limited to networks with a small number of clusters. Recently, Blondel *et al.* introduced an alternative greedy algorithm, which is known as the Louvain method, to find the hierarchical structure of undirected weighted graphs [34]. Compared to other methods, this method performs better in terms of the computation time especially for networks with a large number of nodes [34]. In this paper, we extend this algorithm to directed weighted graphs for community detection.

B. Group Analysis Approaches

In many neuroimaging studies, extracting a common set of features or a representation, e.g., the common community structure for a group of subjects, is important for understanding group behavior [44]. This common structure usually gives us an overall understanding of the group, while individual subject level representations show the subject-specific features. There are three major group analysis strategies that can be employed for community detection: the "virtual-typical-subject" (VTS) approach, the "individual structure" (IS) approach, and the algorithm-based approach [45], [46]. The VTS approach assumes that data from each subject follow the same distribution. It constructs a virtual subject by pooling or averaging the group data and obtains one community structure for the whole group. However, this approach does not consider the intersubject variability and may fail when the behavior from subject to subject is not consistent [45], [46]. The IS approach applies a community detection algorithm to each individual subject and extracts a common community structure from these individual structures. This subject-specific strategy considers diversity across subjects

and integrates the individual structures by averaging, applying voting/consensus algorithms, or finding the most representative subject [44], [47], [48]. The IS approach is usually computationally expensive since it requires the extraction of the community structure for each subject before obtaining a common structure. Both the VTS and IS approaches focus on either preprocessing the data or postprocessing the community structures obtained from each subject. However, neither of these approaches offers a direct extension of the community detection algorithm from the single subject to the multiple subject case [49]. In this paper, we propose a group analysis method by optimizing a common modularity function for directed networks from multiple subjects.

C. Directed Information

Model-based methods and information-theoretic measures are two main approaches to quantifying the effective connectivity between two time series. Model-based methods such as Granger-casuality-based methods, dynamic causal modeling (DCM), and linear non-Gaussian acyclic model, usually are limited to capturing linear relations or requires a *priori* knowledge about the underlying signal models [50]–[53]. However, EEG data are known to have nonlinear dependences, thus information-theoretic measures which are model-free and capable of detecting both linear and nonlinear relationships are suitable for capturing these dependences. Information-theoretic measures [30], [54], [55], such as transfer entropy [54] and directed transinformation, have found numerous applications in neuroscience [31]. However, the implementation of these measures requires the assumption of certain stochastic models, e.g., transfer entropy requires order selection based on a Markov assumption for the system [56]. Recently, directed information has attracted attention for quantifying directional dependences [30], [57]. DI theory was originally developed for the study of communication channels with feedback [30]. In recent years, there has been a growing interest in applying this measure to applications in signal processing, neuroscience and bioinformatics, such as inferring effective connectivity in the brain based on multivariate time series [33], [56]. New theoretical developments have also motivated the use of this measure in quantifying causality between multivariate time series data [29]. In particular, Amblard *et al.* [56] showed how DI and Granger causality are equivalent for linear Gaussian processes and proved key relationships between transfer entropy and the directed information.

The definition of DI for two length n sequences $\mathbf{X} = X^n = (X_1, \ldots, X_n)$ and $\mathbf{Y} = Y^n = (Y_1, \ldots, Y_n)$ is as follows:

$$\mathrm{DI}(X^n \to Y^n) = \sum_{k=1}^{n} I(X^k; Y_k | Y^{k-1}) \tag{2}$$

where $I(X;Y|Z)$ is the conditional mutual information between two random variables X and Y with respect to Z. The computation of DI requires the estimation of joint probabilities of high-dimensional random variables over time. If X and Y are normally distributed, the joint entropy can be estimated based on the covariance matrices. However, for EEG data, the distribution is usually not Gaussian. Therefore, nonparametric entropy and mutual information estimators have to be employed [58], [59]. In this paper, DI estimation based on mutual information is used to estimate the DI directly from EEG data by using the adaptive partitioning method discussed in [59]. However, as the length of the random process increases, the computational complexity, the bias, and variance of these estimators increase immensely with fixed number of realizations for each random process. Recently, we have introduced time-lagged DI for every two samples of X^n and Y^n with different time lags, and have shown that this measure can capture the causality relationship while reducing the computational complexity [60]. DI for every two samples of X^n and Y^n at the kth time sample with a time delay of d ($k > d$) is defined as

$$\begin{aligned}\mathrm{DI}_k(X_{k\text{-}d}X_{k-d+1} \to Y_kY_{k+1}) \\ = I(X_{k\text{-}d}; Y_k) + I(X_{k\text{-}d}X_{k-d+1};\ Y_{k+1}|Y_k)\end{aligned} \tag{3}$$

where $X_{k\text{-}d}$ refers to the $(k-d)$th time sample of the random process $X, k = d+1, \ldots, n-1, d = 0, \ldots, p, p < k, p$ is the maximum lag considered, and n is the length of the signal. For a single-order model with d equal to the actual model order, the sum of time-lagged DI over all time points is equivalent to the actual DI [29]. In practice, for a model with multiple orders, at each time point k, the time-lagged DIs over a certain time window are averaged, i.e., $\overline{\mathrm{DI}_k} = \sum_{d=d_1}^{d_2} \mathrm{DI}_k(X_{k\text{-}d}X_{k-d+1} \to Y_kY_{k+1})/(d_2 - d_1 + 1)$. If X and Y are normally distributed, the complexity of using the original definition of DI is $O(n^4)$ (using LU decomposition [61]), while the complexity of computing DI for two time samples is $O(n)$.

Since $0 \le \mathrm{DI}(X^n \to Y^n) \le I(X^n; Y^n) < \infty$, a normalized version of DI,

$$\rho_{\mathrm{DI}}(X^n \to Y^n) = \frac{\mathrm{DI}(X^n \to Y^n)}{I(X^n; Y^n)} \tag{4}$$

is used for comparing different interactions. For a unidirectional system $\mathbf{X} \to \mathbf{Y}$ with no instantaneous interaction between $\mathbf{X}$ and $\mathbf{Y}$, $\rho_{\mathrm{DI}}(X^n \to Y^n) = 1$ and $\rho_{\mathrm{DI}}(Y^n \to X^n) = 0$.

III. Algorithm for Community Detection

In this section, we first extend the method proposed by Blondel *et al.* to weighted directed networks [34] and then propose an extension for group analysis. The algorithm proposed by Blondel *et al.* is for undirected weighted networks and maximizes the modularity through greedy search. The computation time of this method is $O(N \log N)$, where N is the number of nodes in the network. Compared to other modularity maximization-based clustering methods, this method performs better in terms of computation time and maximizing the modularity for networks with various sizes [34], [38], i.e., higher efficiency and accuracy when partitioning a network into communities.

A. Algorithm for Community Detection in Weighted Directed Networks

In this section, in order to determine the hierarchical community structure of the effective connectivity networks, we propose

to extend Blondel's approach to directed weighted networks. Initially, all nodes of the network are assigned to different communities. The algorithm for determining the community structure of a directed weighted network consists of two steps. First, for each node i, the gain in the modularity ΔQ_j is computed when node i is assigned to the communities of all other nodes j (C_j), where $j = 1, \ldots, N, j \neq i$. The original algorithm only evaluates the change of modularity when node i is assigned to the communities of its neighbors j, where j is defined as the neighbor of i when $A_{i,j} \neq 0$, which may be inaccurate and yield spurious partitions in practical cases [38]. For this reason, we consider the change of modularity with respect to all other nodes in the network. Once ΔQ_j is obtained, where $j = 1, \ldots, N, j \neq i$, the community for which ΔQ_j is positive and largest is chosen as the new community for node i. ΔQ_j, which partly determines the efficiency of the algorithm, can be computed as follows:

$$\Delta Q_j = \frac{1}{W}\sum_{p=1}^{N_j}\left(A_{i,j_p} - \frac{s_i^{\text{out}} s_{j_p}^{\text{in}}}{W} + A_{j_p,i} - \frac{s_i^{\text{in}} s_{j_p}^{\text{out}}}{W}\right) - \frac{1}{W}\sum_{p=1}^{N_i-1}\left(A_{i,i_p} - \frac{s_i^{\text{out}} s_{i_p}^{\text{in}}}{W} + A_{i_p,i} - \frac{s_i^{\text{in}} s_{i_p}^{\text{out}}}{W}\right) \quad (5)$$

where $j_p \in C_j, i_p \in C_i$, and $i_p \neq i, N_j$ (N_i) is the number of nodes in community C_j (C_i), s_i^{in} (s_i^{out}) is the inflow (outflow) of node i, and $W = \sum_{i,j} A_{i,j}$. The first term on the right-hand side of (5) is the modularity contributed by node i when it is assigned to the community of node j, C_j, while the second term is the modularity contributed by node i when it stays in its original community C_i. This process is sequentially and repeatedly applied to all nodes until there is no gain in modularity. At this stage, the first-level partition of the network is obtained. Next, nodes in the same community after the first level partition are used to form metanodes. The number of metanodes is equal to the number of current communities, and the weights between two metanodes are given by the sum of the weights of edges between nodes in the corresponding communities [34]

$$A_{\text{new}}(k,l) = \sum_{i \in C_k}\sum_{j \in C_l} A_{i,j} \quad (6)$$

where $k, l = 1, \ldots, tN$ with tN being the current number of metanodes, and C_k, C_l are the kth and lth communities, respectively. Note that since the network is directed, $A_{\text{new}}(k,l) \neq A_{\text{new}}(l,k)$. These two steps, assigning community labels and forming metanodes, are iterated until the modularity cannot increase anymore, and several levels of partitions are obtained at different resolutions. A low-resolution partition refers to the case where the number of nodes (metanodes) is small. The first-level partition (before the formation of metanodes) has the highest resolution. The modularity is always computed with respect to the initial graph topology such that the two-step iterative procedure is not trapped at a local maximum. This algorithm is summarized in Algorithm 1.

Algorithm 1 Community detection for weighted networks

Input: Weighted adjacency matrix $A \in (0,1)^{N\times N}$, nodes $v_1, \cdots, v_N$, initial community structure $C = \{\{1\}, \cdots, \{N\}\}$, $tN = N$.
Output: M Communities.

```
Compute the modularity Q of the network;
repeat
  ΔQ_total = 0, nchange = 0;
  repeat
    for h = 1 to tN do
      for j = 1 to tN do
        Compute the change of modularity ΔQ_j when node v_h is assigned to C_j;
      end for
      j* = arg max_j ΔQ_j;
      if ΔQ_{j*} > 0 then
        C_{j*} = C_{j*} ∪ v_h;
      end if
    end for
    Compute the change of the modularity ΔQ_total = Q_new − Q;
    if ΔQ_total > 0 then
      Q = Q_new, nchange = nchange + 1;
    end if
  until ΔQ_total ≤ 0
  Nodes in the same community form new meta-nodes;
  tN is equal to the current number of communities;
  Recompute the weighted matrix A ∈ (0,1)^{tN×tN};
until nchange = 0.
```

B. Algorithm for Community Detection for Multiple Subjects

In neuroscience, one of the challenging problems is group analysis when information from multiple subjects needs to be merged. The standard approach to group analysis is to either average the data without considering the intersubject variability or average the detected community structures from all subjects. However, neither of these approaches integrates the information from each subject at the algorithm level, which can take both the intersubject variability and commonality into account. In the algorithm presented above, we only take into account the change of modularity for moving one node to the community of another node for each subject. However, this operation leads to the change of modularity for all subjects. Therefore, in the proposed approach for multiple subjects, we compute the change of modularity ΔQ_j^k for subject k when node i is assigned to the communities of all other nodes j, where $j = 1, \ldots, N, j \neq i$, and $k = 1, \ldots, L$ with L being the number of subjects. To improve the robustness of the clusters across subjects, for a fixed j, we exclude any change in the modularity that is outside the 25th and 75th percentiles of ΔQ_j^k, where $k = 1, \ldots, L$.[1] We then compute the average of the gain in modularity for the remaining L' subjects, $\Delta Q_j = \sum_{k=1}^{L'} \Delta Q_j^k / L'$, where L' is the number of subjects that are between the 25th and 75th percentiles of $\Delta Q_j^k, k = 1, \ldots, L$. We then assign node i to the

[1]These percentiles were selected empirically based on real and simulated data and can be easily modified for different applications.

Algorithm 2 Community detection for multiple weighted networks

Input: Weighted adjacency matrix $A^k \in (0,1)^{N\times N}$, $k = 1, \cdots, L$ with L being the number of subjects, nodes $v_1, \cdots, v_N$, initial community structure $C = \{\{1\}, \cdots, \{N\}\}$, $tN = N$.
Output: M Communities.

```
Compute the modularity Q^k of subject k, k = 1, ..., L;
The modularity of the group is Q = sum_{k=1}^{L} Q^k;
repeat
  dQ_total = 0, nchange = 0;
  repeat
    for h = 1 to tN do
      for j = 1 to tN do
        Compute the change of modularity dQ_j^k when node v_h is assigned to C_j for subject k, k = 1, ..., L;
        Find the outliers beyond the 25th and 75th percentiles of dQ_j^k with k = 1, ..., L ;
        The average change of modularity for the whole group is dQ_j = sum_{k=1}^{L'} dQ_j^k / L', where L' is the number of subjects that are between the 25th and 75th percentiles of dQ_j^k with k = 1, ..., L;
      end for
      j* = arg max_j dQ_j;
      if dQ_{j*} > 0 then
        C_{j*} = C_{j*} U v_h;
      end if
    end for
    Compute the modularity of the whole group Q_new = sum_{k=1}^{L} Q_new^k, Q_new^k is the modularity of subject k;
    Compute the change of the modularity dQ_total = Q_new - Q;
    if dQ_total > 0 then
      Q = Q_new, nchange = nchange + 1;
    end if
  until dQ_total <= 0
  Nodes in the same community form new meta-nodes;
  tN is equal to the current number of communities;
  Recompute the weighted matrix A^k in (0,1)^{tN x tN} of subject k, k = 1, ..., L;
until nchange = 0.
```

community of node j which maximizes ΔQ_j. In this way, the effect of outliers is directly decreased at the algorithm level. The details are shown in Algorithm 2

C. Validation

Once the community structure is determined by the proposed algorithm, we test the agreement of the observed communities with the actual community structure using Cohen's Kappa score [62]. Cohen's Kappa ($\hat{\kappa}$) adjusts the observed proportional agreement by taking into account the amount of agreement which would be expected by chance:

$$\hat{\kappa} = \frac{p - p_e}{1 - p_e} \tag{7}$$

where p is the proportion of units where there is agreement and p_e is the proportion of units expected to agree by chance, determined by Chi-square tests. More specifically, in evaluating our algorithms, suppose that there is a network with N nodes;

TABLE I
OBSERVATION TABLE FOR EVALUATING COMMUNITY DETECTION ALGORITHMS

	$O_{i,j} = 1$	$O_{i,j} = 0$
$M_{i,j} = 1$	TP	FP
$M_{i,j} = 0$	FN	TN

where O is the actual cluster assignment matrix with $O_{i,j} = 1$ if nodes i and j are in the same cluster and 0 otherwise; M is the measured cluster assignment matrix with $M_{i,j} = 1$ if nodes i and j are identified in the same cluster by the algorithm and 0 otherwise. Then, in Table I, TP refers to the number of node pairs which are correctly identified as being in the same cluster by the algorithm; FP refers to the number of node pairs which are falsely identified as being in the same cluster by the algorithm; FN refers to the number of node pairs which are falsely identified as not being in the same cluster by the algorithm; and TN refers to the number of node pairs which are correctly identified as not being in the same cluster by the algorithm. Based on these parameters $p = \frac{\mathrm{TP+TN}}{\mathrm{TP+FP+FN+TN}}$, $p_e = \frac{(\mathrm{TP+FP})(\mathrm{TP+FN})}{(\mathrm{TP+FP+FN+TN})^2} + \frac{(\mathrm{FN+TN})(\mathrm{FP+TN})}{(\mathrm{TP+FP+FN+TN})^2}$.

Compared to other conventional and robust statistics like receiver operating characteristic (ROC) analysis, one additional advantage of using Cohen's Kappa is that the standard error for this statistic is known and the confidence level can be obtained [63]. The standard error of $\hat{\kappa}$ is given by

$$\mathrm{SE}(\hat{\kappa}) = \sqrt{\frac{p(1-p)}{n(1-p_e)^2}} \tag{8}$$

where $n = \frac{(N-1)N}{2}$ is the total number of node pairs. The 95% confidence interval for $\hat{\kappa}$ is $\hat{\kappa} - 1.96 * \mathrm{SE}(\hat{\kappa})$ to $\hat{\kappa} + 1.96 * \mathrm{SE}(\hat{\kappa})$.

IV. MATERIALS

In this section, we briefly explain the datasets used in this paper. We first present the details of EEG data and the computation of effective connectivity for this data. We then describe the two simulated data sets generated for the evaluation of the algorithms presented in this paper. The first dataset is generated to illustrate the importance of edge direction information for revealing the real structure of a directed network. The second dataset is generated for verifying the effectiveness of the proposed group analysis method.

A. EEG Data

Participants: EEG data from ten undergraduates at Michigan State University were drawn from an ongoing study of relationships between the error-related negativity (ERN) and individual differences.[2] All participants retained for analysis made at least six errors since the ERN becomes stable and reliable at six errors [65]. No participants discontinued their involvement once the experiment had begun.

[2]Participants for the present analysis were drawn from samples reported in [64].

Task: Participants completed a letters version of the Eriksen Flanker task [66]. Stimuli were presented on a Pentium R Dual Core computer, using Presentation software (Neurobehavioral systems, Inc.) to control the presentation and timing of stimuli, the determination of response accuracy, and the measurement of reaction times. During the task, participants were presented with a string of five letters. Each five-letter string was either congruent (e.g., FFFFF) or incongruent (e.g., EEFEE) and participants were required to respond to the center letter (target) via the left or right mouse button. Trial types were varied randomly such that 50% of the trials were congruent. Letters were displayed in a standard white font on a black background and subtended 1.3° of visual angle vertically and 9.2° horizontally. A standard fixation mark $(+)$ was presented during the intertrial interval (ITI). Each trial began with the presentation of the flanking letters (i.e., EE EE). Flanking letters remained on the screen for 35 ms and were followed by the target (i.e., EEFEE), which remained for 100 ms (135 ms total presentation time). Each trial was followed by a variable ITI $(1200 - 1700$ ms). The entire experimental session consisted of 480 trials grouped into six blocks of 80 trials each. The letters constituting each string were varied between blocks (e.g., M and N in block 1 and E and F in block 2) and response-mappings were reversed at the midpoint of each block (e.g., left mouse-button click for M through 40 trials of block 1, then right-mouse button click for M for the last 40 trials of block 1) in order to elicit a sufficient number of errors for ERN calculation.

Psychophysiological data recording, reduction, and analysis: Continuous electroencephalographic (EEG) activity was recorded by 64 Ag–AgCl electrodes placed in accordance with the $10/20$ system. Electrodes were fitted in a BioSemi (BioSemi, Amsterdam, The Netherlands) stretch-lycra cap. In addition, two electrodes were placed on the left and right mastoids. The electro-oculogram generated by eye-movements and blinks were recorded by FP1, as well as by electrodes placed below the right eye and on the left and right outer canthi, all approximately 1 cm from the pupil. During data acquisition, the common mode sense active electrode and driven right leg passive electrode formed the ground, as per BioSemi's design specifications [67]. These two electrodes serve a grounding function in place of a typical ground electrode. They provide a feedback loop to keep the subject's average voltage as close to zero as possible. All bioelectric signals were digitized at 512 Hz using ActiView software (BioSemi). Offline analysis was performed using BrainVision Analyzer 2 (BrainProducts, Gilching, Germany). Scalp electrode recordings were rereferenced to the numeric mean of the mastoids and band-pass filtered with cutoffs of 0.1 and 30 Hz (12 dB/oct rolloff). Ocular artifacts were then corrected using the regression method developed by Gratton *et al.* [68]. Response-locked data were segmented into individual epochs beginning 200 ms prior to the response and continued for 1000 ms. Individual trials were rejected on the basis of excessive physiological activity: a voltage step exceeding 50μV between contiguous sampling points, a voltage difference of more than 200μV within a trial, or a maximum voltage difference less than 0.5 mV within a trial. The average correct response negativity (CRN) and ERN across all trials and subjects at FCz electrode are shown in Fig. 1.

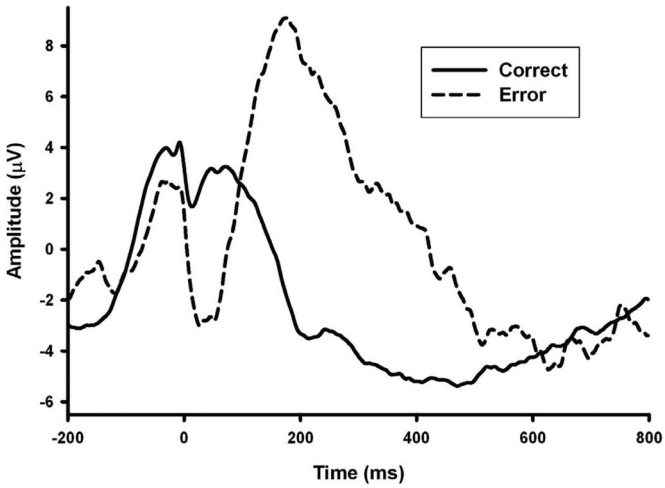

Fig. 1. Average CRN and ERN responses across all trials and subjects at FCz electrode.

Quantification of connectivity: Once the data are collected as described previously, we are interested in studying the effective connectivity in the ERN time window and frequency band. Previous work indicates there is increased information flow associated with ERN for the theta frequency band $(4 - 8$ Hz) and ERN time window $25 - 75$ ms for error responses compared to correct responses in particular between medial prefrontal cortex (mPFC) and lateral prefrontal cortex (lPFC) regions [69]. In the current dataset, the number of error responses for the selected 10 subjects ranged from 28 to 88 with a mean of 49. A random subset of correct trials was selected, to equate the number of error responses for each participant. The EEG data are preprocessed by the spherical spline current source density (CSD) waveforms to sharpen event-related potential (ERP) scalp topographies and reduce volume conduction [70]. The CSD has fewer assumptions than many inverse transforms, attenuates volume conduction, and represents independent sources near the cortical surface [71]. In addition, a zero-phase fourth-order Butterworth bandpass filter is used to obtain signals in the theta band. The effective connectivity is quantified using the time-lagged DI computed over a window corresponding to the ERN response $(0 - 100$ ms after the response), for all trials between each pair of 61 electrodes in the theta band. The time-lagged DI is averaged over a window within $10 - 20$ ms time delay [60], i.e., $d = 5, \ldots, 10$ in (3). Once the connectivity matrices for each response type and each subject are obtained, we used the proposed group analysis algorithm to identify the community structure for each group.

B. Simulated Data

1) Directed Versus Undirected Networks: Most of the existing community detection algorithms are intended for the analysis of undirected and binary networks. However, many networks of interest, such as biological networks, are directed. One approach that has been commonly used for community detection in directed networks is to directly apply the algorithms designed

for undirected networks without considering the edge direction information [27]. To illustrate the importance of edge direction information for community detection, we generated a simulated directed network and compared the performance of two approaches for detecting the community structure of this network. The first approach is to apply the proposed algorithm to the association matrix of the directed network $\mathbf{A}$ directly and the second approach is to ignore the edge direction information and apply the original Blondel's algorithm to the association matrix of the undirected network $\frac{1}{2}(\mathbf{A}+\mathbf{A}^T)$, where $\mathbf{A}^T$ is the transpose of $\mathbf{A}$. The simulated network consisted of 24 nodes and 2 clusters, each cluster with 12 nodes. Nodes 1 to 12 were in the same cluster, while the rest were in the other cluster. The edges within each cluster were generated from a uniform distribution in the interval $[0.3, 0.7]$, while the intercluster edges from cluster 1 to 2 were uniformly distributed in the interval $[0, 0.3]$ and the inter-cluster edges from cluster 2 to 1 were uniformly distributed in the interval $[0.7, 1.0]$. In this case, all of the entries of $\mathbf{A}$ were in the $[0, 1]$ interval, and $\mathbf{A}$ was not symmetric. Without loss of generality, we generated this network 100 times and the average Cohen's Kappa score for both approaches were obtained.

2) Simulated Networks for Group Analysis: To compare the effectiveness of the proposed group analysis method to existing methods, we generated a group of ten directed networks with the same community structure. Each network consisted of 64 nodes and 4 clusters, each cluster with 16 nodes. In order to reflect the statistical properties of the connectivity networks obtained from the EEG data, we used the empirical directed information data from the 10 subjects to select the inter- and intracluster edge values of the simulated networks. Moreover, in order to get well-defined clusters, the intracluster edges should have low variance whereas the intercluster edges should have high variance. For each simulated network, we sorted all of the DI values from the corresponding subject and split them into ten deciles. All edges within one cluster, i.e., intracluster connections, were uniformly sampled with replacement from one of the four highest deciles. The edges between clusters, on the other hand, were sampled with replacement from the whole population of DI values where the sampling weights were generated from a normal distribution. The mean of the normal distribution is equal to the mean of DI values in the lowest decile, which increases the probability of choosing lower edge values. This procedure ensures that the intercluster edges simultaneously have a low mean and a high variance. Without loss of generality, we generated 100 simulations of the networks to compute the average Cohen's Kappa score.

We also compared the performance of the proposed algorithm with VTS and IS for a case with outliers. For this purpose, we generated a group of ten directed networks with eight of them having the same community structure as described above. The other two networks each consisted of a single cluster, where all of the edge values in each network were uniformly sampled with replacement across the 4th and 5th deciles of the corresponding subject's DI values. Without loss of generality, we generated 300 simulations of the networks to compute the average Cohen's Kappa score.

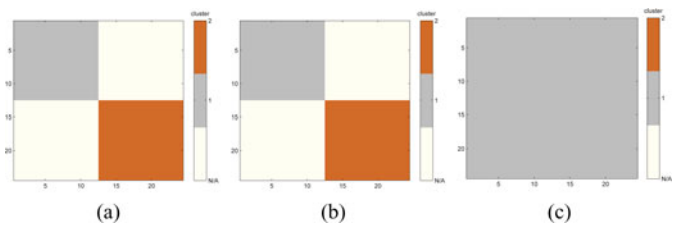

Fig. 2. Community membership matrices for (a) actual community structure, (b) community structure obtained from the proposed algorithm which considers edge direction information, and (c) community structure obtained from Blondel's original algorithm which does not consider edge direction. Ivory indicates that the corresponding node pairs are not in the same cluster (cluster N/A). Gray indicates that the corresponding node pairs are in cluster 1. Brown indicates that the corresponding node pairs are in cluster 2.

V. Results

A. Directed Versus Undirected Networks

To verify the importance of edge direction information, the proposed algorithm is applied to the directed network described in Section IV-B1. The community detection results are shown in Fig. 2. We observe that the proposed algorithm can detect the community structure of the network if we use the edge direction information [see Fig. 2(b)], whereas the original algorithm (Blondel's algorithm) designed for undirected networks fails to capture the actual community structure [see Fig. 2(c)]. Cohen's Kappa score is used to evaluate the performance of these two approaches. If the edge direction is taken into account, the mean of $\hat{\kappa}$ is 1 and 0.0011 otherwise. The average standard error $\mathrm{SE}(\hat{\kappa})$ is 0 for the proposed algorithm and 0.0417 otherwise. Therefore, our algorithm can reveal the real structure of the directed network better than conventional clustering algorithms designed for undirected networks. In fact, when the association matrix of a network is strongly asymmetric, the community detection results between directed and undirected representations of the network will be quite different. In these situations, the community detection algorithms based on the directed weighted graphs can reveal the real community structure of the network, whereas conventional community detection algorithms, which detect communities without considering the edge direction, blur the underlying community structure.

B. Group Analysis on Synthetic Data

Three different group analysis strategies, Algorithm 2 (with and without excluding the outliers), VTS, and IS, are applied on the simulated data described in Section IV-B2 with the results presented in Tables II and III. The proposed algorithm is applied in two different ways: excluding the changes in modularity outside the 25%th and 75%th percentiles (as described in Section III-B) and including all of the networks in the computation of change in modularity.

Table II presents the performance of the different algorithms when applied to the group of ten networks with the same structure. We observe that the proposed algorithm and the VTS both outperform the IS approach with a higher mean Cohen's Kappa score and a lower standard error, which indicates that the proposed method can provide more robust solutions to the group analysis problem. Moreover, the proposed algorithm without

TABLE II
AVERAGE COHEN'S KAPPA FOR GROUP ANALYSIS (NO OUTLIERS IN THE DATA)

$\hat{\kappa}$	Algorithm 2 (modularity change in the range of 25%th to 75%th percentiles)	Algorithm 2 (total change in modularity)	VTS	IS (voting)
mean	0.9943	1	0.9943	0.9853
$SE(\hat{\kappa})$ $\times(10^{-3})$	0.2362	0	0.2362	0.6630

TABLE III
AVERAGE COHEN'S KAPPA FOR GROUP ANALYSIS METHODS WITH TWO OUTLIERS IN THE GROUP

$\hat{\kappa}$	Algorithm 2 (modularity change in the range of 25%th to 75%th percentiles)	Algorithm 2 (total change in modularity)	VTS	IS (voting)
mean	0.9609	0.9419	0.8743	0.9356
$SE(\hat{\kappa})$	0.0018	0.0024	0.0052	0.0027

excluding any of the networks performs the best, since in this case the community structure is stable across the group and including more samples in the computation of change in modularity improves the accuracy of the community detection algorithm. In addition, a two-sided Wilcoxon rank sum test is used to test the null hypothesis that the distributions of the Cohen's Kappa values across 100 simulations using Algorithm 2 and that using VTS or IS are from the same continuous distributions with equal medians. The p-values for these two comparisons are 1 and 9.2924×10^{-45}, respectively, which indicates that the null hypothesis cannot be rejected for the comparison between Algorithm 2 and VTS at 5% significance level. Conversely, the null hypothesis that the distributions of Cohen's Kappa values from Algorithm 2 and IS are the same is rejected at 5% significance level.

For the case with two outliers in a group of ten networks, the results are presented in Table III. We observe that the proposed algorithm, with and without excluding the outliers, has the highest mean Cohen's Kappa score and the lowest standard error. In addition, Algorithm 2 with the outliers outside of the 25%th and 75%th percentiles excluded performs the best, since the removal of the outliers impacts the change of the modularity function for the whole group and reveals a common community structure that is more consistent across networks. These results indicate that Algorithm 2 reduces the effect of outliers without making any assumptions about the data. Since all of our data are generated from empirical DI values, this finding also suggests that our method is the most suitable for group analysis of EEG data. The increased performance of our method can be attributed to two key factors. First, in VTS, averaging the data may reduce the variance of modularity across subjects and blur the community structure. On the other hand, IS may easily be affected by outliers or networks with nondistinct community structure. Second, our method maximizes the modularity at the algorithmic level at each step and thus reduces the effect of outliers.

Similar to the previous analysis, a two-sided Wilcoxon rank sum test is used to test the null hypothesis that Cohen's Kappa values across 300 simulations using Algorithm 2 and that using VTS or IS are from the same continuous distributions with equal medians. The obtained p-values are 7.1711×10^{-15} and 1.7886×10^{-124}, respectively, which indicates that the null hypothesis that the distribution of the Cohen's Kappa values from Algorithm 2 and VTS(IS) are the same can be rejected at 5% significance level for both comparisons.

C. Group Analysis on EEG Data

Previous work indicates that there is increased information flow associated with ERN for the theta frequency band $(4-8$ Hz) and ERN time window $(25-75)$ ms for error responses compared to correct responses in particular between mPFC and lPFC regions [69]. We applied Algorithm 2 to the association matrices quantified by DI for the ten subjects for error and correct responses. The clustering results at all levels for each group are presented in Fig. 3. The number of partition levels for both CRN and ERN are 3. Since the modularity function for each response type achieves its maximum at the top (final) level, we interpret the partitions at the top (final) level for each group. The third level of partition for CRN has four large clusters, i.e., anterior-central, left-posterior, right posterior, and some nodes in the central region. The third level of partition for ERN has three large modules, i.e., left anterior-central region, right anterior-central region, and the posterior region. Compared to CRN, central nodes and anterior nodes are more closely connected for ERN, which indicates that the medial cortex and right lPFC regions work together when an error occurs. In addition, for ERN the left and right lateral nodes are assigned to separate clusters indicating functional specialization. We also notice that the clusters in the posterior region for the two response types are different. For ERN, the posterior nodes are more tightly connected resulting in a single cluster compared to CRN, which indicates close cooperation among nodes in this region.

In addition, we investigate the significance of the directionality of the interactions within clusters for the final level of partitioning for each response type. We use Y-bootstrapping, which is a stationary bootstrap method proposed in [72] and used for assessing the significance of causality in EEG studies [73]. In order to test the null hypothesis of nondirectional information flow, the causal structure between time series $\mathbf{X}$ and $\mathbf{Y}$ is destroyed. To be specific, for each electrode pair with time series $\mathbf{X}$ and $\mathbf{Y}$, if the significance of the directionality $\mathbf{X} \rightarrow \mathbf{Y}$ is tested, the time series $\mathbf{X}$ is resampled on a block basis 100 times to generate new observations $\mathbf{X}_m^*, m = 1, \ldots, 100$. The block-based resampling is performed as follows. First, blocks of length L_i are randomly drawn with replacement from the original time series. The length of each block is drawn from a geometric distribution. Data are wrapped in a circular way such that at the end of a time series a block can be drawn by adding observations from the beginning of the time series. The p-value of DI$(\mathbf{X} \rightarrow \mathbf{Y})$ is computed from the bootstrap

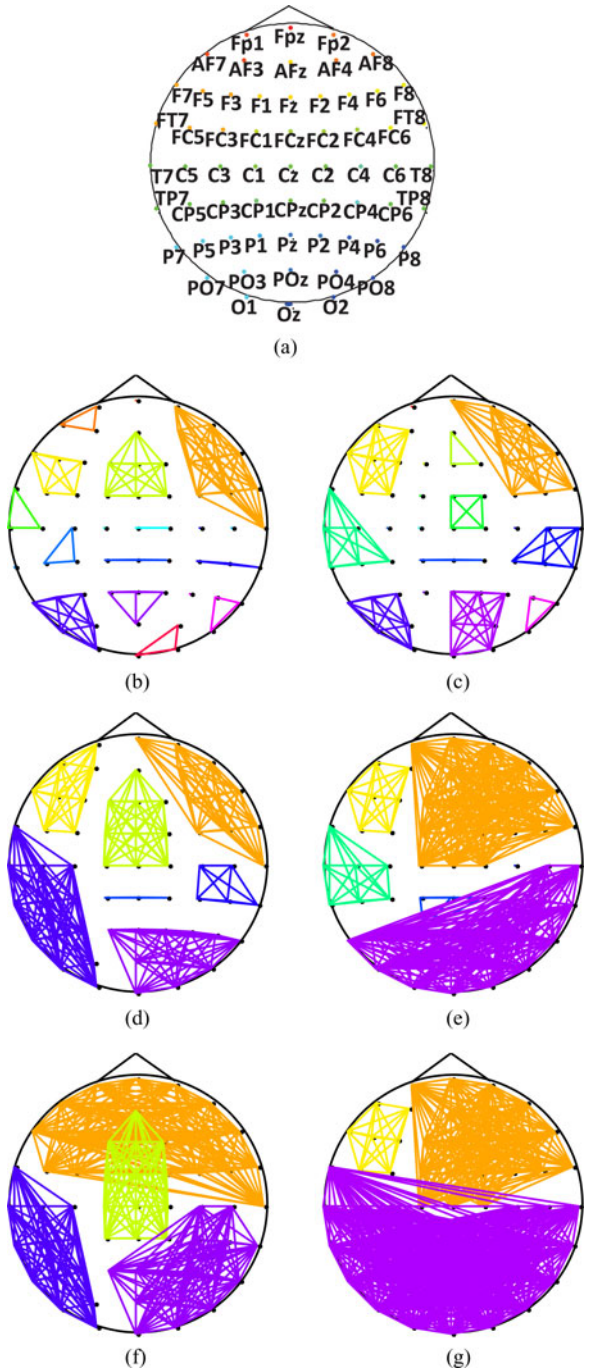

Fig. 3. Results of the hierarchical multisubject community detection algorithm for ten subjects in each group (CRN (b), (d), and (f) and ERN (c), (e), and (g)). (a) EEG 10–20 system. (b) CRN: level 1. (c) ERN: level 1. (d) CRN: level 2. (e) ERN: level 2. (f) CRN: level 3. (g) ERN: level 3.

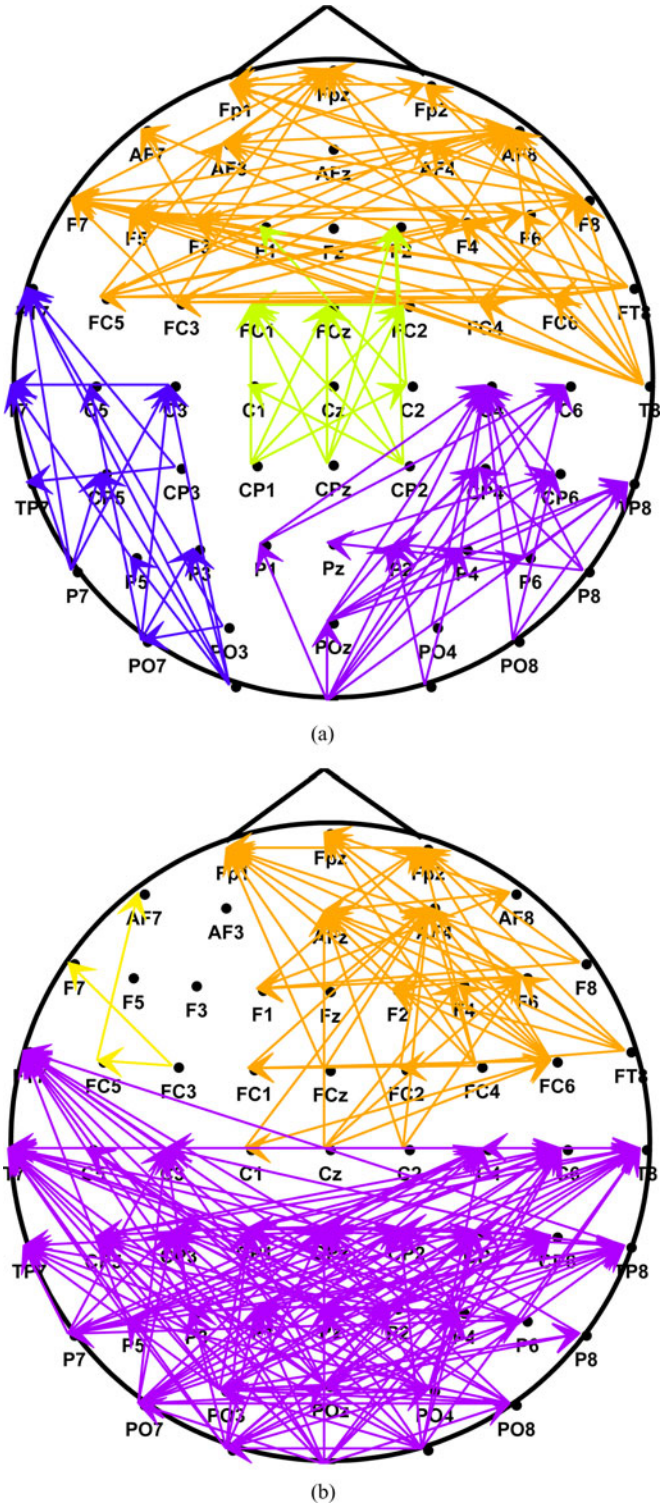

Fig. 4. Directionality of information flow within clusters. (a) Directional information flow for the last level of partitioning for CRN. (b) Directional information flow for the last level of partitioning for ERN.

distribution, $\mathrm{DI}(\mathbf{X}_m^* \rightarrow \mathbf{Y}), m = 1, \ldots, 100$. In order to control the error rates for multiple hypothesis testing for all pairs of electrodes within a cluster, the method proposed by Genovese *et al.* is used [74]. To implement this procedure, for electrode pairs in the same cluster, their p-values across subjects are pooled and sorted from smallest to largest, $p_1 \leq p_2 \cdots \leq p_V$, where V is the total number of electrode pairs in that cluster times the number of subjects. To control the false discovery rate bound q at 0.05, the largest i for which $p_i \leq \frac{i}{V} q$ is computed and is selected as the threshold, p_r, for this sequence of p-values. For each subject, if the p-value for $\mathrm{DI}(\mathbf{X} \rightarrow \mathbf{Y})$ is less than p_r, then the directed information flow from $\mathbf{X}$ to $\mathbf{Y}$ is significant; otherwise, it is not significant. Electrode pairs between which the information flow is significant in all ten subjects are shown in Fig. 4. We observe that the information flow is directional for

both CRN and ERN. For ERN, there is a clear directional flow from the central to the frontal electrodes in the anterior-central region.

Although we detect the community structure for both CRN and ERN at different resolutions, results in Fig. 3 do not show the amount and direction of information flow within and between clusters. We define the overall intercluster and intracluster information flow for all subjects, respectively, as follows:

$$\mathrm{IFL}_{C_k,C_k} = \frac{1}{N_k^2} \sum_{i,j \in C_k} \mathrm{TA}_{i,j}$$

$$\mathrm{IFL}_{C_k,C_p} = \frac{1}{2 \times N_k \times N_p} \sum_{i \in C_k} \sum_{j \in C_p} \mathrm{TA}_{i,j} \qquad (9)$$

where N_k (N_p) is the number of nodes in cluster C_k (C_p), $k, p = 1, \ldots, K$ with K being the total number of clusters, $TA = \frac{1}{L} \sum_{m=1}^{L} A^m$, where A^m is the association matrix of subject m in a group (ERN or CRN) and L is the number of subjects in that group. The results are shown in Fig. 5. In all cases, the information flow within clusters is stronger than the flow between clusters, indicating that nodes within a cluster share more information compared to nodes between clusters, consistent with the definition of a cluster. For CRN, the information flow within the left-posterior region is higher than the information exchange in the other regions. While for ERN, there is high information exchange within the anterior-central region, and this value is higher compared to CRN. In addition, we determine the asymmetry of information flow between cluster pairs by applying a sign test to test the null hypothesists that the median flow from cluster C_i to C_j is equal to that from cluster C_j to C_i. The Bonferroni method is used for multiple testing correction [74]. For both CRN and ERN, there is no significant difference between the information flow in two directions between clusters which indicates that the information flow between clusters is almost symmetric. However, we should note this is an average information flow across subjects and regions, and does not necessarily mean that all regions exchange information in a symmetric manner. Another reason for close to symmetric information flow might be the computation of time-lagged DI over a small time window.

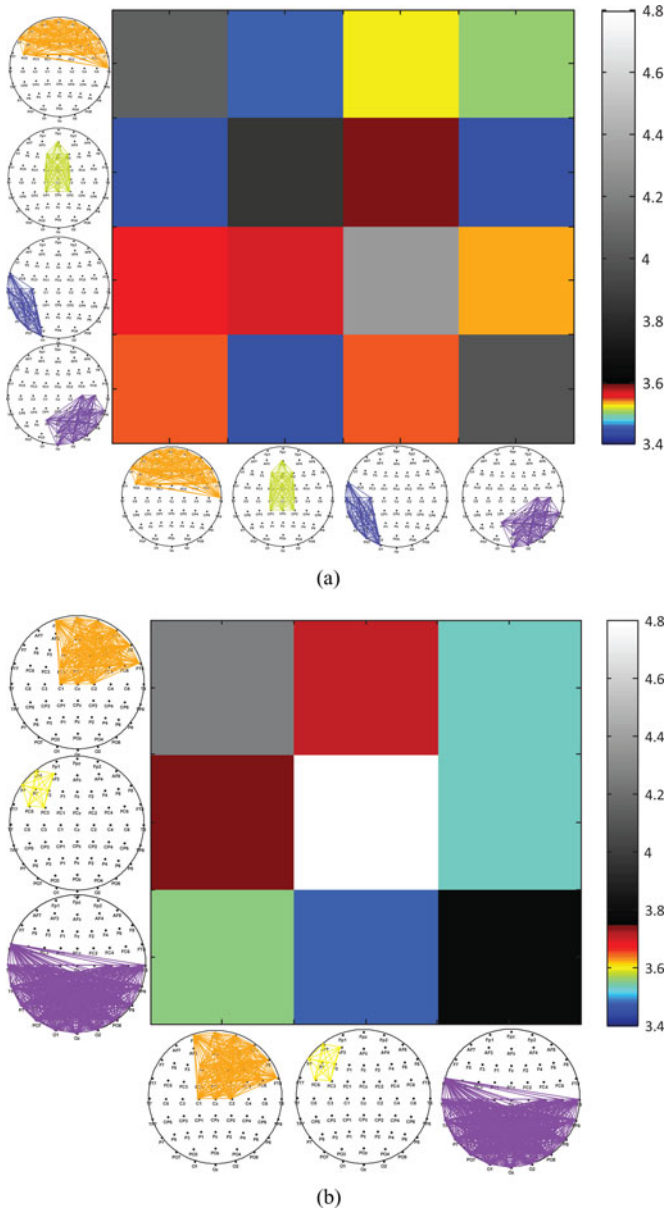

Fig. 5. Information flow within and between clusters. The direction of flow is from row to column. (a) Information flow for the last partition level for CRN. (b) Information flow for the last partition level for ERN.

VI. Conclusion

In this paper, we proposed a new method to identify the communities in effective brain connectivity networks. In order to achieve this goal, first, we applied the directed information measure to EEG data involving a study of ERN to obtain the association matrix of the network. Directed information can effectively detect the nonlinear causal relationships between different brain regions. In addition, we extended a modularity-based community detection algorithm proposed by Blondel *et al.* to weighted directed networks. Compared to current community detection methods, our proposed algorithm discovers the actual structure of a directed network by employing the edge direction information. Finally, we proposed a group analysis method to obtain a common community structure across subjects to address the problem of variability across subjects. This strategy decreases the effect of outliers without making any assumptions about the data. The proposed group analysis method has been shown to achieve higher accuracy in community detection compared to standard approaches, such as VTS and IS. It is also applied to EEG data and is shown to discriminate between error and correct responses in terms of the resulting community structures. Our findings that errors are characterized by a directed network involving central and lateral frontal regions is consistent with models of cognitive control in which the mPFC serves to monitor behavior and sends error information to lPFC in order to modulate behavior and improve performance [75], [76]. In particular, our directional findings that the flow of information in the error network moves from the central to lateral anterior locations is strong support (maybe the first of its kind) for the notion that the monitor (mPFC) is engaged first and then passes

critical performance information on to the controller (lPFC) that implements the control.

Future work will focus on improvements to the implementation of the proposed community detection algorithm. The proposed algorithm is based on the optimization of modularity. However, modularity optimization encounters the problem of resolution limit, which indicates that it may miss detecting clusters whose size is comparatively small to the whole graph [38]. Therefore, it would be of interest to investigate and extend methods that do not depend on modularity optimization, e.g., the Markov Clustering algorithm based on the dynamical evolution of walkers moving on the graph [77], to find the common communities across a group of weighted directed networks. In addition, one can consider overlapping communities by extending the current framework to consider multiple community memberships. The proposed community detection framework for effective connectivity networks can also be applied to networks obtained through other measures of causality such as Granger causality-based measures or phase slope index [78]. This work can also be extended to dynamic networks and detect the change of modules across time and frequency [79]. Finally, the proposed methods are applied to signals obtained through a CSD transformation of the scalp EEG data, and thus the inferred effective connectivity patterns are limited in terms of spatial localization. This problem can be addressed by estimating the cortical current density distribution from the scalp EEG [80], and then applying DI which would allow for a stronger inference of the inter-regional connectivity in the source domain.

REFERENCES

[1] Y. Liu, J. Moser, and S. Aviyente, "Community detection for directional neural networks inferred from EEG data," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2011, pp. 7155–7158.

[2] B. He, L. Yang, C. Wilke, and H. Yuan, "Electrophysiological imaging of brain activity and connectivity challenges and opportunities," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 7, pp. 1918–1931, Jul. 2011.

[3] K. J. Friston, "Modalities, modes, and models in functional neuroimaging," *Science*, vol. 326, no. 5951, pp. 399–403, 2009.

[4] B. He, T. Coleman, G. M. Genin, G. Glover, X. Hu, N. Johnson, T. Liu, S. Makeig, P. Sajda, and K. Ye, "Grand challenges in mapping the human brain: NSF workshop report," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 11, pp. 2983–2992, 2013.

[5] S. Mori, K. Oishi, and A. V. Faria, "White matter atlases based on diffusion tensor imaging," *Curr. Opin. Neurol.*, vol. 22, no. 4, pp. 362–369, 2009.

[6] K. Friston, "Functional and effective connectivity: A review," *Brain Connect.*, vol. 1, no. 1, pp. 13–36, 2011.

[7] E. Pereda, R. Q. Quiroga, and J. Bhattacharya, "Nonlinear multivariate analysis of neurophysiological signals," *Progr. Neurobiol.*, vol. 77, pp. 1–37, 2005.

[8] O. David, D. Cosmelli, and K. J. Friston, "Evaluation of different measures of functional connectivity using a neural mass model," *Neuroimage*, vol. 21, no. 2, pp. 659–673, 2004.

[9] G. L. Gerstein and D. H. Perkel, "Simultaneously recorded trains of action potentials: analysis and functional interpretation," *Science*, vol. 164, no. 3881, pp. 828–830, 1969.

[10] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nat. Rev. Neurosci.*, vol. 10, no. 3, pp. 186–198, 2009.

[11] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, 2010.

[12] D. Bassett and E. Bullmore, "Small-world brain networks," *Neuroscientist*, vol. 12, no. 6, pp. 512–523, 2006.

[13] O. Sporns and J. D. Zwi, "The small world of the cerebral cortex," *Neuroinformatics*, vol. 2, no. 2, pp. 145–162, 2004.

[14] S. Achard, R. Salvador, B. Whitcher, J. Suckling, and E. Bullmore, "A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs," *J. Neurosci.*, vol. 26, no. 1, pp. 63–72, 2006.

[15] S. Achard and E. Bullmore, "Efficiency and cost of economical brain functional networks," *PLoS Computat. Biol.*, vol. 3, no. 2, e17, 2007. doi:10.1371/journal.pcbi.0030017.

[16] E. Ravasz and A.-L. Barabási, "Hierarchical organization in complex networks," *Phys. Rev. E*, vol. 67, no. 2, pp. 026112-1–026112-12, 2003. doi: 10.1103/PhysRevE.67.026112.

[17] C. Zhou, L. Zemanová, G. Zamora, C. C. Hilgetag, and J. Kurths, "Hierarchical organization unveiled by functional connectivity in complex brain networks," *Phys. Rev. Lett.*, vol. 97, no. 23, pp. 238103-1–238103-4, 2006. doi: 10.1103/PhysRevLett.97.238103.

[18] M. Barthelemy, "Betweenness centrality in large complex networks," *Eur. Phys. J. B-Condens. Matter Complex Syst.*, vol. 38, no. 2, pp. 163–168, 2004.

[19] O. Sporns, C. J. Honey, and R. Kötter, "Identification and classification of hubs in brain networks," *PloS one*, vol. 2, no. 10, e1049, 2007, doi:10.1371/journal.pone.0001049.

[20] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 12, pp. 7821–7826, 2002.

[21] M. Chavez, M. Valencia, V. Navarro, V. Latora, and J. Martinerie, "Functional modularity of background activities in normal and epileptic brain networks," *Phys. Rev. Lett.*, vol. 104, no. 11, 118701, 4 pp., 2010. doi: 10.1103/PhysRevLett.104.118701.

[22] A. J. Schwarz, A. Gozzi, and A. Bifone, "Community structure and modularity in networks of correlated brain activity," *Magnet. Reson. Imag.*, vol. 26, no. 7, pp. 914–920, 2008.

[23] M. E. J. Newman, "Detecting community structure in networks," *Eur. Phys. J. B-Condens. Matter Complex Syst.*, vol. 38, no. 2, pp. 321–330, 2004.

[24] L. Ferrarini, I. M. Veer, E. Baerends, M.-J. van Tol, R. J. Renken, N. J.A. van der Wee, D. J. Veltman, A. Aleman, F. G. Zitman, B. W. J. H. Penninx, M. A. van Buchem, J. H. C. Reiber, S. A. R. B. Rombouts, and J. Milles, "Hierarchical functional modularity in the resting-state human brain," *Human Brain Mapp.*, vol. 30, no. 7, pp. 2220–2231, 2009.

[25] D. A. Fair, A. L. Cohen, J. D. Power, N. U. Dosenbach, J. A. Church, F. M. Miezin, B. L. Schlaggar, and S. E. Petersen, "Functional brain networks develop from a "local to distributed" organization," *PLoS Computat. Biol.*, vol. 5, no. 5, pp. 1–14, 2009.

[26] D. Meunier, S. Achard, A. Morcom, and E. Bullmore, "Age-related changes in modular organization of human brain functional networks," *Neuroimage*, vol. 44, no. 3, pp. 715–723, 2009.

[27] E. A. Leicht and M. E. J. Newman, "Community structure in directed networks," *Phys. Rev. Lett.*, vol. 100, no. 11, 118703, 4 pp., 2008. doi: 10.1103/PhysRevLett.100.118703.

[28] Y. Liu and S. Aviyente, "Directed information measure for quantifying the information flow in the brain," in *Proc. Int. Conf. IEEE, Eng. Med. Biol. Soc.*, 2009, pp. 2188–2191.

[29] Y. Liu and S. Aviyente, "Quantification of effective connectivity in the brain using a measure of directed information," *Computat. Math. Methods Med.*, vol. 2012, pp. 1–16, 2012.

[30] J. Massey, "Causality, feedback, and directed information," in *Proc. Symp. Inf. Theory Appl.*, 1990, pp. 27–30.

[31] H. Hinrichs, T. Noesselt, and H. J. Heinze, "Directed information flow model free measure to analyze causal interactions in event related EEG-MEG-experiments," *Human Brain Mapp.*, vol. 29, no. 2, pp. 193–206, 2008.

[32] A. Rao, A. O. Hero, III, D. J. States, and J. D. Engel, "Using directed information to build biologically relevant influence networks," in *Proc. Computat. Syst. Bioinformat.*, 2007, pp. 145–156.

[33] C. J. Quinn, T. P. Coleman, N. Kiyavash, and N. G. Hatsopoulos, "Estimating the directed information to infer causal relationships in ensemble neural spike train recordings," *J. Comput. Neurosci.*, vol. 30, no. 1, pp. 17–44, 2011.

[34] V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Statist. Mech., Theory Exp.*, vol. 2008, P10008, 12 pp., 2008.

[35] M. E. J. Newman, "Modularity and community structure in networks," *Proc. Nat. Acad. Sci.*, vol. 103, no. 23, pp. 8577–8582, 2006.

[36] C. Dorso and A. Medus, "Community detection in networks," *Int. J. Bifurcat. Chaos*, vol. 20, no. 02, pp. 361–367, 2010.

[37] J. Duch and A. Arenas, "Community detection in complex networks using extremal optimization," *Phys. Rev. E*, vol. 72, no. 2, 027104, 4 pp., 2005. doi: 10.1103/PhysRevE.72.027104.

[38] S. Fortunato, "Community detection in graphs," *Phys. Rep.*, vol. 486, no. 3-5, pp. 75–174, 2010.
[39] U. Von Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395–416, 2007.
[40] A. Arenas, J. Duch, A. Fernández, and S. Gómez, "Size reduction of complex networks preserving modularity," *New J. Phys.*, vol. 9, 176, 15 pp., 2007.
[41] M. E. J. Newman, "Fast algorithm for detecting community structure in networks," *Phys. Rev. E*, vol. 69, no. 6, 066133, 5 pp., 2004. doi: 10.1103/PhysRevE.69.066133.
[42] M. E. Newman and E. A. Leicht, "Mixture models and exploratory analysis in networks," *Proc. Nat. Acad. Sci.*, vol. 104, no. 23, pp. 9564–9569, 2007.
[43] R. Guimerà, M. Sales-Pardo, and L. A. N. Amaral, "Module identification in bipartite and directed networks," *Phys. Rev. E*, vol. 76, no. 3, 036102, 8 pp., 2007. doi: 10.1103/PhysRevE.76.036102.
[44] D. Meunier, R. Lambiotte, A. Fornito, K. D. Ersche, and E. T. Bullmore, "Hierarchical modularity in human brain functional networks," *Front. Neuroinformat.*, vol. 3, pp. 1–12, 2009.
[45] J. Li, Z. J. Wang, S. J. Palmer, and M. J. McKeown, "Dynamic Bayesian network modeling of fMRI: A comparison of group-analysis methods," *NeuroImage*, vol. 41, no. 2, pp. 398–407, 2008.
[46] A. Mechelli, W. D. Penny, C. J. Price, D. R. Gitelman, and K. J. Friston, "Effective connectivity and intersubject variability: Using a multisubject network to test differences and commonalities," *Neuroimage*, vol. 17, no. 3, pp. 1459–1469, 2002.
[47] G. Karypis and V. Kumar, "A fast and high quality multilevel scheme for partitioning irregular graphs," *SIAM J. Scientif. Comput.*, vol. 20, no. 1, pp. 359–392, 1998.
[48] A. Strehl and J. Ghosh, "Cluster ensembles—A knowledge reuse framework for combining multiple partitions," *J. Mach. Learn. Res.*, vol. 3, pp. 583–617, 2003.
[49] N. M. Correa, T. Adali, Y. O. Li, and V. D. Calhoun, "Canonical correlation analysis for data fusion and group inferences," *IEEE Signal Process. Mag.*, vol. 27, no. 4, pp. 39–50, 2010.
[50] C. W. J. Granger, "Testing for causality: A personal viewpoint," *J. Econ. Dyn. Control*, vol. 2, pp. 329–352, 1980.
[51] L. A. Baccala and K. Sameshima, "Partial directed coherence: A new concept in neural structure determination," *Biolog. Cybern.*, vol. 84, no. 6, pp. 463–474, 2001.
[52] K. Friston, L. Harrison, and W. Penny, "Dynamic causal modelling," *Neuroimage*, vol. 19, no. 4, pp. 1273–1302, 2003.
[53] S. Shimizu, P. Hoyer, A. Hyvärinen, and A. Kerminen, "A linear non-gaussian acyclic model for causal discovery," *J. Mach. Learn. Res.*, vol. 7, pp. 2003–2030, 2006.
[54] T. Schreiber, "Measuring information transfer," *Phys. Rev. Lett.*, vol. 85, no. 2, pp. 461–464, 2000.
[55] Y. Saito and H. Harashima, *Recent Advances in EEG and EMG Data Processing*. Amsterdam, The Netherlands: Elsevier, 1981.
[56] P. O. Amblard and O. J. Michel, "On directed information theory and granger causality graphs," *J. Computat. Neurosci.*, vol. 30, no. 1, pp. 7–16, 2010.
[57] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 4–21, Jan. 2003.
[58] E. G. Miller, "A new class of entropy estimators for multi-dimensional densities," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2003, vol. 3, pp. 297–300.
[59] G. A. Darbellay and I. Vajda, "Estimation of the information by an adaptive partitioning of the observation space," *IEEE Trans. Inf. Theory*, vol. 45, no. 4, pp. 1315–1321, May 1999.
[60] Y. Liu and S. Aviyente, "Time-lagged directed information," in *Proc. IEEE Intl. Conf. Acoust., Speech Signal Process.*, 2011, pp. 3864–3867.
[61] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vettering, *Numerical recipes: The art of scientific computing*, 3rd ed. ed. Cambridge, U.K.: Cambridge Univ. Press, 2007.
[62] J. Cohen, "Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit.," *Psycholog. Bull.*, vol. 70, no. 4, pp. 213–220, 1968.
[63] J. Garner, "The standard error of cohen's kappa," *Statist. Med.*, vol. 10, no. 5, pp. 767–775, 2006.
[64] J. S. Moser, H. S. Schroder, C. Heeter, T. P. Moran, and Y.-H. Lee, "Mind your errors: Evidence for a neural mechanism linking growth mindset to adaptive post-error adjustments," *Psycholog. Sci.*, vol. 22, no. 12, pp. 1484–1489, 2011.
[65] D. M. Olvet and G. Hajcak, "The stability of error-related brain activity with increasing trials," *Psychophysiology*, vol. 46, no. 5, pp. 957–961, 2009.
[66] B. A. Eriksen and C. W. Eriksen, "Effects of noise letters upon the identification of a target letter in a nonsearch task," *Percept. Psychophys.*, vol. 16, no. 1, pp. 143–149, 1974.
[67] [Online]. Available:http://www.biosemi.com/faq/cms&drl.htm
[68] G. Gratton, M. G. H. Coles, and E. Donchin, "A new method for off-line removal of ocular artifact," *Electroencephalogr. Clin. Neurophysiol.*, vol. 55, no. 4, pp. 468–484, 1983.
[69] S. Aviyente, E. Bernat, W. Evans, and S. Sponheim, "A phase synchrony measure for quantifying dynamic functional integration in the brain," *Human Brain Mapp.*, vol. 32, no. 1, pp. 80–93, 2011.
[70] J. Kayser and C. E. Tenke, "Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns—Part I: Evaluation with auditory oddball tasks," *Clin. Neurophysiol.*, vol. 117, no. 2, pp. 348–368, 2006.
[71] C. Tenke and J. Kayser, "Generator localization by current source density (CSD): Implications of volume conduction and field closure at intracranial and scalp resolutions," *Clin. Neurophysiol.*, vol. 123, no. 12, pp. 2328–2345, 2012.
[72] D. N. Politis and J. P. Romano, "The stationary bootstrap," *J. Amer. Statist. Assoc.*, vol. 89, no. 428, pp. 1303–1313, 1994.
[73] M. Chávez, J. Martinerie, and M. Le Van Quyen, "Statistical assessment of nonlinear causality: Application to epileptic EEG signals," *J. Neurosci. Methods*, vol. 124, no. 2, pp. 113–128, 2003.
[74] C. R. Genovese, N. A. Lazar, and T. Nichols, "Thresholding of statistical maps in functional neuroimaging using the false discovery rate," *Neuroimage*, vol. 15, no. 4, pp. 870–878, 2002.
[75] E. K. Miller and J. D. Cohen, "An integrative theory of prefrontal cortex function," *Annu. Rev. Neurosci.*, vol. 24, pp. 167–202, 2001.
[76] M. Botvinick, T. Braver, D. Barch, C. Carter, and J. Cohen, "Conflict monitoring and cognitive control," *Psycholog. Rev.*, vol. 108, no. 3, pp. 624–652, 2001.
[77] R. Sinatra, F. de Vico Fallani, L. Astolfi, F. Babiloni, F. Cincotti, D. Mattia, and V. Latora, "Cluster structure of functional networks estimated from high-resolution EEG data," *Intl. J. Bifurcat. Chaos*, vol. 19, no. 2, pp. 665–676, 2009.
[78] S. Haufe, V. Nikulin, K. Müller, and G. Nolte, "A critical assessment of connectivity measures for EEG data: A simulation study," *NeuroImage*, vol. 64, pp. 120–133, 2012.
[79] S. Boccaletti, M. Ivanchenko, V. Latora, A. Pluchino, and A. Rapisarda, "Detecting complex network modularity by dynamical clustering," *Phys. Rev. E*, vol. 75, no. 4, 045102, 4 pp., 2007, doi: 10.1103/PhysRevE.75.045102.
[80] Z. Liu and B. He, "fMRI-EEG integrated cortical source imaging by use of time-variant spatial constraints," *NeuroImage*, vol. 39, no. 3, pp. 1198–1214, 2008.

Authors' photographs and biographies not available at the time of publication.