

# Stability of Network Communities as a Function of Task Complexity

Stefan Fuertinger and Kristina Simonyan

## Abstract

■ The analysis of the community architecture in functional brain networks has revealed important relations between specific behavioral patterns and characteristic features of the associated functional organization. Numerous studies have assessed changes in functional communities during different states of awareness, learning, information processing, and various behavioral patterns. The robustness of detected communities within a network has been an often-discussed topic in complex systems research. However, our knowledge regarding the intersubject stability of functional communities in the human brain while performing different tasks is still lacking. In this study, we examined the variability of functional communities in weighted undirected graphs based on fMRI recordings of healthy participants across three conditions: the resting state, syllable production as a simple vocal motor task, and meaningful speech production

representing a complex behavioral pattern with cognitive involvement. On the basis of the constructed empirical networks, we simulated a large cohort of artificial graphs and performed a leave-one-out stability analysis to assess the sensitivity of communities in the group-averaged networks with respect to perturbations in the averaging cohort. We found that the stability of partitions derived from group-averaged networks depended on task complexity. The determined community architecture in mean networks reflected within-behavior network stability and between-behavior flexibility of the human functional connectome. The sensitivity of functional communities increased from rest to syllable production to speaking, which suggests that the approximation quality of the community structure in the average network to reflect individual per-participant partitions depends on task complexity. ■

## INTRODUCTION

Considerable advancements in imaging technology drastically improved the quality and level of detail of neuroimaging data sets collected over the past years (Van Essen et al., 2013). Although this development presented obvious advantages for knowledge generation, it also came with a number of problems arising in the analysis of complex high-dimensional data. Thus, many fMRI studies started utilizing abstract mathematical concepts rooted in graph theory (Bondy & Murty, 1976) to represent and subsequently analyze brain activity data in terms of large-scale graphs (Wang et al., 2014; van den Heuvel & Sporns, 2011; Lynall et al., 2010; Bassett & Bullmore, 2006; Humphries, Gurney, & Prescott, 2006). The mathematical concept of a graph, which represents a collection of nodes and edges, proved to be particularly well suited to comprehensively assess the whole-brain functional connectome (De Vico Fallani, Richiardi, Chavez, & Achard, 2014; Bullmore & Sporns, 2009). The analysis of the community architecture within these graphs revealed a number of previously unknown brain activity patterns, such as the context-dependent global reorganization of brain activity not necessarily restricted to underlying white matter connections (Andric & Hasson, 2015), the predictive power of

nodal assignments to functional modules for learning performance (Bassett et al., 2011), the estimation of cognitive performance based on the modular organization of a cognitive control network at rest (Stevens, Tappan, Garg, & Fair, 2012), and the modular disorganization and a breakdown of normal network architecture preceding an epileptic seizure (Fuertinger, Simonyan, Sperling, Sharan, & Hamzei-Sichani, 2016). The inherent structure of these functional patterns most likely would have remained undetected without using the theoretical underpinning of graph partitions.

However, finding a graph's community structure as well as analyzing and comparing network partitions present several technical challenges itself (Sporns & Betzel, 2015). One of the most widely used approaches to partition a network into communities, or modules, is the technique of modularity maximization (Porter, Onnela, & Mucha, 2009; Newman & Girvan, 2004). Maximizing a network's Newman modularity score is a computationally efficient way to simultaneously determine both the number of nodal clusters and their members (Newman, 2006). However, modularity maximization also experiences two well-known conceptual problems: resolution limit (Fortunato & Barthelemy, 2007), which, under certain conditions, prevents the detection of small communities in a graph, and module degeneracy (Good, de Montjoye, & Clauset, 2010), caused by numerous different network partitions with near-optimal modularity scores. The latter aspect is closely related

to the problem of comparing network partitions and assessing their quality in reflecting the “true” underlying community structure of a graph. These methodological issues, together with the fact that neuroimaging studies traditionally focus on group-level inferences, are likely reasons for many graph-theory-based functional imaging studies to center on analyzing the community structure of group-averaged networks (Bertolero, Yeo, & D’Esposito, 2015; Betzel et al., 2015; Godwin, Barry, & Marois, 2015; Onoda & Yamaguchi, 2013). A few studies have demonstrated pronounced effects of intersubject variability on functional coupling patterns, particularly during complex cognitive tasks (Miranda-Dominguez et al., 2014; Grabner et al., 2007; Newman, Carpenter, Varma, & Just, 2003; Rypma & D’Esposito, 1999). However, the impact of individual variability in functional network architecture on the community layout of group-level mean networks across different experimental conditions is still largely unknown.

In this study, we analyzed the variability of functional network communities across 27 healthy adult participants with respect to a group-wide reference given by the modular decomposition of the group-averaged networks. We assessed three different conditions: the resting state as a baseline status, syllable production representing a simple vocal motor task with minimal cognitive involvement, and the production of meaningful sentences as a highly complex speech motor behavior. Furthermore, we performed an in-depth assessment of the effects of perturbations in the averaging process on the architecture of the mean networks by simulating 100 graphs per condition and conducted a comprehensive leave-one-out analysis of the stability of mean network communities. Using different variations of community detection algorithms, some studies have examined the robustness of communities within a single network by probing well-studied prototype graphs, such as the karate club network (Karrer, Levina, & Newman, 2008; Gfeller, Chappelier, & De Los Rios, 2005; Wu & Huberman, 2004; Zachary, 1977). The aim of this study was to analyze the within-group stability and sensitivity of functional communities with respect to the experimental condition. We hypothesized that the community architecture was both specific and sensitive to each condition and driven by its complexity.

## METHODS

### Ethics Statement

All participants provided written informed consent before study participation, which was approved by the institutional review boards of the Icahn School of Medicine at Mount Sinai and the National Institute of Neurological Disorders and Stroke, National Institutes of Health.

### Data Acquisition

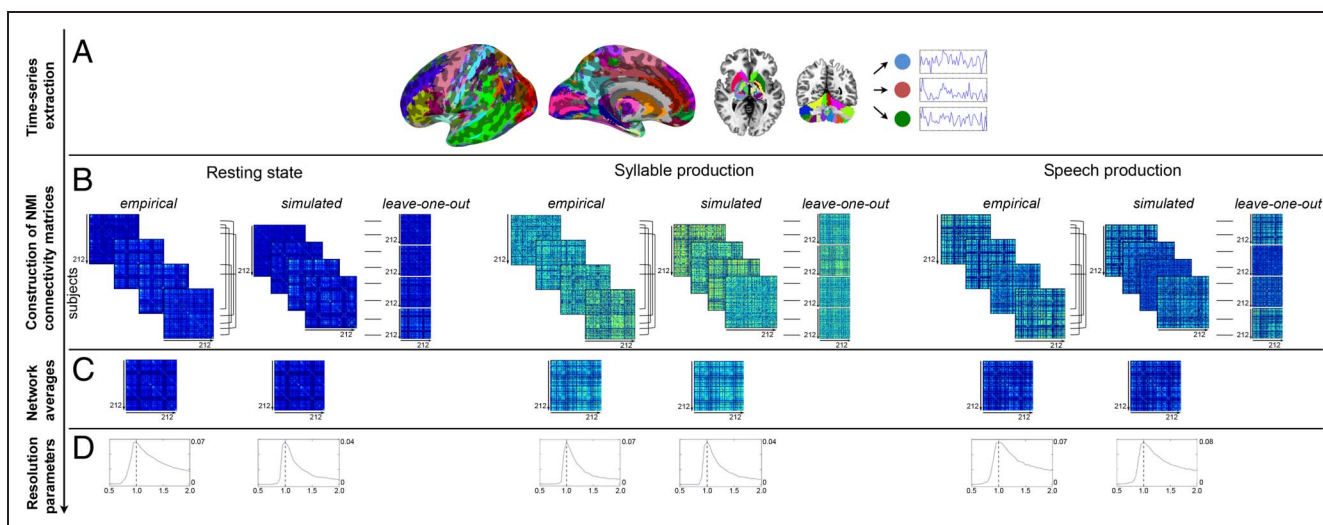
Twenty-seven right-handed monolingual English-speaking healthy participants (18 women, 9 men; mean age =

52.2 years,  $SD = 11.3$  years) without any history of neurological, psychiatric, or laryngeal disorders were recruited for this study.

Brain images were acquired on a 3-T GE scanner (Milwaukee, WI). Whole-brain resting-state fMRI images were acquired before the task-production fMRI within the same scanning session. Participants were instructed to rest without specific thoughts, with their eyes closed in an environment with dimmed light. Participants were continuously monitored during scanning and debriefed after the scan; no participant reported falling asleep during the scanning session. Functional images were acquired with a gradient-weighted EPI pulse sequence (repetition time = 2 sec, 150 contiguous volumes, echo time [TE] = 30 msec, flip angle [FA] = 90°, 33 slices with 3.75-mm in-plane resolution, slice thickness = 4 mm). Syllable and sentence production fMRI data were acquired with a gradient-weighted EPI pulse sequence (total repetition time = 10.6 sec, comprising 3.6-sec delay for stimulus presentation, 5-sec delay for task production, and 2-sec image acquisition; TE = 30 msec; FA = 90°; 36 slices with 3.75-mm in-plane resolution; slice thickness = 4 mm) using a BOLD contrast and an event-related sparse-sampling design. As a simple vocal motor task, all participants produced the same syllables /i?i-i?i/, which is composed of the vowel /i/ (as in “beet”) and the glottal stop /ʔ/ followed again by /i/, which maximized vocal fold adduction while minimizing orofacial articulation. As a more cognitively and linguistically complex speech motor task, all participants produced the same 10 meaningful, grammatically correct English sentences (e.g., “We are always away,” “Tom is in the army”), one per acquisition volume (Fuertinger, Horwitz, & Simonyan, 2015). The speech production task was designed to encompass higher-order cognitive and linguistic processing, whereas syllable production represented a simple vocal motor task that was relevant to speaking but had minimal cognitive and linguistic involvement (Fuertinger et al., 2015; Simonyan & Fuertinger, 2015; Simonyan, Herscovitch, & Horwitz, 2013). All sample stimuli were acoustically presented one at a time and performed by the participant one per acquisition volume. Thirty-six trials per each task (syllable or sentence production) and 24 resting fixations as a baseline were acquired over the five scanning sessions in each participant with the tasks pseudorandomized between sessions and participants. A high-resolution T1-weighted image was acquired as an anatomical reference using a magnetization-prepared rapid acquisition gradient echo sequence (inversion time = 450 msec, TE = 3.0 msec, FA = 10, 128 slices with 1.2-mm thickness).

### Construction of Functional Networks

Functional images were preprocessed using AFNI software (Cox, 1996) following a standard analysis pipeline as described earlier (Fuertinger et al., 2015). A whole-brain parcellation based on the cytoarchitectonic maximum



**Figure 1.** Illustration of the utilized analytic strategy. (A) On the basis of a 212-region whole-brain parcellation, regionally averaged BOLD time series were extracted from fMRI recordings acquired at rest, during syllable production, and while speaking. (B) NMI coefficients were calculated for each pair of ROI time series and assembled in connectivity matrices to establish empirical functional networks. By imposing a Gaussian null model on randomly sampled empirical per-subject networks, we simulated 100 artificial networks for each condition (rest, syllable, and speech). The simulated networks were used in a leave-one-out analysis paradigm to compute 100 leave-one-out mean networks by omitting one graph at a time and averaging across the remaining cohort. (C) Group-averaged networks were computed for each condition in both empirical and simulated network cohorts and were considered as points of reference within the associated group. (D) The established mean networks were used to compute group-specific resolution parameters for all subsequent modularity maximization calculations.

probability maps and macrolabel atlas (Eickhoff et al., 2005) was used to extract regionally averaged task-production and resting-state BOLD fMRI residual time series for all participants (Figure 1A). The whole-brain parcellation was composed of  $n = 212$  ROIs, including 142 cortical, 36 sub-cortical, and 34 cerebellar areas.

The interaction of ROIs during each condition (i.e., rest, syllable, and speech) was estimated by computing the normalized mutual information (NMI; Strehl & Ghosh, 2002) for each pair of regional time series. Normalizing the classic mutual information (Cover & Thomas, 1991)  $I(x,y)$  of two random variables  $x$  and  $y$  by the geometric mean of the associated Shannon entropies (Shannon, 1948)  $H(x)$  and  $H(y)$ , respectively, resulted in a metric, which assumed values close to 0 in case of statistical independence, whereas pairwise mutual dependence yielded NMI coefficients close to 1. Hence, NMI values were (unlike the classic mutual information) bounded from above by 1, allowing for quantitative comparisons across different data sets.

Following Sturges' (1926) rule, we used eight bins to discretize probability distributions for computing NMI coefficients, which resulted in an estimated inflation of the mutual information because of sample size by less than 0.24 bits (Steuer, Kurths, Daub, Weise, & Selbig, 2002). This finding indicates high approximation quality of the empirical mutual information values within each condition with negligible influence of data acquisition differences. It is, therefore, unlikely that the observed differences in the examined static functional networks were driven by variations in length of task and rest data acquisition.

Symmetric  $N \times N$  matrices were constructed by arranging the computed pairwise NMI coefficients by region.

Because of the absence of negative values, any NMI matrix can be readily interpreted as the connectivity matrix of a weighted undirected graph with ROIs representing nodes and pairwise NMI values between ROIs serving as edge weights.

In contrast, classic Pearson correlation matrices are not necessarily nonnegative, and this has to be accounted for in a graph theoretical framework. Negative values may be discarded by either considering only their absolute magnitudes or removing the corresponding edges from the graph entirely, resulting in a loss or potential corruption of the underlying correlative information. Instead, the positive and negative components of a graph may be analyzed separately, although this approach might result in unconnected subgraphs. Alternatively, classic network metrics may be redefined and extended to account for negative edge weights; however, the potential existence of negative cycles proved to be especially problematic for measures relying on the notion of paths through a graph (Fakcharoenphol & Rao, 2006). To circumvent these issues and their inherent impact on network architecture, we chose to use the NMI as a statistical dependence metric to construct functional networks. In this manner, three groups of functional networks were constructed,  $(F_k^r)_{k \geq 1}$ ,  $(F_k^{\text{sy}l})_{k \geq 1}$ , and  $(F_k^{\text{sp}})_{k \geq 1}$  (with  $k$  denoting the participant number), corresponding to the resting state, syllable production, and speaking, respectively (Figure 1B).

In a second step, we simulated 300 networks, including 100 for each resting  $(G_k^r)_{k \geq 1}$ , syllable production  $(G_k^{\text{sy}l})_{k \geq 1}$ , and speaking  $(G_k^{\text{sp}})_{k \geq 1}$  conditions, by imposing a Gaussian random graph model on randomly sampled empirical networks from the corresponding groups (Figure 1B; Wang

et al., 2011; Ghosh, Rho, McIntosh, Kotter, & Jirsa, 2008). Specifically, let  $F$  be a fixed but arbitrary empirical network, and let  $F(m,n)$  denote the edge weight between nodes  $m$  and  $n$ . Given  $F$ , a simulated network  $G$  is obtained by imposing a Gaussian null graph with independent, identically distributed edges of zero mean and constant variance given by the scaled median of edge weights within each group (rest, syllable, and speech). Then, the expected edge weight between nodes  $m$  and  $n$  in  $G$  is given by  $E(G(m,n) | s_1, \dots, s_N) = (s_m + s_n)/(N - 2) - 2S/[(N - 1)(N - 2)]$  for  $m \neq n$  and  $E(G(m,n) | s_1, \dots, s_N) = 0$ ; otherwise, where  $s_m = \sum_n F(m,n)$  denotes the strength of node  $m$  in  $F$  and  $S = \frac{1}{2} \sum_m s_m$  is the total nodal strength in  $F$  (Chang, Leahy, & Pantazis, 2012). Imposing a Gaussian graph on  $F$  as a null model to generate  $G$  ensures that the basic topology of the empirical network  $F$ , including its degree distribution, is preserved, while simultaneously introducing stochastic variability in the simulated network cohort (Chang et al., 2012). Consequently, this strategy injects stochastic alterations in the modular architecture of  $F$  to construct a simulated graph  $G$  of comparable but randomly adapted modular architecture.

### Network Preprocessing

We will use the shorthand notation  $(F_k)_{k \geq 1}$  and  $(G_k)_{k \geq 1}$  to refer to all (rest, syllable, speech) empirical and simulated networks, respectively. The connection density of both  $(F_k)_{k \geq 1}$  and  $(G_k)_{k \geq 1}$  was calculated in every graph as the number of actual edges divided by the maximum possible number of edges in the graph. Thus, the density of an undirected graph with  $N$  nodes is given by  $2K/(N^2 - N)$ , with  $K$  denoting the actual edge count in the network. All networks (empirical and simulated) showed a density  $\geq 99\%$ , indicating that most graphs were almost fully connected. To allow an assessment of edge clustering and connection patterns, a density-dependent iterative thresholding approach was applied to every graph. Edges with weight less than a subsequently incremented percentage of the maximum weight in the graph (starting with 1%) were removed from the network. The iteration was stopped if the removal of an edge would fully disconnect a node from the rest of the network. It has been shown that network architecture becomes increasingly random above a density of 50% (Lynall et al., 2010; Humphries et al., 2006); thus, the connection density of all graphs was reduced to the within-group minimum density, which was calculated as the lowest common density to which every network in the group could be thresholded without disconnecting a node. In this way, empirical graphs were reduced to 65% (rest), 63% (syllable), and 50% (speech) densities, whereas the thresholding of simulated networks resulted in densities of 60% (rest), 68% (syllable), and 47% (speech). Given the size of the considered networks ( $N > 200$ ), the obtained within-group minimum densities were comparable across conditions (van Wijk, Stam, & Daffertshofer, 2010).

One of the most widely used ways to represent group effects in graph theoretical assessments of neuroimaging

data is to consider group-averaged networks (Ginestet, Fournel, & Simmons, 2014), which we thus used as group-specific points of reference in our subsequent topological graph analysis. Group-averaged networks were computed by calculating mean edge weights within each condition (rest, syllable, and speech) in both empirical and simulated networks (Figure 1C). We denote group-averaged networks by the subscript “0” ( $F_0^r$ ,  $G_0^r$ , etc.) and drop the superscript “r,” “syl,” and “sp” to indicate mean networks of each condition, namely, the resting state, syllable production, and speaking ( $F_0$  and  $G_0$ ). To further assess the effects of perturbations in the averaging set on the topology of the mean network, we implemented a leave-one-out analysis paradigm in the simulated network cohort, which was composed of the most samples per group (rest, syllable, and speech). Although an  $n$ -fold cross-validation approach would have been computationally more efficient, it was a priori not clear how to choose the number of withheld samples  $n$  appropriately to not introduce an estimation bias (Hastie, Tibshirani, & Friedman, 2009). Hence, we performed a leave-one-out validation using a large cohort of simulated networks to minimize any undue influence of a single empirical subject on the community layout in the group average. Thus, we computed 100 leave-one-out mean networks per condition ( $(G_{0 \setminus k}^r)_{k \geq 1}$ ,  $(G_{0 \setminus k}^{\text{syl}})_{k \geq 1}$ , and  $(G_{0 \setminus k}^{\text{sp}})_{k \geq 1}$ , corresponding to the resting state, syllable, and speech production, respectively, by averaging across the graphs  $(G_j^r)$ ,  $(G_j^{\text{syl}})$ , and  $(G_j^{\text{sp}})$ , respectively, for  $j = 1, \dots, 100, j \neq k$  (Figure 1B).

### Analysis of Network Topology

The architecture of each network was analyzed by estimating its optimal modular decomposition. The terms “community” and “module” will be used interchangeably in this article to denote groups of nodes that form densely interconnected subnetworks within a graph. Most community detection strategies that are used to analyze functional brain networks yield partitions, in which each node is unambiguously assigned to exactly one module (Sporns & Betzel, 2015). In line with this, we defined a modular decomposition of a graph to be a partition that divided a network into nonoverlapping nodal communities. One of the most widely used techniques to compute a network’s modular decomposition consists of maximizing a graph’s Newman modularity score  $Q$  (Newman & Girvan, 2004). The modularity score of a network is a statistic that quantifies the degree to which a network can be divided into nodal groups that show a higher internal connection density than would be expected by chance. Thus, maximizing  $Q$  ideally uncovers the partition with the lowest between-module and densest within-group connectivity, which thus best approximates the network’s inherent community structure.

However, the global null model used to predict between-group connectivity in the calculation of  $Q$  is based on the assumption that every node can establish connections to

any other node in the network. As a consequence, merging two small but clearly delineated communities in a sufficiently large network might increase the value of  $Q$  (Fortunato & Barthelemy, 2007). To counteract this effect known as the resolution limit of modularity, we used a multiresolution approach (Bassett et al., 2013) for community detection. A resolution parameter  $\gamma$  was introduced that scaled the null model in the calculation of  $Q$ , so that smaller communities were detected for  $\gamma > 1$ , whereas  $0 < \gamma \leq 1$  resulted in larger modules. To balance modularity resolution for each of the six groups (rest, syllable, and speech conditions in the empirical and simulated network cohorts), the optimal modular decomposition of each group-averaged network was approximated by maximizing  $Q(\gamma)$  for  $\gamma$  in  $[0.5, 2.0]$  (30 values in increments of 0.05). Simultaneously, we computed a modularity score  $Q^{\text{rnd}}(\gamma)$  for each value of  $\gamma$  as the average modularity across 20 random networks with the same degree distribution as the corresponding group-averaged graph. Maximizing the quality function  $f(\gamma) = Q(\gamma) - Q^{\text{rnd}}(\gamma)$ , that is, the difference between each network's actual score and a corresponding random graph's modularity score, favored the associated modular architecture that deviated the most from a partition expected by chance under the null model. To reduce computational load and ensure that the scale of partitions was consistent within each group, the maximizer  $\gamma = \text{argmax} f(\gamma)$  for the respective mean network was used as the resolution parameter for all graphs in the underlying group (Figure 1D).

Another methodological drawback of modularity maximization is rooted in its nature as global optimization problem. With increasing network size and community count, the modularity score  $Q$  is characterized by numerous local maxima, causing a degeneracy of near-optimal but dissimilar network partitions (Good et al., 2010). To address this phenomenon, we approximated the optimal modular decomposition of a graph using consensus communities, based on clustering a partition association matrix (Lancichinetti & Fortunato, 2012). A heuristic optimization strategy based on a multi-iterative generalization of the Louvain community detection algorithm (Rubinov & Sporns, 2010; Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) was used to maximize  $Q$ , given the calculated group-specific resolution parameters. In this manner, 100 near-optimal modular decompositions were calculated for every network, which formed the basis of an association matrix  $A$ , with elements  $A_{ij}$  representing the number of times nodes  $i$  and  $j$  were assigned to the same community, weighted by each partition's quality as quantified by its modularity score  $Q$ . To eliminate small entries of  $A$  and thus improve clustering results, we applied an automated thresholding strategy (Bassett et al., 2013). The cutoff threshold  $\tau$  was defined as the maximal element across 10,000 null association matrices constructed based on random permutations of the previously calculated 100 partitions. Every element  $A_{ij} < \tau$  was set to zero, and the resulting association matrix was partitioned again 100 times using the Louvain

algorithm, generating a new association matrix. If necessary, this step was repeated until partitions converged to a single-limit partition, which eventually represented the final consensus decomposition of the network.

Because  $Q$  can be maximized even for synthetic graphs with random topology, the successful computation of a network partition does not necessarily imply that a graph is really modular. To assess whether the computed consensus communities were indicative of a truly modular architecture of the underlying networks, we compared the partition of every empirical and simulated group-averaged network,  $F_0$  and  $G_0$ , respectively, to random graphs with identical degree distributions. As a first step, the  $Q$  modularity of each network was transformed to a  $z$  score  $Q_z$  by subtracting the average modularity of 100 corresponding random graphs  $\mu(Q_k^{\text{rnd}})$  and dividing the result by the corresponding standard deviation  $\sigma(Q_k^{\text{rnd}})$ . Thus,  $Q_z > 2$  would suggest a significantly higher modularity of a group-averaged network than expected by chance. However, it has been shown that some highly modular networks fail to exhibit a significant  $Q_z$ , whereas the Newman modularity of other computer-generated graphs (Girvan & Newman, 2002), which lack any community structure, is much greater than expected by chance (Karrer et al., 2008). Thus, we used robustness testing as an additional means to ensure the quality of the computed consensus partitions. Although the community layout of a truly modular network is usually stable against small perturbations of its link structure, rewiring even a few edges in a random graph generally results in substantial changes in its modular decomposition (Karrer et al., 2008; Gfeller et al., 2005). Thus, similar community layouts of original and perturbed networks are suggestive of a robust modular makeup of a graph, whereas community dynamics that resemble random networks suggest the lack of any natural underlying nodal clustering within a graph. Following Karrer et al. (2008), both empirical and simulated mean networks,  $F_0$  and  $G_0$ , respectively, were randomized with increasing rewiring probability  $\alpha$  (40 steps from 0 to 1). For each value of  $\alpha$ , perturbed graphs  $F_0(\alpha)$  and  $G_0(\alpha)$  were computed by averaging across 100 rewired networks ( $F_0^k(\alpha)$ ) and ( $G_0^k(\alpha)$ ), respectively. Random graphs  $R_0$  were constructed by averaging across 100 candidate null model networks with the same degree distribution as  $F_0$  and  $G_0$ , respectively. The constructed random graphs were similarly perturbed, where, for each value of  $\alpha$ , the network  $R_0(\alpha)$  was determined by averaging across 100 rewired random networks ( $R_0^k(\alpha)$ ). The similarity of community layouts in two graphs  $T$  and  $T'$  was quantified by introducing a partition distance  $p_d$  defined as the normalized variation of information between the partitions  $P$  and  $P'$  of  $T$  and  $T'$ , respectively, that is,  $p_d(T, T') = (H(P) + H(P') - 2I(P, P')) / \log(N)$  (Meila, 2007), where  $H$  and  $I$  denote the Shannon entropy and mutual information, respectively. Thus,  $p_d(T, T') = 0$  indicates identical partitions, whereas  $p_d(T, T') = 1$  points to maximally different community layouts. If the evolution of  $p_d(F_0, F_0(\alpha))$  and  $p_d(G_0, G_0(\alpha))$  departed

significantly from the perturbation dynamics of the corresponding null models  $p_d(R_0, R_0(\alpha))$ , it would thus suggest a clear nonrandom community architecture of the networks  $F_0$  and  $G_0$ . Because of its high computational demand, a robustness analysis was only performed for the group-averaged networks  $F_0$  and  $G_0$ , which were the a priori focus of analysis in this study.

### Statistical Evaluation and Computational Environment

To assess whether deviations in the partitions of the group-averaged networks from the modular decompositions of the underlying cohort were related to the observed condition (i.e., rest, syllable, and speech) or were identified by chance only, we performed a linear mixed effect model analysis (Bates, Machler, Bolker, & Walker, 2015). Three linear models with condition as fixed effect and an intercept for subjects or leave-one-out averages as random effect were fit to the data given by the partition distances  $p_d(F_0, F_k)$ ,  $p_d(G_0, G_k)$ , and  $p_d(G_0, G_{0:k})$ , respectively. We used likelihood ratio tests at a Bonferroni-corrected  $p \leq .05$  to assess whether the full models with the observed condition (rest, syllable, and speech) as the main effect differed significantly from a null model without the effect. The constructed models were subjected to simultaneous two-sided all-to-all general linear hypothesis tests in a Tukey contrast at  $p \leq .05$  adjusted for multiple comparisons to determine the statistical significance of differences in reference-to-subject partition distances across conditions.

Network processing and visualization codes were written in Python using the open-source packages NumPy, SciPy (van der Walt, Colbert, & Varoquaux, 2011), and Matplotlib (Hunter, 2007). Community analyses were performed in MATLAB (The MathWorks, Inc., Natick, MA) using the Brian Connectivity Toolbox (Rubinov & Sporns, 2010). The statistical analysis was done in R (R-Core-Team, 2016) using the packages lme4 (Bates et al., 2015) and multcomp (Hothorn, Bretz, & Westfall, 2008). Three-dimensional renderings of the networks embedded in reference brain models were generated using the BrainNet Viewer (Xia, Wang, & He, 2013). Network processing and visualization codes used in this study are made available at [research.mssm.edu/simonyanlab/analytical-tools/](http://research.mssm.edu/simonyanlab/analytical-tools/).

### Alternative Strategies for Network Construction

To ensure that our results were not biased by differences in the sampling rate between resting state and task production, we sparse-sampled the original resting-state time series to match the sparse-sampling design of the task production scans. We applied the same community analysis pipeline to the networks based on this new set of sparse-sampled resting-state data and compared our results with the findings reported in the main study. We

found that all networks showed identical within-group partition sensitivity (empirical networks:  $p_d(F_0, F_k) = 0.35 \pm 0.03$ , artificial networks:  $p_d(G_0, G_k) = 0.36 \pm 0.03$ , leave-one-out averages:  $p_d(G_0, G_{0:k}) = 0.01 \pm 0.01$ ). We subsequently constructed linear mixed effect models using partition data from sparse-sampled and original resting-state networks with data source (sparse sampled vs. original) as fixed effect and an intercept for subjects/leave-one-out averages as random effects. General linear hypothesis tests showed no statistically significant differences between sparse-sampled and original resting-state partitions (empirical networks:  $p = .7$ , artificial networks:  $p = .6$ , leave-one-out averages:  $p = .9$ ). This additional analysis demonstrated that the sparse-sampling sequence during task production did not differentially influence our data in terms of mapping network communities.

To illustrate that the validity of our findings did not depend on our choice of the NMI as the statistical dependence metric, we performed a second set of numerical experiments, in which we based the constructed networks on pairwise zero-lag Pearson correlation coefficients. We first subjected the Pearson networks to the same relative thresholding strategy as the NMI graphs by iteratively eliminating edges with low weights while keeping all networks connected (i.e., all nodes must have degree  $\geq 1$  after thresholding). This reduced the densities of empirical networks to 65% for resting state, 85% for syllable production, and 72% for speech production and similarly for artificial networks to 67% for resting state, 86% for syllable production, and 74% for speech production. To further lower network densities to the levels of the NMI graphs, we used a maximum spanning tree approach. For each graph, we computed a maximum spanning tree and populated it with the strongest connections found in the original networks until a target density was met (Hidalgo, Klinger, Barabasi, & Hausmann, 2007). To keep the results consistent, we used the densities of the NMI networks as target values. Because of the presence of negative weights in the graphs, we used an adapted implementation of the Louvain algorithm to perform the community detection analysis (Traag & Bruggeman, 2009).

Compared with the findings reported below for the NMI graphs, the results obtained using these Pearson networks were highly similar (empirical graphs [rest/syllable/speech]:  $Q_z = 39.1/50.0/57.4$ ,  $p_d(F_0, F_k) = 0.33 \pm 0.03/0.37 \pm 0.05/0.37 \pm 0.03$ ; artificial graphs [rest/syllable/speech]:  $Q_z = 56.1/36.1/47.0$ ,  $p_d(G_0, G_k) = 0.37 \pm 0.03/0.37 \pm 0.05/0.39 \pm 0.03$ ;  $p_d(G_0, G_{0:k}) = 0.02 \pm 0.04/0.09 \pm 0.06/0.12 \pm 0.06$ ). Likelihood ratio tests of linear mixed effect models constructed based on empirical as well as artificial graphs confirmed that deviations of per-subject partitions from the group-averaged reference were condition related (both  $ps < .0007$ ) and not by chance, similar to our main finding in NMI networks. Within-group stability of mean network partitions showed identical characteristics as in NMI networks. Deviations of empirical per-subject network communities from the

group-averaged partition showed significant deviations between resting state and task production (both  $p$ s < .003) but not across tasks (speech vs. syllable:  $p = .9$ ). Finally, leave-one-out partitions also exhibited an increase in community sensitivity with increasing task complexity (all  $p$ s <  $10^{-5}$ ) similar to NMI networks.

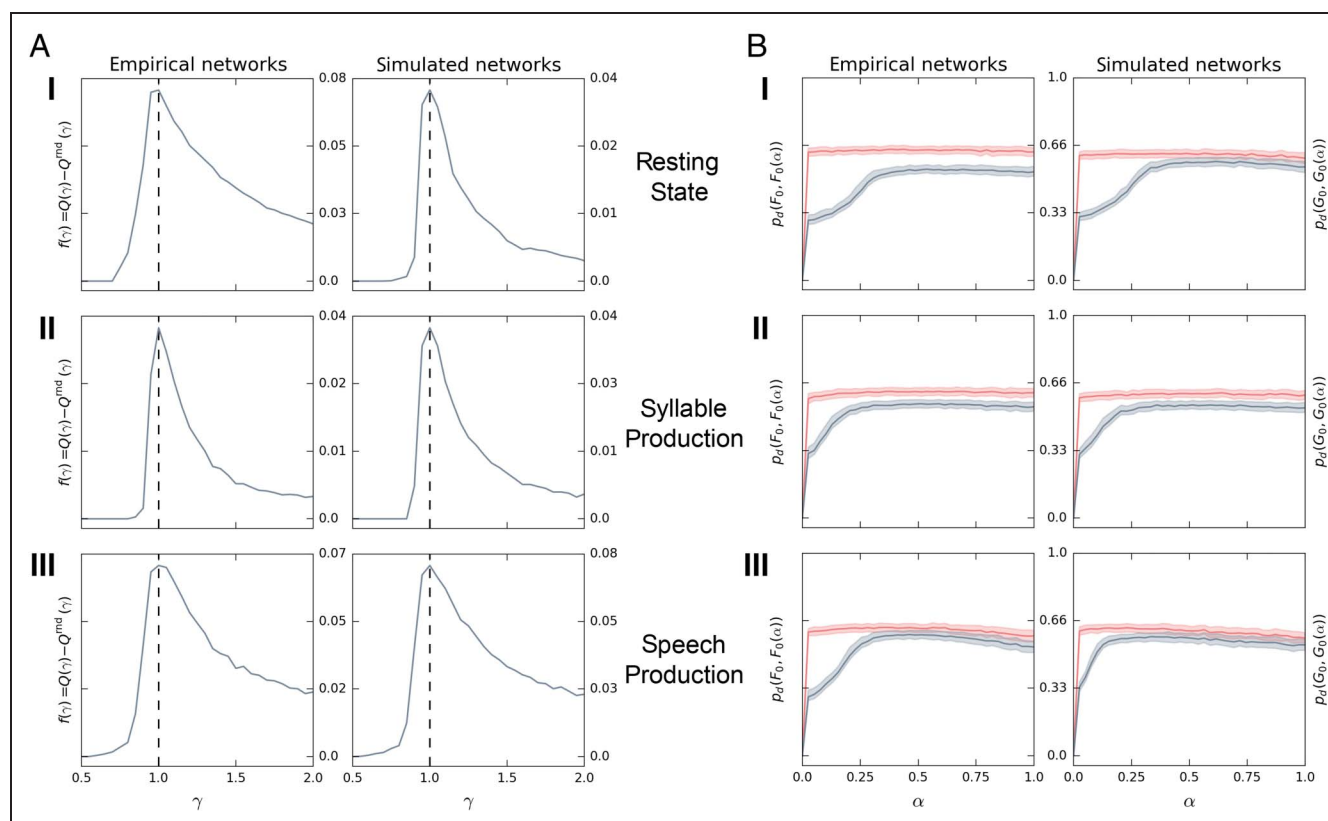
In conclusion, these data demonstrated that our main findings below are not influenced by our choice of the mutual information as the statistical dependence measure. However, using the NMI instead of Pearson correlation coefficients ensured the reliability of our results as it simultaneously accounted for nonlinear co-dependencies in the data. Moreover, the use of a maximum spanning tree approach with Pearson's networks may yield ambiguous results because a network has a unique maximum spanning tree if and only each edge weight in the graph is unique (Kepner & Gilbert, 2011). This condition may be violated by most functional networks, yielding numerous plausible maximum spanning trees per graph. Thus, although our data showed similarities between NMI and

Pearson networks, using NMI networks is overall more robust for providing reliable outcome, as described below.

## RESULTS

### Empirical Networks

As a first step in our analysis, we performed a detailed assessment of the community architecture in empirical large-scale functional networks based on pairwise NMI values between regionally averaged BOLD time series. A comparative assessment of the modularity score of the empirical group-averaged networks  $F_0$  and corresponding null model graphs yielded  $\gamma = 1$  as the resolution parameter, for which  $Q$  was maximally different from  $Q^{\text{md}}$  in all conditions (Figure 2A). Thus,  $\gamma = 1$  was used in all following modularity maximization calculations. The threshold values in the subsequent consensus partitionings were chosen to be 0.0625 for  $F_0^r$  and  $F_0^{\text{syll}}$  and 0.044



**Figure 2.** Group-specific resolution parameters were defined as maximizers of a quality function and robustness of the modular structure in group-averaged networks. (A) The resolution parameter  $\gamma$  was defined as the solution to an optimization problem posed for every condition (rest, syllable, and speech: I–III) in both empirical (left column) and simulated (right column) networks. The value of  $\gamma$  that maximized a quality function  $f(\gamma)$  reflecting the Newman modularity  $Q(\gamma)$  of the respective group-averaged network with respect to the modularity score of an associated random graph  $Q^{\text{md}}(\gamma)$  was used in all subsequent modular decomposition calculations in the corresponding group. Thus, condition-specific resolution parameters were defined as  $\text{argmax } f(\gamma)$  (vertical dashed lines). (B) The quality of the computed reference partitions of the group-averaged networks was assessed using perturbation testing. Randomizing empirical (left column) and simulated (right column) mean networks,  $F_0$  and  $G_0$ , respectively, with rewiring probability  $\alpha$  yielded perturbed graphs  $F_0(\alpha)$  and  $G_0(\alpha)$ , respectively. Shown is the mean partition distance (averaged across 100 rewired networks for each value of  $\alpha$  with shaded areas indicating standard deviations from the mean) between actual and perturbed original state (I), syllable (II), and speech production (III) networks (blue) in comparison with the average partition distance between the associated original and randomized null model graphs  $p_d(R_0, R_0(\alpha))$  (red).

for  $F_0^{\text{SP}}$ , which were based on the respective maxima in the corresponding null association matrices.

A likelihood ratio test of the constructed linear mixed effect model against an associated null model confirmed that deviations of per-subject partitions from the group-averaged reference were indeed condition related ( $p = 3.2 \times 10^{-5}$ ) and not by chance.

An in-depth assessment of the modularity score  $F_0$  and the results of robustness testing confirmed the highly modular structure of all empirical mean networks (all  $Q_z \geq 32.8$ ; see Table 1). Compared with equivalent random graphs, the perturbation curves  $p_d(F_0, F_0(\alpha))$  of all empirical functional networks increased markedly slower and deviated significantly from the corresponding null models, indicating that the detected communities were relatively robust against perturbations, unlike modules detected in the respective random graphs (Figure 2B).

The resting-state network  $F_0^r$  was composed of four modules, including frontoparietal (red), thalamic–BG (green), central–temporal (purple), and occipital–cerebellar (orange; Figure 3A-I). Individual per-subject networks  $(F_k^r)_{k \geq 1}$  showed a similar modular makeup that closely resembled the group-averaged partition ( $p_d(F_0^r, F_k^r) = 0.35 \pm 0.03$ ; see Table 1).

Compared with the resting state, syllable production was characterized by significantly larger deviations of per-subject partitions from the group-averaged reference ( $p_d(F_0^{\text{sy1}}, F_k^{\text{sy1}}) = 0.39 \pm 0.03, p = .002$ ). The community architecture of the mean network  $F_0^{\text{sy1}}$  was composed of four modules, including mainly left-lateralized (red), mainly right-lateralized (purple), thalamic–BG (green), and cerebellar (orange; Figure 3B-I).

Similar to syllable networks, speech production was characterized by pronounced changes in the layout of per-subject communities as compared with the modular structure of the mean network ( $p_d(F_0^{\text{SP}}, F_k^{\text{SP}}) = 0.4 \pm 0.04$ ; see Table 1). These deviations were significantly larger than the corresponding resting-state variations ( $p < 10^{-4}$ ) but comparable with the observed modular inter-

subject variability during syllable production ( $p = .93$ ). The mean network  $F_0^{\text{SP}}$  consisted of five communities, including two mainly left- and right-lateralized central–temporal (red and purple, respectively), thalamic–BG (green), fronto-occipital (orange), and cerebellar (blue; Figure 3C-I).

The low standard deviations of  $p_d(F_0, F_k)$  yielded very small maximal standard errors of the average subject-to-reference partition distances across all conditions ( $SE_r = 0.007, SE_{\text{sy1}} = 0.007, SE_{\text{SP}} = 0.009$ ), indicating that the chosen sample size of 100 simulated graphs  $(G_k)_{k \geq 1}$  was adequate to perform a modular stability analysis.

## Simulated Networks

On the basis of the empirical per-subject networks, 100 artificial graphs were constructed per condition, namely, resting state, meaningless syllable production, and meaningful speaking. Analogous to the analysis performed for the empirical networks, the mean networks  $G_0$  were used to algorithmically determine condition-specific resolution parameters and consensus partitioning thresholds. As for the empirical networks,  $\gamma = 1$  maximized the difference between  $Q$  and  $Q^{\text{rnd}}$  in all groups (Figure 2A), whereas the consensus thresholds were determined to be 0.046, 0.044, and 0.0625 for  $G_0^r, G_0^{\text{sy1}}$ , and  $G_0^{\text{SP}}$ , respectively. Deviations in the partitions of the individual 100 simulated networks from the corresponding group-averaged references were related to the analyzed conditions (rest, syllable, and speech) as revealed by a likelihood ratio test of the associated linear mixed effect model against a corresponding null model ( $p < 10^{-6}$ ), suggesting that the basic topological properties of the empirical networks were preserved in the simulated graphs. This observation was further confirmed when assessing the quality of the modular structure in the group-averaged networks. All graphs  $G_0$  showed a normalized modularity score significantly greater than two ( $Q_z \geq 27.1$ ; see Table 1) as well as smoothly inclining perturbation curves  $p_d(G_0, G_0(\alpha))$  that differed markedly from corresponding random networks (Figure 2B).

The modular decomposition of the simulated resting-state network  $G_0^r$  closely resembled the community architecture of the corresponding empirical graph ( $p_d(F_0^r, G_0^r) = 0.14$ ); however, an additional fifth module (blue) was detected in  $G_0^r$  (Figure 3A-III). Notably, this new module recruited nodes exclusively from the thalamic–BG module in the empirical network (green in Figure 3A-I) and thus was composed of mostly subcortical nodes, including the subdivisions of the bilateral amygdala and thalamus, nucleus accumbens, and hippocampus. Thus, the additional fifth module in  $G_0^r$  took the position of the thalamic–BG community (green) in  $F_0^r$ , whereas some remaining subcortical regions migrated from the central–temporal module (purple) in  $F_0^r$  to the thalamic–BG community (green) in  $G_0^r$  (Figure 3A-II). A within-group analysis revealed that the community architecture of the generated 100 simulated

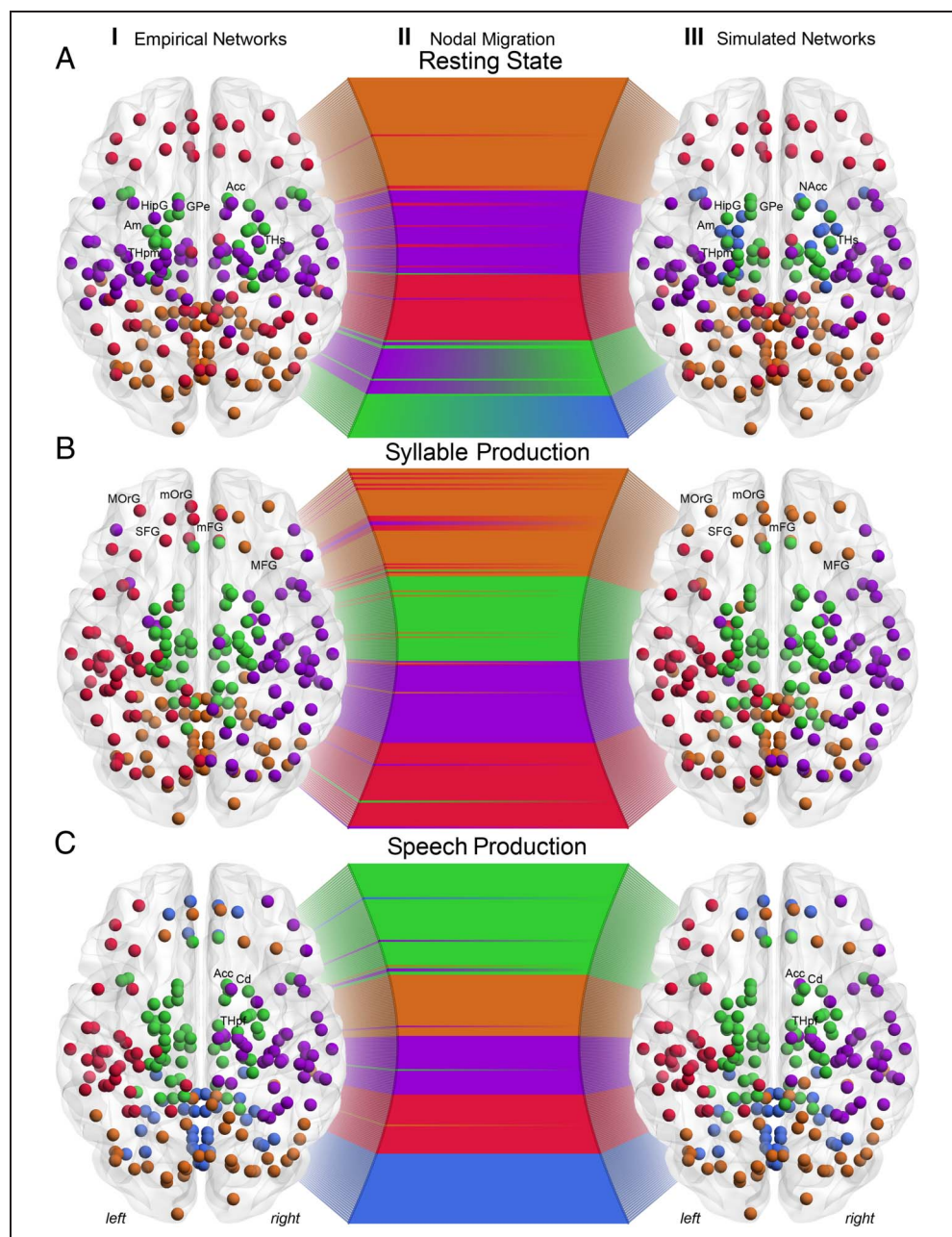
**Table 1.** Summary Statistics of Modular Decompositions

|                     | <i>Rest</i>     | <i>Syllable</i> | <i>Speech</i>   |
|---------------------|-----------------|-----------------|-----------------|
| Empirical networks  |                 |                 |                 |
| $Q_z$               | 53.5            | 32.8            | 47.2            |
| $p_d(F_0, F_k)$     | $0.35 \pm 0.03$ | $0.39 \pm 0.03$ | $0.4 \pm 0.04$  |
| Artificial networks |                 |                 |                 |
| $Q_z$               | 59.0            | 27.1            | 51.5            |
| $p_d(G_0, G_k)$     | $0.36 \pm 0.03$ | $0.4 \pm 0.03$  | $0.4 \pm 0.04$  |
| $p_d(G_0, G_{0,k})$ | $0.01 \pm 0.02$ | $0.05 \pm 0.04$ | $0.07 \pm 0.04$ |

Calculated  $Q$  modularity  $z$  scores and within-group partition distances between per-subject networks/leave-one-out samples and the respective group-averaged reference graphs in empirical and artificial network cohorts across the resting state, syllable production, and speaking.



**Figure 3.** Spatial distribution of communities in the group-averaged empirical and simulated networks. The modular structure of empirical (I) and simulated (III) group-averaged networks at rest (A), while uttering syllables (B) and during speech production (C), is shown on 3-D brain renderings in axial orientation with nodal colors illustrating regional community affiliations. Changes in module membership between empirical ( $F_0$ ) and simulated ( $G_0$ ) networks are visualized by migration flow diagrams (II). The nodes of  $F_0$  and  $G_0$  are arranged on the left and right arches, respectively, color coded by module membership. Nodal migration between communities in the empirical and simulated networks is illustrated by color gradients between the opposing arches. Acc = nucleus accumbens; Am = amygdala; Cd = caudate; mFG/MFG/SFG = medial/middle/superior frontal gyrus; GPe = external segment of globus pallidus; HipG = hippocampal gyrus; MORg/mOrG = medial/middle orbital gyrus; THpf/THpm/THs = prefrontal/premotor/somatosensory part of the thalamus.



graphs was similar to the respective group-averaged partition ( $p_d(G_0^r, G_k^r) = 0.36 \pm 0.03$ ; Table 1), consistent with the findings for the empirical networks.

The community architecture of the simulated syllable production network  $G_0^{syl}$  closely approximated the partition of the empirical network ( $p_d(F_0^{syl}, G_0^{syl}) = 0.15$ ), showing four modules of almost identical structure as their respective empirical counterparts (Figure 3B-III). Interestingly, the most pronounced community membership fluctuations were found in the frontal lobe with nine nodes (subdivisions of the bilateral orbital and frontal gyri) migrating from the left-lateralized module in the empirical graph  $F_0^{syl}$  (red) to the cerebellar module in the simulated network  $G_0^{syl}$  (orange; Figure 3B-II). In accordance with

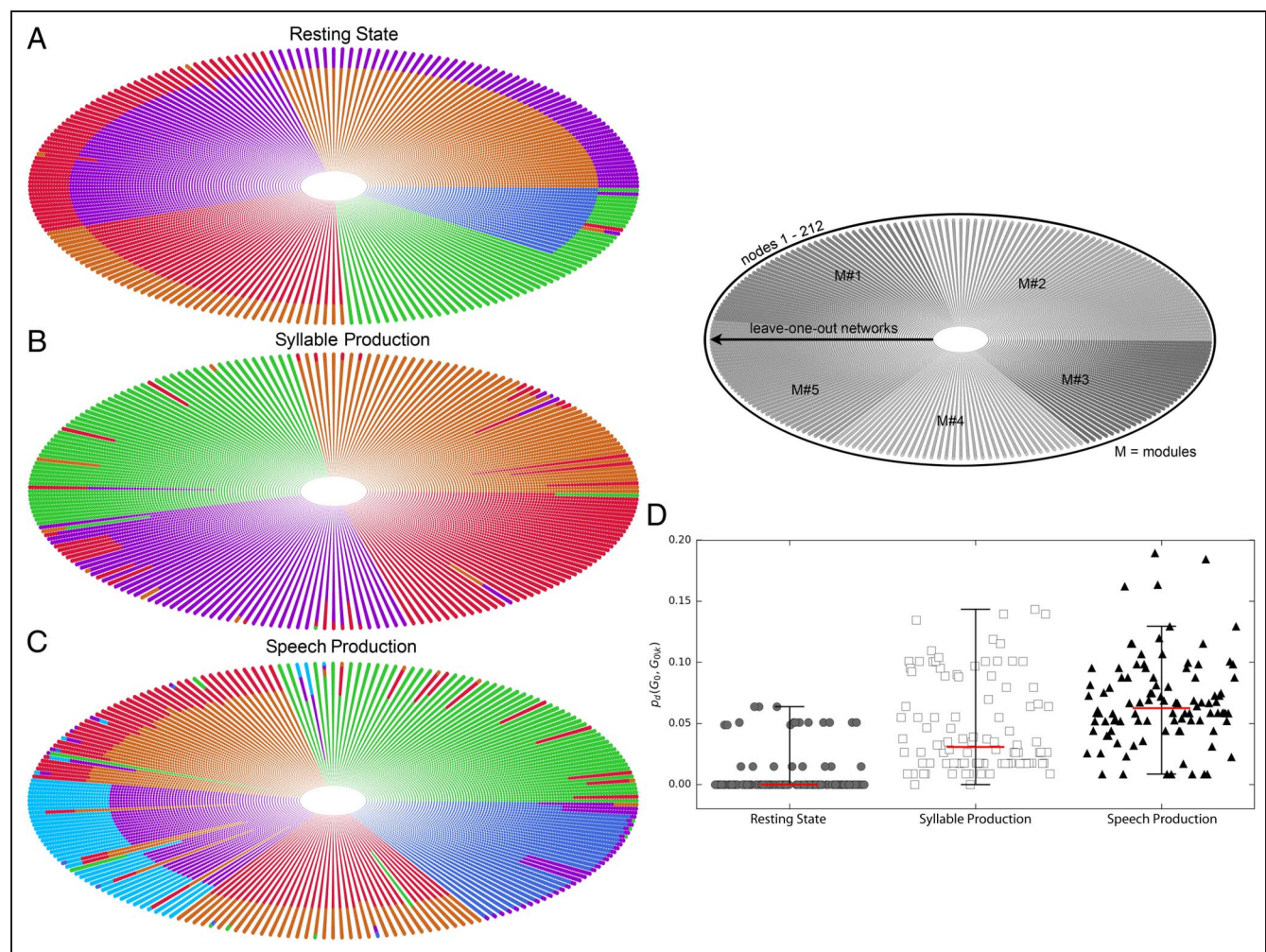
the results obtained for the empirical networks, the constructed 100 syllable production graphs showed significantly larger deviations from the respective group-averaged reference than the simulated resting-state networks ( $p_d(G_0^{syl}, G_k^{syl}) = 0.4 \pm 0.03$ ,  $p < 10^{-4}$ ).

The within-group analysis of simulated speech production graphs revealed a similar pattern with deviations of individual network partitions from the group-averaged reference, which were significantly more pronounced than for the simulated resting-state graphs ( $p_d(G_0^{sp}, G_k^{sp}) = 0.4 \pm 0.04$ ,  $p < 10^{-4}$ ) but comparable with the simulated syllable production networks ( $p = .6$ ). Remarkably, the community architecture of the empirical and simulated speech production graphs,  $F_0^{sp}$  and  $G_0^{sp}$ , respectively, was

very similar, showing nodal communities of almost equal topology ( $p_d(F_0^{SP}, G_0^{SP}) = 0.06$ ; Figure 3C-III). Both networks were composed of five modules, which differed only by isolated nodes on module boundaries changing their affiliation from  $F_0^{SP}$  to  $G_0^{SP}$  (Figure 3C-II). The thalamic–BG module in  $G_0^{SP}$  showed most deviations from its empirical counterpart by recruiting some additional subcortical regions (right nucleus accumbens, external globus pallidus). However, this module was also the largest in both networks, comprising 26% and 28% of all nodes in  $F_0^{SP}$  and  $G_0^{SP}$ , respectively.

To further assess the impact of intersubject variability on the partitions of the group-averaged graphs, we per-

formed a leave-one-out stability analysis using the 100 simulated networks constructed for each condition. Group-averaged resting-state network partitions were characterized by the highest robustness against perturbations within the sample cohort ( $p_d(G_0^r, G_{0/k}^r) = 0.01 \pm 0.02$ ; Figure 4D and E, Table 1) and exhibited a relatively stable community architecture across all leave-one-out networks (Figure 4A). The mean syllable production network was significantly more sensitive to variations in the underlying cohort than the resting-state graph ( $p_d(G_0^{sy1}, G_{0/k}^{sy1}) = 0.05 \pm 0.04$ ,  $p < 10^{-6}$ ; Figure 4D), with nodal community affiliations fluctuating notably between leave-one-out graphs (Figure 4B). Compared with the resting state and syllable



**Figure 4.** Sensitivity of community layouts with respect to the group-averaged reference partitions. Visualization of communities in the group-averaged networks  $G_0$  (innermost ellipse) and in the constructed 100 leave-one-out graphs ( $G_{0,k}$ ) $_{k \geq 1}$  (surrounding concentric ellipses) at rest (A), during syllable production (B), and while speaking (C). Each ellipse consists of 212 circular disks, with each disk representing a network node, color coded based on community affiliation. Nodes are ordered following the group-averaged reference partition to make the display consistent across ellipses, such that each individual node always occupies the same relative position in all ellipses, from center to boundary. (D) Perturbations in community structure across leave-one-out graphs ( $G_{0,k}$ ) $_{k \geq 1}$  with respect to the reference partition of the mean networks  $G_0$  across the resting state (gray circles), syllable production (white squares), and speaking (black triangles) to the respective mean network community architecture quantified by the partition distance  $p_d$ . Horizontal red lines indicate median values across the respective cohort. The range of partition distance values within each group is illustrated by whiskers extending from medians to upper and lower quartiles plus 1.5 times the respective interquartile range. General linear hypothesis testing of the constructed linear mixed effect models in an all-to-all Tukey contrast at  $p \leq .05$  yielded significant differences in within-group perturbation sensitivity between rest and task as well as between tasks (all  $p_s \leq 7.1 \times 10^{-6}$ ).

production, communities of the group-averaged speech production network  $G_0^{\text{SP}}$  showed the highest variability across the testing set ( $p_d(G_0^{\text{SP}}, G_{0\setminus k}^{\text{SP}}) = 0.07 \pm 0.04$ ; Figure 4D). The modular makeup of  $(G_{0\setminus k}^{\text{SP}})_{k \geq 1}$  changed drastically from sample to sample (Figure 4C), resulting in significantly higher deviations of leave-one-out network partitions from the group-averaged reference as compared with rest and syllable production (all  $p$ s  $\leq 7.1 \times 10^{-6}$ ).

## DISCUSSION

We performed a stability analysis of nodal communities in large-scale functional brain networks during three conditions of increasing complexity: the resting state, production of meaningless syllables, and production of meaningful speech. Our findings indicate that task-related intersubject variability impacts mean network community structure and its aptitude to represent the modular makeup of the underlying condition. Pronounced interindividual differences in functional connectivity as usually observed during task conditions (Hermundstad et al., 2013) are potentially confounding within-group stability of task-based functional modules.

The calibration of the resolution parameter used in the Louvain algorithm resulted in  $\gamma = 1$  for all group-averaged empirical networks  $F_0$ , which implies that the scale of functional modules was balanced across all examined conditions. The same result was also obtained for the simulated networks  $G_0$ , demonstrating that the resolution of empirical communities was preserved in the simulated graphs. These findings indicate that the determined differences in community stability within empirical  $(F_k)_{k \geq 1}$ , simulated  $(G_k)_{k \geq 1}$ , and leave-one-out  $(G_{0\setminus k})_{k \geq 1}$  networks were a function of the assessed condition and likely not caused by variations in the resolution of the modularity function.

Empirical networks  $F_0$  were characterized by a highly modular structure as shown by robustness testing (Figure 2B), which confirmed the nonrandom nature of the computed partitions. The follow-up assessment of the modular architecture in the simulated network cohort  $(G_k)_{k \geq 1}$  further confirmed that the employed strategy for generating these graphs indeed preserved the topological characteristics seen in the empirical networks. Imposing Gaussian random graphs on the per-subject networks  $(F_k)_{k \geq 1}$  to construct  $(G_k)_{k \geq 1}$  introduced perturbations in the group-averaged graphs  $G_0$ , which in turn led to changes in their respective modular architecture as compared with the empirical networks  $F_0$ . Thus, assessing the migration of nodes from communities in  $F_0$  to modules in the simulated group-averaged networks  $G_0$  (Figure 3-II) represents a module perturbation analysis by itself. The subsequently performed leave-one-out analysis of the generated networks further quantified the task-dependent sensitivity of functional communities.

## Topological Characteristics of Resting-State Networks

Resting-state communities in  $F_0^r$  showed a spatial distribution that was balanced across hemispheres with no evident lateralization pattern (Figure 3A-I). Compared with both tasks, resting-state community showed more prominent differences between empirical  $F_0^r$  and simulated  $G_0^r$  networks (Figure 3A), with an additional fifth module formed in  $G_0^r$  compared with the four modules in  $F_0^r$ . The fifth module in  $G_0^r$  (blue; Figure 3A-III) was mainly composed of subcortical nodes migrating the central-temporal module of  $F_0^r$  (purple). The layout of resting-state modules with respect to communities in  $G_0^r$  showed the highest similarity across leave-one-out averages  $(G_{0\setminus k}^r)_{k \geq 1}$  compared with the two analyzed tasks (Figure 4D and E), indicating that subject-specific contributions in the construction of the group-averaged network  $G_0^r$  did not noticeably influence its community structure. This finding motivates the hypothesis that the modular decomposition of a group-averaged functional resting-state network is well suited to approximate the community layouts in the underlying network cohort. Clearly, this proposition premises a sufficient number of graphs to construct the average network. However, the empirical networks analyzed here showed a very small maximal standard error of average subject-to-reference partition distances across all conditions (maximal  $SE = 0.009$ ), suggesting that the validity of the stated conjecture does not require an unrealistically large sample size.

## Topological Characteristics of Task-Production Networks

The architecture of empirical speech and syllable production networks was shaped by two communities with clearly delineated hemispheric dominance (red and purple; Figure 3A-I and B-I), which in total comprised 51% ( $F_0^{\text{syll}}$ ) and 37% ( $F_0^{\text{SP}}$ ) of all nodes in the respective graphs. Both syllable and speech production graphs were characterized by a markedly lower similarity of per-subject network partitions to the group-averaged reference as compared with the resting state. In other words, functional communities exhibited higher within-group variability while performing the two tasks than during the resting state. Given that network partitions (in every  $F_0$  as well as in all  $F_k$ ) were calculated in terms of consensus communities, the observed increase in topological variability of modules was thus most likely a consequence of individual task-related functional adaptations.

Interestingly, although the simulated and empirical networks were very similar across all conditions ( $p_d(F_0, G_0) \leq 0.15$ ), confirming the robustness of the detected communities with respect to perturbations of edges (Figure 2B), the simulated speech production network  $G_0^{\text{SP}}$  exhibited the smallest deviation from the corresponding empirical graph  $F_0^{\text{SP}}$  (Figure 3C). On the other hand, similar to the

associated empirical networks, simulated speech production communities were characterized by large within-group variations with respect to the reference partition of the mean network and showed the highest sensitivity to leave-one-out sampling (Figure 4). This observation may be indicative of two effects. First, within-group variations present in the empirical cohort  $(F_k^{SP})_{k \geq 1}$  were conserved by the randomization process that generated the simulated graphs  $(G_k^{SP})_{k \geq 1}$ . Second, the perturbations introduced in  $(F_k^{SP})_{k \geq 1}$  to construct  $(G_k^{SP})_{k \geq 1}$  had a negligible effect on the layout of functional communities in the mean network  $G_0^{SP}$  and were much less pronounced than edge-weight deviations because of intersubject variability. Thus, community structure translated well from empirical  $F_0^{SP}$  to simulated group-averaged  $G_0^{SP}$  despite pronounced within-group variability in both cohorts  $(F_k^{SP})_{k \geq 1}$  and  $(G_k^{SP})_{k \geq 1}$ .

By comparison, the community structure of the simulated syllable production network  $G_0^{SYL}$  deviated noticeably from the empirical graph  $F_0^{SYL}$ . Particularly, regions in the pFC (orange; Figure 3B), which have been shown to play a key role in the cognitive aspects of speech control (Burton, Locasto, Krebs-Noble, & Gullapalli, 2005; Giraud et al., 2004; Benson et al., 2001; Muller et al., 1997) and which are of less importance for the processing of meaningless syllables (Fuertinger et al., 2015), exhibited differing community assignments in  $G_0^{SYL}$  as compared with  $F_0^{SYL}$ . Thus, like in the resting state, modular deviations in the simulated graph  $G_0^{SP}$  from the community layout in  $F_0^{SP}$  were not evenly distributed across the entire network but predominantly affected specific, weakly integrated functional components of the graphs.

Whereas simulated resting-state networks  $(G_k^r)_{k \geq 1}$  closely followed the community architecture of the respective group-averaged reference graph  $G_0^r$ , the partitions of syllable  $(G_k^{SYL})_{k \geq 1}$  and speech production  $(G_k^{SP})_{k \geq 1}$  networks showed significantly higher within-group variability with respect to the corresponding mean graphs. Single-subject changes in the computation of the analyzed task-related  $G_0^{SYL}$  and  $G_0^{SP}$  networks altered their modular topology significantly more as compared with the resting state (both  $ps < 10^{-5}$ ), with communities in the speech production network  $G_0^{SP}$  showing the highest sensitivity to perturbations in cohort averaging, as revealed by leave-one-out sampling (Figure 4D and E).

Although recent studies reported that resting-state functional architecture was related to anatomical connectivity patterns, with functional modules corresponding to dense anatomical connections (Shih et al., 2015; Shen et al., 2012; Wang, Sporns, & Burkhalter, 2012), task production induces a behaviorally driven reconfiguration of the functional architecture of the brain on a global scale (Fuertinger et al., 2015, 2016; Stanley, Dagenbach, Lyday, Burdette, & Laurienti, 2014) and may therefore be more sensitive to intersubject variability. Thus, the observed high within-group stability of resting-state functional partitions as compared with task-related modular layouts might be a

consequence of the underlying neuroanatomical architecture. An interesting question for future studies might be what the functional architecture of the brain may look like during human multitasking, that is, in participants performing more than one task at the same time.

## Conclusions

We analyzed the sensitivity of community architecture in large-scale functional brain networks with respect to task complexity. A randomization strategy was used to simulate artificial subjects that conserved the modular layout of the original empirical graphs while simultaneously introducing enough randomness to ensure sufficient variability across networks for studying topological stability. We demonstrated that resting-state functional network partitions were characterized by low intersubject variability with respect to a group-averaged reference, whereas speech and syllable production graph architectures exhibited significantly larger within-group deviations from the community layout of the associated mean networks. Thus, the sensitivity of nodal modular topology with respect to group averaging increased from rest to syllable production to speaking, which suggests that the approximation quality of the community structure in the average network to reflect individual per-subject partitions depends on task complexity.

## Acknowledgments

This work was supported by the grants from the National Institute on Deafness and Other Communication Disorders (R01DC012545) and the National Institute of Neurological Disorders and Stroke (R01NS088160) to K. S.

Reprint requests should be sent to Kristina Simonyan, Department of Neurology, One Gustave L. Levy Place, Box 1137, Icahn School of Medicine at Mount Sinai, New York, NY 10029, or via e-mail: kristina.simonyan@mssm.edu.

## REFERENCES

- Andric, M., & Hasson, U. (2015). Global features of functional brain networks change with contextual disorder. *Neuroimage*, *117*, 103–113.
- Bassett, D. S., & Bullmore, E. (2006). Small-world brain networks. *Neuroscientist*, *12*, 512–523.
- Bassett, D. S., Porter, M. A., Wymbs, N. F., Grafton, S. T., Carlson, J. M., & Mucha, P. J. (2013). Robust detection of dynamic community structure in networks. *Chaos*, *23*, 013142.
- Bassett, D. S., Wymbs, N. F., Porter, M. A., Mucha, P. J., Carlson, J. M., & Grafton, S. T. (2011). Dynamic reconfiguration of human brain networks during learning. *Proceedings of the National Academy of Sciences, U.S.A.*, *108*, 7641–7646.
- Bates, D. M., Machler, M., Bolker, B. M., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.
- Benson, R. R., Whalen, D. H., Richardson, M., Swainson, B., Clark, V. P., Lai, S., et al. (2001). Parametrically dissociating speech and nonspeech perception in the brain using fMRI. *Brain and Language*, *78*, 364–396.

- Bertolero, M. A., Yeo, B. T., & D'Esposito, M. (2015). The modular and integrative functional architecture of the human brain. *Proceedings of the National Academy of Sciences, U.S.A.*, *112*, E6798–E6807.
- Betz, R. F., Misisic, B., He, Y., Rumschlag, J., Zuo, X.-N., & Sporns, O. (2015). Functional brain modules reconfigure at multiple scales across the human lifespan. arXiv preprint:1510.08045.
- Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*, P10008.
- Bondy, J. A., & Murty, U. S. R. (1976). *Graph theory with applications*. New York: North Holland.
- Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, *10*, 186–198.
- Burton, M. W., Locasto, P. C., Krebs-Noble, D., & Gullapalli, R. P. (2005). A systematic investigation of the functional neuroanatomy of auditory and visual phonological processing. *Neuroimage*, *26*, 647–661.
- Chang, Y. T., Leahy, R. M., & Pantazis, D. (2012). Modularity-based graph partitioning using conditional expected models. *Physical Review E*, *85*, 016109.
- Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York: Wiley.
- Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*, 162–173.
- De Vico Fallani, F., Richiardi, J., Chavez, M., & Achard, S. (2014). Graph analysis of functional brain networks: Practical issues in translational neuroscience. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, *369*.
- Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage*, *25*, 1325–1335.
- Fakcharoenphol, J., & Rao, S. (2006). Planar graphs, negative weight edges, shortest paths, and near linear time. *Journal of Computer and System Sciences*, *72*, 868–889.
- Fortunato, S., & Barthelemy, M. (2007). Resolution limit in community detection. *Proceedings of the National Academy of Sciences, U.S.A.*, *104*, 36–41.
- Fuertinger, S., Horwitz, B., & Simonyan, K. (2015). The functional connectome of speech control. *PLoS Biology*, *13*, e1002209.
- Fuertinger, S., Simonyan, K., Sperling, M. R., Sharan, A. D., & Hamzei-Sichani, F. (2016). High frequency brain networks undergo modular breakdown during epileptic seizures. *Epilepsia*, *57*, 1097–1108.
- Gfeller, D., Chappelier, J. C., & De Los Rios, P. (2005). Finding instabilities in the community structure of complex networks. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, *72*, 056135.
- Ghosh, A., Rho, Y., McIntosh, A. R., Kotter, R., & Jirsa, V. K. (2008). Noise during rest enables the exploration of the brain's dynamic repertoire. *PLoS Computational Biology*, *4*, e1000196.
- Ginestet, C. E., Fournel, A. P., & Simmons, A. (2014). Statistical network analysis for functional MRI: Summary networks and group comparisons. *Frontiers in Computational Neuroscience*, *8*, 51.
- Giraud, A. L., Kell, C., Thierfelder, C., Sterzer, P., Russ, M. O., Preibisch, C., et al. (2004). Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cerebral Cortex*, *14*, 247–255.
- Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences, U.S.A.*, *99*, 7821–7826.
- Godwin, D., Barry, R. L., & Marois, R. (2015). Breakdown of the brain's functional network modularity with awareness. *Proceedings of the National Academy of Sciences, U.S.A.*, *112*, 3799–3804.
- Good, B. H., de Montjoye, Y. A., & Clauset, A. (2010). Performance of modularity maximization in practical contexts. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, *81*, 046106.
- Grabner, R. H., Ansari, D., Reishofer, G., Stern, E., Ebner, F., & Neuper, C. (2007). Individual differences in mathematical competence predict parietal brain activation during mental calculation. *Neuroimage*, *38*, 346–356.
- Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). New York: Springer.
- Hermundstad, A. M., Bassett, D. S., Brown, K. S., Aminoff, E. M., Clewett, D., Freeman, S., et al. (2013). Structural foundations of resting-state and task-based functional connectivity in the human brain. *Proceedings of the National Academy of Sciences, U.S.A.*, *110*, 6169–6174.
- Hidalgo, C. A., Klinger, B., Barabasi, A. L., & Hausmann, R. (2007). The product space conditions the development of nations. *Science*, *317*, 482–487.
- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*, 346–363.
- Humphries, M. D., Gurney, K., & Prescott, T. J. (2006). The brainstem reticular formation is a small-world, not scale-free, network. *Proceedings of the Biological Sciences*, *273*, 503–511.
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*, 90–95.
- Karrer, B., Levina, E., & Newman, M. E. (2008). Robustness of community structure in networks. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, *77*, 046119.
- Kepner, J. V., & Gilbert, J. R. (2011). *Graph algorithms in the language of linear algebra*. Philadelphia: Society for Industrial and Applied Mathematics.
- Lancichinetti, A., & Fortunato, S. (2012). Consensus clustering in complex networks. *Scientific Reports*, *2*, 336.
- Lynall, M. E., Bassett, D. S., Kerwin, R., McKenna, P. J., Kitzbichler, M., Muller, U., et al. (2010). Functional connectivity and brain networks in schizophrenia. *Journal of Neuroscience*, *30*, 9477–9487.
- Meila, M. (2007). Comparing clusters—An information based distance. *Journal of Multivariate Analysis*, *98*, 873–895.
- Miranda-Dominguez, O., Mills, B. D., Carpenter, S. D., Grant, K. A., Kroenke, C. D., Nigg, J. T., et al. (2014). Connectotyping: Model based fingerprinting of the functional connectome. *PLoS One*, *9*, e111048.
- Muller, R. A., Rothermel, R. D., Behen, M. E., Muzik, O., Mangner, T. J., & Chugani, H. T. (1997). Receptive and expressive language activations for sentences: A PET study. *NeuroReport*, *8*, 3767–3770.
- Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, *69*, 026113.
- Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences, U.S.A.*, *103*, 8577–8582.
- Newman, S. D., Carpenter, P. A., Varma, S., & Just, M. A. (2003). Frontal and parietal participation in problem solving in the Tower of London: fMRI and computational modeling of planning and high-level perception. *Neuropsychologia*, *41*, 1668–1682.

- Onoda, K., & Yamaguchi, S. (2013). Small-worldness and modularity of the resting-state functional brain network decrease with aging. *Neuroscience Letters*, *556*, 104–108.
- Porter, M. A., Onnela, J.-P., & Mucha, P. J. (2009). Communities in networks. *Notices of the AMS*, *36*, 1082–1166.
- R-Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage*, *52*, 1059–1069.
- Rypma, B., & D'Esposito, M. (1999). The roles of prefrontal brain regions in components of working memory: Effects of memory load and individual differences. *Proceedings of the National Academy of Sciences, U.S.A.*, *96*, 6558–6563.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 623–656.
- Shen, K., Bezgin, G., Hutchison, R. M., Gati, J. S., Menon, R. S., Everling, S., et al. (2012). Information processing architecture of functionally defined clusters in the macaque cortex. *Journal of Neuroscience*, *32*, 17465–17476.
- Shih, C. T., Sporns, O., Yuan, S. L., Su, T. S., Lin, Y. J., Chuang, C. C., et al. (2015). Connectomics-based analysis of information flow in the Drosophila brain. *Current Biology*, *25*, 1249–1258.
- Simonyan, K., & Fuertinger, S. (2015). Speech networks at rest and in action: Interactions between functional brain networks controlling speech production. *Journal of Neurophysiology*, *113*, 2967–2978.
- Simonyan, K., Herscovitch, P., & Horwitz, B. (2013). Speech-induced striatal dopamine release is left lateralized and coupled to functional striatal circuits in healthy humans: A combined PET, fMRI and DTI study. *Neuroimage*, *70*, 21–32.
- Sporns, O., & Betzel, R. F. (2015). Modular brain networks. *Annual Review of Psychology*, *67*, 613–640.
- Stanley, M. L., Dagenbach, D., Lyday, R. G., Burdette, J. H., & Laurienti, P. J. (2014). Changes in global and regional modularity associated with increasing working memory load. *Frontiers in Human Neuroscience*, *8*, 954.
- Steuer, R., Kurths, J., Daub, C. O., Weise, J., & Selbig, J. (2002). The mutual information: Detecting and evaluating dependencies between variables. *Bioinformatics*, *18(Suppl. 2)*, S231–S240.
- Stevens, A. A., Tappon, S. C., Garg, A., & Fair, D. A. (2012). Functional brain network modularity captures inter- and intra-individual variation in working memory capacity. *PLoS One*, *7*, e30468.
- Strehl, A., & Ghosh, J. (2002). Cluster ensembles—A knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, *3*, 583–617.
- Sturges, H. A. (1926). The choice of a class interval case I computations involving a single. *Journal of the American Statistical Association*, *21*, 65–66.
- Traag, V. A., & Bruggeman, J. (2009). Community detection in networks with positive and negative links. *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics*, *80*, 036115.
- van den Heuvel, M. P., & Sporns, O. (2011). Rich-club organization of the human connectome. *Journal of Neuroscience*, *31*, 15775–15786.
- van der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: A structure for efficient numerical computation. *Computing in Science & Engineering*, *13*, 22–30.
- Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., et al. (2013). The WU-Minn Human Connectome Project: An overview. *Neuroimage*, *80*, 62–79.
- van Wijk, B. C. M., Stam, C. J., & Daffertshofer, A. (2010). Comparing brain networks of different size and connectivity density using graph theory. *Plos One*, *5*, e13701.
- Wang, J., Qiu, S., Xu, Y., Liu, Z., Wen, X., Hu, X., et al. (2014). Graph theoretical analysis reveals disrupted topological properties of whole brain functional networks in temporal lobe epilepsy. *Clinical Neurophysiology*, *125*, 1744–1756.
- Wang, J. H., Zuo, X. N., Gohel, S., Milham, M. P., Biswal, B. B., & He, Y. (2011). Graph theoretical analysis of functional brain networks: Test-retest evaluation on short- and long-term resting-state functional MRI data. *PLoS One*, *6*, e21976.
- Wang, Q., Sporns, O., & Burkhalter, A. (2012). Network analysis of corticocortical connections reveals ventral and dorsal processing streams in mouse visual cortex. *Journal of Neuroscience*, *32*, 4386–4399.
- Wu, F., & Huberman, B. A. (2004). Finding communities in linear time: A physics approach. *European Physical Journal B*, *38*, 331–338.
- Xia, M. R., Wang, J. H., & He, Y. (2013). BrainNet viewer: A network visualization tool for human brain connectomics. *Plos One*, *8*, e68910.
- Zachary, W. W. (1977). Information-flow model for conflict and fission in small-groups. *Journal of Anthropological Research*, *33*, 452–473.