Published in final edited form as:

J Cogn Neurosci. 2016 December ; 28(12): 2030–2043. doi:10.1162/jocn_a_01026.

Stability of Network Communities as a Function of Task Complexity

Stefan Fuertinger¹ and Kristina Simonyan^{1,2,*}

¹Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, NY, USA

²Department of Otolaryngology, Icahn School of Medicine at Mount Sinai, New York, NY, USA

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 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 inter-subject 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 subjects 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. Based on 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-subject 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). While 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 functional magnetic resonance imaging (fMRI) studies started utilizing abstract mathematical concepts rooted in graph theory (Bondy & Murty, 1976) to represent

* Corresponding author: Kristina Simonyan, MD, PhD, Department of Neurology, One Gustave L. Levy Place, Box 1137, Icahn School of Medicine at Mount Sinai, New York, NY 10029, Tel.: (212) 241-0656, Fax: (646) 537-8628, kristina.simonyan@mssm.edu.

and subsequently analyze brain activity data in terms of large-scale graphs (Bassett & Bullmore, 2006; Humphries, Gurney, & Prescott, 2006; Lynall et al., 2010; van den Heuvel & Sporns, 2011; J. Wang et al., 2014). 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 (Bullmore & Sporns, 2009; De Vico Fallani, Richiardi, Chavez, & Achard, 2014). The analysis of the community architecture within these graphs revealed a number of previously unknown brain activity patterns, such as the context-dependent global re-organization 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), estimation of cognitive performance based on the modular organization of a cognitive control network at rest (Stevens, Tappon, Garg, & Fair, 2012), and 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 presents 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 (M. E. Newman & Girvan, 2004; Porter, 2009). Maximizing a network's Newman modularity score is a computationally efficient way to simultaneously determine both the number of nodal clusters as well as their members (M. E. J. Newman, 2006). However, modularity maximization also suffers from two well-known conceptual problems: the 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, 2015; Godwin, Barry, & Marois, 2015; Onoda & Yamaguchi, 2013). A few studies have demonstrated pronounced effects of inter-subject variability on functional coupling patterns, particularly during complex cognitive tasks (Grabner et al., 2007; Miranda-Dominguez et al., 2014; S. D. 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 subjects with respect to a group-wide reference given by the modular decomposition of the group-averaged network. 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. Further, 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 (Gfeller, Chappelier, & De Los Rios, 2005; Karrer, Levina, & Newman, 2008; 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 subjects provided written informed consent prior to study participation, which was approved by the Institutional Review Boards of the Icahn School of Medicine at Mount Sinai and National Institute of Neurological Disorders and Stroke, National Institutes of Health.

2.1 Data acquisition

Twenty-seven right-handed monolingual English-speaking healthy subjects (18 females, 9 males, age 52.2 ± 11.3 years [mean $\pm$ SD]) without any history of neurological, psychiatric, or laryngeal disorders were recruited for this study.

Brain images were acquired on a 3T GE scanner (Milwaukee, WI). Whole brain resting-state fMRI images were acquired before the task-production fMRI within the same scanning session. Subjects 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 subject reported falling asleep during the scanning session. Functional images were acquired with a gradient-weighted echo planar imaging (EPI) pulse sequence (TR = 2 s, 150 contiguous volumes, TE = 30 ms, FA = 90 degrees, 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 TR = 10.6 s comprising 3.6-s delay for stimulus presentation, 5-s delay for task production and 2-s image acquisition, TE = 30 ms, FA = 90 degrees, 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 subjects produced the same syllables /i?i-i?i/, which comprised 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 subjects produced the same ten meaningful, grammatically correct English sentences (e.g., “We are always away”, “Tom is in the army”), one per volume acquisition (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 subject one per acquisition volume. A total of 36 trials per each task (syllable or sentence production) and 24 resting fixations as a baseline were acquired over the five scanning sessions in each subject 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 (MPRAGE) sequence (TI = 450 ms, TE = 3.0 ms, FA = 10, 128 slices with 1.2 mm thickness).

Construction of Functional Networks

Functional images were pre-processed 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 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 subjects (Fig. 1-I). The whole-brain parcellation comprised $N = 212$ regions of interest (ROIs), including 142 cortical, 36 subcortical 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, 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 zero in case of statistical independence, while pair-wise mutual dependence yielded NMI coefficients close to one. Hence, NMI values were (unlike the classic mutual information) bounded from above by one, allowing for quantitative comparisons across different data sets.

Following Sturges' rule (Sturges, 1926), we used 8 bins to discretize probability distributions for computing NMI coefficients, which resulted in an estimated inflation of the mutual information due to 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 pair-wise NMI coefficients by region. Due to 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 pair-wise NMI values between ROIs serving as edge weights.

In contrast, classic Pearson correlation matrices are not necessarily non-negative, which has to be accounted for in a graph theoretical framework. Negative values may be discarded by either considering only their absolute magnitudes or by removing the corresponding edges from the graph entirely, resulting in a loss or potential corruption of the underlying correlative information. Instead, positive and negative components of a graph may be

analyzed separately, though this approach might result in unconnected sub-graphs. Alternatively, classic network metrics may be re-defined 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 statistical dependence metric to construct functional networks. In this manner, three groups of functional networks were constructed, $(F_k^r)_{k=1}$, $(F_k^{syl})_{k=1}$, and $(F_k^{sp})_{k=1}$ (with k denoting the subject number), corresponding to the resting state, syllable production, and speaking (Fig 1-II).

In a second step, we stimulated a total of 300 networks, including 100 for each resting $(G_k^r)_{k=1}$, syllable production $(G_k^{syl})_{k=1}$ and speaking $(G_k^{sp})_{k=1}$ conditions, by imposing a Gaussian random graph model on randomly sampled empirical networks from the corresponding groups (Fig. 1-II) (Ghosh, Rho, McIntosh, Kotter, & Jirsa, 2008; J. H. Wang et al., 2011). 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 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 $(F_k)_{k=1}$ and $(G_k)_{k=1}$ to refer to all (rest, syllable, speech) empirical and simulated networks, respectively. The connection density of both $(F_k)_{k=1}$ and $(G_k)_{k=1}$ 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 $\approx 99\%$, 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% (Humphries et al., 2006; Lynall et al., 2010), 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) density, 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, speech) in both empirical and simulated networks (Fig. 1-III). 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, i.e., 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 comprised the most samples per group (rest, syllable, speech). While a 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 ($(G_{0 \setminus k}^r)_{k=1}$, $(G_{0 \setminus k}^{syl})_{k=1}$, and $(G_{0 \setminus k}^{sp})_{k=1}$, corresponding to the resting state, syllable and speech production, respectively, by averaging across the graphs (G_j^r) , (G_j^{syl}) and (G_j^{sp}) , respectively, for $j = 1, \dots, 100$, $j \neq k$ (Fig. 1-II).

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 sub-networks 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 non-overlapping 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 (M. E. 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 multi-resolution approach (Bassett et al., 2013) for community detection. A resolution parameter γ was introduced that scaled the null model in the calculation of Q , so that smaller communities were detected for $\gamma > 1$, whereas $0 < \gamma < 1$ resulted in larger modules. To balance modularity resolution for each of the six groups (rest, syllable, 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 γ in $[0.5, 2.0]$ (30 values in increments of 0.05). Simultaneously, we computed a modularity score $Q^{\text{nd}}(\gamma)$ for each value of γ as average modularity across 20 random networks with the same degree distribution as the corresponding group-averaged graph. Maximizing the quality function $f(\gamma) = Q(\gamma) - Q^{\text{nd}}(\gamma)$, i.e., the difference between each network's actual 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 $\gamma = \text{argmax}_{\gamma} f(\gamma)$ for the respective mean network was used as resolution parameter for all graphs in the underlying group (Fig. 1-IV).

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 (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008; Rubinov & Sporns, 2010) 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 τ 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.

Since 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 really is 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 while the Newman modularity of other computer generated graphs (Girvan & Newman, 2002), that 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. While 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 (Gfeller et al., 2005; Karrer et al., 2008). Thus, similar community layouts of original and perturbed networks are suggestive of a robust modular makeup of a graph, while 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 α (40 steps from 0 to 1). For each value of α , 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 α , 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, i.e., $p_d(T, T') = (H(P) + H(P') - 2I(P, P')) / \log(N)$ (Meila, 2007), where H and I denote the Shannon entropy and mutual information, respectively. Thus, $p_d(T, T') = 0$ indicates identical partitions, while $p_d(T, T') = 1$ 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 non-random community architecture of the networks F_0 and G_0 . Due to 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, speech) or were identified by chance only, we performed a linear mixed-effect model analysis (Bates, 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 $p = 0.05$ to assess whether the full models with the observed condition (rest, syllable, speech) as 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 hypotheses tests in a Tukey contrast at $p = 0.05$ 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, 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 <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 to 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: $p = 0.7$; artificial networks: $p = 0.6$, leave-one-out averages: $p = 0.9$). 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 statistical dependence metric, we performed a second set of numerical experiments, in which we based the constructed networks on pair-wise 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 ≥ 1 after thresholding). This reduced the densities of empirical networks to 65% for the resting state, 85% for syllable production, and 72% for speech production and similarly for artificial networks to 67% for the 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 results consistent we used the densities of the NMI

networks as target values. Due to 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 to the findings reported below for the NMI graphs, 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): $Q_z = 56.1/36.1/47.0$; $p_d(G_0, G_k) = 0.37 \pm 0.03/0.37 \pm 0.05/0.39 \pm 0.03$; $p_d(G_0, G_{0 \setminus k}) = 0.02 \pm 0.04/0.09 \pm 0.06/0.12 \pm 0.06$]. Likelihood ratio tests of linear mixed effects 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 $p < 0.0007$) 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 $p < 0.003$) but not across tasks (speech vs. syllable: $p = 0.9$). Lastly, leave-one-out partitions also exhibited an increase in community sensitivity with increasing task complexity (all $p < 10^{-5}$) 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 statistical dependence measure. However, using the NMI instead of Pearson correlation coefficients ensured the reliability of our results as it simultaneously accounted for non-linear 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, while 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 pair-wise 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^{nd} in all conditions (Fig. 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^{sy} and 0.044 for F_0^{sp} , which was 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 to 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 (Fig. 2B).

The resting state network F_0^r comprised four modules, including fronto-parietal (red), thalamic-basal ganglia (green), central-temporal (purple) and occipital-cerebellar (orange) (Fig. 3A–I). Individual per-subject networks $(F_k^r)_{k=1}$ 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 to the resting state, syllable production was characterized by significantly larger deviations of per-subject partitions from the group-averaged reference ($p_d(F_0^{syl}, F_k^{syl}) = 0.39 \pm 0.03$, $p = 0.002$). The community architecture of the mean network F_0^{syl} comprised four modules, including mainly left-lateralized (red), mainly right-lateralized (purple), thalamic-basal ganglia (green) and cerebellar (orange) (Fig. 3B–I).

Similar to syllable networks, speech production was characterized by pronounced changes in the layout of per-subject communities as compared to the modular structure of the mean network ($p_d(F_0^{sp}, F_k^{sp}) = 0.4 \pm 0.04$, see Table 1). These deviations were significantly larger than the corresponding resting-state variations ($p < 10^{-4}$) but comparable to the observed modular inter-subject variability during syllable production ($p = 0.93$). The mean network F_0^{sp} consisted of five communities, including two mainly left- and right-lateralized central-temporal (red and purple, respectively), thalamic-basal ganglia (green), fronto-occipital (orange) and cerebellar (blue) (Fig. 3C–I).

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

Simulated Networks

Based on the empirical per-subject networks, 100 artificial graphs were constructed per condition, i.e., 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^{nd} in all groups (Fig. 2A), while the consensus thresholds were determined to be 0.046, 0.044, and 0.0625 for G_0^r , G_0^{syl} , and G_0^{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, 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 ($Q_z = 27.1$, see Table 1) as well as smoothly inclining perturbation curves $p_d(G_0, G_0(\alpha))$ that differed markedly from corresponding random networks (Fig. 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 (Fig. 3A–III). Notably, this new module recruited nodes exclusively from the thalamic-basal ganglia module in the empirical network (green in Fig. 3A–I) and thus comprised 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-basal ganglia community (green) in F_0^r , while some remaining subcortical regions migrated from the central-temporal module (purple) in F_0^r to the thalamic-basal ganglia community (green) in G_0^r (Fig. 3A–II). A within-group analysis revealed that the community architecture of the generated 100 simulated 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 (Fig. 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, Fig. 3B–II). In accordance with 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-average reference, that 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 to the simulated syllable production networks ($p = 0.6$). 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 ($p_d(F_0^{sp}, G_0^{sp}) = 0.06$, Fig. 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} (Fig. 3C–II). The thalamic-basal ganglia 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 inter-subject variability on the partitions of the group-averaged graphs, we performed 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

($p_d(G_0^r, G_{0/k}^r) = 0.01 \pm 0.02$, Fig. 4D–E, Table 1) and exhibited a relatively stable community architecture across all leave-one-out networks (Fig. 4A). The mean syllable production network was significantly more sensitive to variations in the underlying cohort than the resting-state graph ($p_d(G_0^{\text{syll}}, G_{0/k}^{\text{syll}}) = 0.05 \pm 0.04$, $p < 10^{-6}$, Fig. 4D), with nodal community affiliations fluctuating notably between leave-one-out graphs (Fig. 4B). Compared to the resting state and syllable production, communities of the group-averaged speech production network G_0^{sp} showed the highest variability across the testing set ($p_d(G_0^{\text{sp}}, G_{0/k}^{\text{sp}}) = 0.07 \pm 0.04$, Fig. 4D). The modular makeup of $(G_{0/k}^{\text{sp}})_{k=1}$ changed drastically from sample to sample (Fig. 4C), resulting in significantly higher deviations of leave-one-out network partitions from the group averaged reference as compared to rest and syllable production (all $p < 7.1 \times 10^{-6}$).

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 inter-subject variability impacts mean network community structure and its aptitude to represent the modular makeup of the underlying condition. Pronounced inter-individual 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 $(F_k)_{k=1}$, simulated $(G_k)_{k=1}$ and leave-one-out $(\tilde{G}_{0/k})_{k=1}$ 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 (Fig. 2B), which confirmed the non-random nature of the computed partitions. The follow-up assessment of the modular architecture in the simulated network cohort $(G_k)_{k=1}$ further confirmed that the employed strategy for generating these graphs indeed preserved the topological characteristics seen in the empirical the networks. Imposing Gaussian random graphs on the per-subject networks $(F_k)_{k=1}$ to construct $(G_k)_{k=1}$ introduced perturbations in the group-averaged graphs G_0 , which in turn led to changes in their respective modular architecture as compared to 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 (Fig. 3II) 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 (Fig. 3A–I). Compared to both tasks, resting state community showed more prominent differences between empirical F_0^r and simulated G_0^r networks (Fig. 3A) with an additional fifth module formed in G_0^r compared to four modules in F_0^r . The fifth module in G_0^r (blue, Fig. 3A–III) comprised mainly 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 $(G_{0,k}^r)_{k=1}$ compared to the two analyzed tasks (Fig. 4D–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, Fig. 3AB–I), which in total comprised 51% (F_0^{sy}) and 37% (F_0^{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 to the resting state. In other words, functional communities exhibited higher within-group variability while performing the two analyzed 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, while simulated and empirical networks were very similar across all conditions ($p_d(F_0, G_0) = 0.15$), confirming the robustness of the detected communities with respect to perturbations of edges (Fig. 2B), the simulated speech production network G_0^{sp} exhibited the smallest deviation from the corresponding empirical graph F_0^{sp} (Fig. 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 (Fig. 4). This observation may be indicative of two effects. First, within-group variations present in the empirical cohort $(F_k^{sp})_{k=1}$ were conserved by the randomization process that generated the simulated graphs $(G_k^{sp})_{k=1}$. Second, the perturbations introduced in $(F_k^{sp})_{k=1}$ to construct $(G_k^{sp})_{k=1}$ had a negligible effect on the layout of functional communities in the mean network G_0^{sp} and were hence much less pronounced than edge-weight deviations due to inter-subject variability. Thus, community structure translated well

from empirical F_0^{SP} to simulated group-average G_0^{SP} despite pronounced within-group variability in both cohorts $(F_k^{SP})_{k=1}$ and $(G_k^{SP})_{k=1}$.

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 prefrontal cortex (orange, Fig. 3B), which have been shown to play a key role in the cognitive aspects of speech control (Benson et al., 2001; Burton, Locasto, Krebs-Noble, & Gullapalli, 2005; Giraud et al., 2004; 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 to 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.

While simulated resting state networks $(G_k^r)_{k=1}$ closely followed the community architecture of the respective group-averaged reference graph G_0^r , partitions of syllable $(G_k^{syl})_{k=1}$ and speech production $(G_k^{SP})_{k=1}$ 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 to the resting state (both $p < 10^{-5}$), 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 (Fig. 4D–E).

While recent studies reported that resting-state functional architecture was related to anatomical connectivity patterns, with functional modules corresponding to dense anatomical connections (Shen et al., 2012; Shih et al., 2015; Q. 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; Fuertinger et al., 2016; Stanley, Dagenbach, Lyday, Burdette, & Laurienti, 2014) and may therefore be more sensitive to inter-subject variability. Thus, the observed high within-group stability of resting-state functional partitions as compared to 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 multi-tasking, i.e., in subjects 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 inter-subject 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

Supported by the grants from the National Institute on Deafness and Other Communication Disorders (R01DC011805) and the National Institute of Neurological Disorders and Stroke (R01NS088160) to KS.

References

- Andric M, Hasson U. Global features of functional brain networks change with contextual disorder. *Neuroimage*. 2015; 117:103–113. [PubMed: 25988223]
- Bassett DS, Bullmore E. Small-world brain networks. *Neuroscientist*. 2006; 12(6):512–523. [PubMed: 17079517]
- Bassett DS, Porter MA, Wymbs NF, Grafton ST, Carlson JM, Mucha PJ. Robust detection of dynamic community structure in networks. *Chaos*. 2013; 23(1):013142. [PubMed: 23556979]
- Bassett DS, Wymbs NF, Porter MA, Mucha PJ, Carlson JM, Grafton ST. Dynamic reconfiguration of human brain networks during learning. *Proc Natl Acad Sci U S A*. 2011; 108(18):7641–7646. [PubMed: 21502525]
- Bates DM, Machler M, Bolker BM, Walker S. Fitting Linear Mixed-Effects Models using lme4. *ArXiv e-print*. 2015
- Benson RR, Whalen DH, Richardson M, Swainson B, Clark VP, Lai S, et al. Parametrically dissociating speech and nonspeech perception in the brain using fMRI. *Brain Lang*. 2001; 78(3):364–396. [PubMed: 11703063]
- Bertolero MA, Yeo BT, D'Esposito M. The modular and integrative functional architecture of the human brain. *Proc Natl Acad Sci U S A*. 2015
- Betzal RF, Misic B, He Y, Rumschlag J, Zuo X-N, Sporns O. Functional brain modules reconfigure at multiple scales across the human lifespan. 2015
- Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics-Theory and Experiment*. 2008
- Bondy, JA.; Murty, USR. *Graph theory with applications*. New York: North Holland; 1976.
- Bullmore E, Sporns O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci*. 2009; 10(3):186–198. [PubMed: 19190637]
- Burton MW, Locasto PC, Krebs-Noble D, Gullapalli RP. A systematic investigation of the functional neuroanatomy of auditory and visual phonological processing. *Neuroimage*. 2005; 26(3):647–661. [PubMed: 15955475]
- Chang YT, Leahy RM, Pantazis D. Modularity-based graph partitioning using conditional expected models. *Physical Review E*. 2012; 85(1)
- Cover, TM.; Thomas, JA. *Elements of Information Theory*. New York: Wiley; 1991.
- Cox RW. AFNI: software for analysis and visualisation of functional magnetic resonance neuroimages. *Comput Biomed Res*. 1996; 29:162–173. [PubMed: 8812068]
- De Vico Fallani F, Richiardi J, Chavez M, Achard S. Graph analysis of functional brain networks: practical issues in translational neuroscience. *Philos Trans R Soc Lond B Biol Sci*. 2014; 369(1653)
- Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, et al. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage*. 2005; 25(4):1325–1335. [PubMed: 15850749]
- Fakcharoenphol J, Rao S. Planar graphs, negative weight edges, shortest paths, and near linear time. *Journal of Computer and System Sciences*. 2006; 72(5):868–889.
- Fortunato S, Barthelemy M. Resolution limit in community detection. *Proc Natl Acad Sci U S A*. 2007; 104(1):36–41. [PubMed: 17190818]
- Fuertinger S, Horwitz B, Simonyan K. The Functional Connectome of Speech Control. *PLoS Biol*. 2015; 13(7):e1002209. [PubMed: 26204475]

- Fuertinger S, Simonyan K, Sperling MR, Sharan AD, Hamzei-Sichani F. High frequency brain networks undergo modular breakdown during epileptic seizures. *Epilepsia*. 2016 In press.
- Gfeller D, Chappelier JC, De Los Rios P. Finding instabilities in the community structure of complex networks. *Phys Rev E Stat Nonlin Soft Matter Phys*. 2005; 72(5 Pt 2):056135. [PubMed: 16383716]
- Ghosh A, Rho Y, McIntosh AR, Kotter R, Jirsa VK. Noise during rest enables the exploration of the brain's dynamic repertoire. *PLoS Comput Biol*. 2008; 4(10):e1000196. [PubMed: 18846206]
- Ginestet CE, Fournel AP, Simmons A. Statistical network analysis for functional MRI: summary networks and group comparisons. *Front Comput Neurosci*. 2014; 8:51. [PubMed: 24834049]
- Giraud AL, Kell C, Thierfelder C, Sterzer P, Russ MO, Preibisch C, et al. Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cereb Cortex*. 2004; 14(3):247–255. [PubMed: 14754865]
- Girvan M, Newman MEJ. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*. 2002; 99(12):7821–7826. [PubMed: 12060727]
- Godwin D, Barry RL, Marois R. Breakdown of the brain's functional network modularity with awareness. *Proc Natl Acad Sci U S A*. 2015; 112(12):3799–3804. [PubMed: 25759440]
- Good BH, de Montjoye YA, Clauset A. Performance of modularity maximization in practical contexts. *Phys Rev E Stat Nonlin Soft Matter Phys*. 2010; 81(4 Pt 2):046106. [PubMed: 20481785]
- Grabner RH, Ansari D, Reishofer G, Stern E, Ebner F, Neuper C. Individual differences in mathematical competence predict parietal brain activation during mental calculation. *Neuroimage*. 2007; 38(2):346–356. [PubMed: 17851092]
- Hastie, T.; Tibshirani, R.; Friedman, JH. *The elements of statistical learning : data mining, inference, and prediction*. 2nd. New York, NY: Springer; 2009.
- Hermundstad AM, Bassett DS, Brown KS, Aminoff EM, Clewett D, Freeman S, et al. Structural foundations of resting-state and task-based functional connectivity in the human brain. *Proc Natl Acad Sci U S A*. 2013; 110(15):6169–6174. [PubMed: 23530246]
- Hidalgo CA, Klinger B, Barabasi AL, Hausmann R. The product space conditions the development of nations. *Science*. 2007; 317(5837):482–487. [PubMed: 17656717]
- Hothorn T, Bretz F, Westfall P. Simultaneous inference in general parametric models. *Biometrical Journal*. 2008; 50(3):346–363. [PubMed: 18481363]
- Humphries MD, Gurney K, Prescott TJ. The brainstem reticular formation is a small-world, not scale-free, network. *Proc Biol Sci*. 2006; 273(1585):503–511. [PubMed: 16615219]
- Hunter JD. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*. 2007; 9(3): 90–95.
- Karrer B, Levina E, Newman ME. Robustness of community structure in networks. *Phys Rev E Stat Nonlin Soft Matter Phys*. 2008; 77(4 Pt 2):046119. [PubMed: 18517702]
- Kepner, JV.; Gilbert, JR. *Graph algorithms in the language of linear algebra*. Philadelphia: Society for Industrial and Applied Mathematics; 2011.
- Lancichinetti A, Fortunato S. Consensus clustering in complex networks. *Sci Rep*. 2012; 2:336. [PubMed: 22468223]
- Lynall ME, Bassett DS, Kerwin R, McKenna PJ, Kitzbichler M, Muller U, et al. Functional connectivity and brain networks in schizophrenia. *J Neurosci*. 2010; 30(28):9477–9487. [PubMed: 20631176]
- Meila M. Comparing Clusters - An Information Based Distance. *Journal of Multivariate Analysis*. 2007; 98:873–895.
- Miranda-Dominguez O, Mills BD, Carpenter SD, Grant KA, Kroenke CD, Nigg JT, et al. Connectotyping: model based fingerprinting of the functional connectome. *PLoS One*. 2014; 9(11):e111048. [PubMed: 25386919]
- Muller RA, Rothermel RD, Behen ME, Muzik O, Mangner TJ, Chugani HT. Receptive and expressive language activations for sentences: a PET study. *Neuroreport*. 1997; 8(17):3767–3770. [PubMed: 9427367]

- Newman ME, Girvan M. Finding and evaluating community structure in networks. *Phys Rev E Stat Nonlin Soft Matter Phys.* 2004; 69(2 Pt 2):026113. [PubMed: 14995526]
- Newman MEJ. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America.* 2006; 103(23):8577–8582. [PubMed: 16723398]
- Newman SD, Carpenter PA, Varma S, Just MA. Frontal and parietal participation in problem solving in the Tower of London: fMRI and computational modeling of planning and high-level perception. *Neuropsychologia.* 2003; 41(12):1668–1682. [PubMed: 12887991]
- Onoda K, Yamaguchi S. Small-worldness and modularity of the resting-state functional brain network decrease with aging. *Neurosci Lett.* 2013; 556:104–108. [PubMed: 24157850]
- Porter MA, Onnela J-P, Mucha PJ. Communities in Networks. *Notices of the AMS.* 2009; 36(9):1082–1166.
- R-Core-Team. R: A Language and Environment for Statistical Computing. 2016
- Rubinov M, Sporns O. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage.* 2010; 52(3):1059–1069. [PubMed: 19819337]
- Rypma B, D'Esposito M. The roles of prefrontal brain regions in components of working memory: effects of memory load and individual differences. *Proc Natl Acad Sci U S A.* 1999; 96(11):6558–6563. [PubMed: 10339627]
- Shannon CE. A Mathematical Theory of Communication. *Bell System Technical Journal.* 1948; 27(4): 623–656.
- Shen K, Bezgin G, Hutchison RM, Gati JS, Menon RS, Everling S, et al. Information processing architecture of functionally defined clusters in the macaque cortex. *J Neurosci.* 2012; 32(48): 17465–17476. [PubMed: 23197737]
- Shih CT, Sporns O, Yuan SL, Su TS, Lin YJ, Chuang CC, et al. Connectomics-based analysis of information flow in the *Drosophila* brain. *Curr Biol.* 2015; 25(10):1249–1258. [PubMed: 25866397]
- Simonyan K, Fuertinger S. Speech networks at rest and in action: interactions between functional brain networks controlling speech production. *J Neurophysiol.* 2015; 113(7):2967–2978. [PubMed: 25673742]
- Simonyan K, Herscovitch P, Horwitz B. 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.* 2013; 70:21–32. [PubMed: 23277111]
- Sporns O, Betzel RF. Modular Brain Networks. *Annu Rev Psychol.* 2015
- Stanley ML, Dagenbach D, Lyday RG, Burdette JH, Laurienti PJ. Changes in global and regional modularity associated with increasing working memory load. *Front Hum Neurosci.* 2014; 8:954. [PubMed: 25520639]
- Steuer R, Kurths J, Daub CO, Weise J, Selbig J. The mutual information: detecting and evaluating dependencies between variables. *Bioinformatics.* 2002; 2:S231–S240.
- Stevens AA, Tappin SC, Garg A, Fair DA. Functional brain network modularity captures inter- and intra-individual variation in working memory capacity. *PLoS One.* 2012; 7(1):e30468. [PubMed: 22276205]
- Strehl A, Ghosh J. Cluster ensembles - a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research.* 2002; 3:583–617.
- Sturges HA. The choice of a class interval Case I Computations involving a single. *Journal of the American Statistical Association.* 1926; 21:65–66.
- Traag VA, Bruggeman J. Community detection in networks with positive and negative links. *Phys Rev E Stat Nonlin Soft Matter Phys.* 2009; 80(3 Pt 2):036115. [PubMed: 19905188]
- van den Heuvel MP, Sporns O. Rich-Club Organization of the Human Connectome. *Journal of Neuroscience.* 2011; 31(44):15775–15786. [PubMed: 22049421]
- van der Walt S, Colbert SC, Varoquaux G. The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science & Engineering.* 2011; 13(2):22–30.
- Van Essen DC, Smith SM, Barch DM, Behrens TE, Yacoub E, Ugurbil K, et al. The WU-Minn Human Connectome Project: an overview. *Neuroimage.* 2013; 80:62–79. [PubMed: 23684880]

- van Wijk BCM, Stam CJ, Daffertshofer A. Comparing Brain Networks of Different Size and Connectivity Density Using Graph Theory. *Plos One*. 2010; 5(10)
- Wang J, Qiu S, Xu Y, Liu Z, Wen X, Hu X, et al. Graph theoretical analysis reveals disrupted topological properties of whole brain functional networks in temporal lobe epilepsy. *Clin Neurophysiol*. 2014; 125(9):1744–1756. [PubMed: 24686109]
- Wang JH, Zuo XN, Gohel S, Milham MP, Biswal BB, He Y. Graph theoretical analysis of functional brain networks: test-retest evaluation on short- and long-term resting-state functional MRI data. *PLoS One*. 2011; 6(7):e21976. [PubMed: 21818285]
- Wang Q, Sporns O, Burkhalter A. Network analysis of corticocortical connections reveals ventral and dorsal processing streams in mouse visual cortex. *J Neurosci*. 2012; 32(13):4386–4399. [PubMed: 22457489]
- Wu F, Huberman BA. Finding communities in linear time: a physics approach. *European Physical Journal B*. 2004; 38(2):331–338.
- Xia MR, Wang JH, He Y. BrainNet Viewer: A Network Visualization Tool for Human Brain Connectomics. *Plos One*. 2013; 8(7)
- Zachary WW. Information-Flow Model for Conflict and Fission in Small-Groups. *Journal of Anthropological Research*. 1977; 33(4):452–473.

Figure 1. Illustration of the utilized analytic strategy

(I) Based on 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. (II) 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, 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. (III) 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. (IV) The established mean networks were used to compute group-specific resolution parameters for all subsequent modularity maximization calculations.

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 γ was defined as the solution to an optimization problem posed for every condition (rest, syllable, speech, I-III) in both empirical (left column) and simulated (right column) networks. The value of γ 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 $Q^{nd}(\gamma)$ was used in all subsequent modular decomposition calculations in the corresponding group. Thus, condition-specific resolution parameters were defined as $\arg\max f(\gamma)$ (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 α 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 α with shaded areas indicating standard deviations from the mean) between actual and perturbed resting state (I), syllable- (II) and speech-production (III) networks (blue) in comparison to the average partition distance between the associated original and randomized null model graphs $p_d(R_0, R_0(\alpha))$ (red).

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 networks (G_0) are visualized by migration flow diagrams (II). The nodes of F_0 and G_0 are arranged on the left and right arch, 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. *Abbreviations:* 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; THpdf/THpm/THs = prefrontal/premotor/somatosensory part of the thalamus.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 4. Sensitivity of community layouts with respect to the group-averaged reference partitions

Visualization of communities in the group-averaged networks G_0 (inner-most ellipse) and in the constructed 100 leave-one-out graphs $(G_{0 \setminus k})_{k=1}^{100}$ (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 $(G_{0 \setminus k})_{k=1}^{100}$ 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 inter-quartile range. General linear hypothesis testing of the constructed linear mixed effect models in an all-to-all Tukey contrast at $p = 0.05$ yielded significant differences in within-

group perturbation sensitivity between rest and task as well as between tasks (all p 7.1×10^{-6}).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Summary statistics of modular decompositions

		Rest	Syllable	Speech
Empirical networks	Q_z	53.5	32.8	47.2
	$p_d(F_0, F_k)$	0.35±0.03	0.39±0.03	0.4±0.04
Artificial networks	Q_z	59.0	27.1	51.5
	$p_d(G_0, G_k)$	0.36±0.03	0.4±0.03	0.4±0.04
	$p_d(G_0, G_{0,k})$	0.01±0.02	0.05±0.04	0.07±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.