NeuroImage 197 (2019) 24–36

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

A consistent organizational structure across multiple functional subnetworks of the human brain☆

Paul E. Stillman [a,*], James D. Wilson [b], Matthew J. Denny [c], Bruce A. Desmarais [c], Skyler J. Cranmer [d], Zhong-Lin Lu [d]

[a] Yale University, New Haven, CT, 06511, USA
[b] University of San Francisco, San Francisco, CA, 94117, USA
[c] Penn State University, University Park, PA, 16802, USA
[d] The Ohio State University, Columbus, OH, 43210, USA

ARTICLE INFO

Keywords:
Network neuroscience
cGERGM
Cognitive neuroscience

ABSTRACT

A recurrent theme of both cognitive and network neuroscience is that the brain has a consistent subnetwork structure that maps onto functional specialization for different cognitive tasks, such as vision, motor skills, and attention. Understanding how regions in these subnetworks relate is thus crucial to understanding the emergence of cognitive processes. However, the organizing principles that guide how regions within subnetworks communicate, and whether there is a common set of principles across subnetworks, remains unclear. This is partly due to available tools not being suited to precisely quantify the role that different organizational principles play in the organization of a subnetwork. Here, we apply a joint modeling technique – the correlation generalized exponential random graph model (cGERGM) – to more completely quantify subnetwork structure. The cGERGM models a correlation network, such as those given in functional connectivity, as a function of activation motifs – consistent patterns of coactivation (i.e., connectivity) between collections of nodes that describe how the regions within a network are organized (e.g., clustering) – and anatomical properties – relationships between the regions that are dictated by anatomy (e.g., Euclidean distance). By jointly modeling all features simultaneously, the cGERGM models the unique variance accounted for by each feature, as well as a point estimate and standard error for each, allowing for significance tests against a random graph and between graphs. Across eight functional subnetworks, we find remarkably consistent organizational properties guiding subnetwork architecture, suggesting a fundamental organizational basis for subnetwork communication. Specifically, all subnetworks displayed greater clustering than would be expected by chance, but lower preferential attachment (i.e., hub use). These findings suggest that human functional subnetworks follow a segregated highway structure, in which tightly clustered subcommunities develop their own channels of communication rather than relying on hubs.

Cognitive neuroscientists have identified a roughly consistent set of distinct functional subregions, also known as subnetworks, that are reliably involved in cognitive functions (e.g., Bressler and Menon, 2010). These subnetworks are consistent with those identified via community detection techniques (Power et al., 2011; Yeo et al., 2011) used in network neuroscience (Medaglia et al., 2015; Betzel and Bassett, 2017; Sporns, 2011; Bullmore and Sporns, 2009; Bassett and Sporns, 2017). This consistency suggests that these subnetworks reflect meaningful subdivisions of the brain. Understanding how these subnetworks operate may therefore be crucial to understanding how the brain balances local versus distributed processing. In the present paper, we model functional connectivity using the correlation generalized exponential random graph model (cGERGM, Stillman et al., 2017) in order to better understand the organizational properties of these subnetworks.

☆ MD was supported by NSF Grant DGE-1144860. BD was supported by NSF grants SES-1558661, SES-1619644, SES-1637089, and CISE-1320219. SJC was supported by NSF grants SES-1357622, SES-1461493, and SES-1514750, NIH R-34 DA043079, and the Alexander von Humboldt Foundation's Fellowship for Experienced Researchers. ZLL and PES were supported by NSF-SMA 1533500. JDW was supported by NSF grant DMS-1830547. The authors have intellectual property considerations for using cGERGM in a medical context. SJC, JW, ZLL, and MD are co-founders and equity holders in Cerenetics, Inc., a start-up company commercializing the technology described in this paper.

* Corresponding author. 5415C Evans Hall, 165 Whitney Ave, Yale University, New Haven, CT, 06511, USA.
E-mail address: paul.e.stillman@gmail.com (P.E. Stillman).

https://doi.org/10.1016/j.neuroimage.2019.03.036
Received 7 December 2018; Received in revised form 15 February 2019; Accepted 17 March 2019
Available online 28 March 2019

When studying functional connectivity, researchers first divide the brain into regions (i.e., nodes) and then gauge connectivity between nodes via the correlation (or some metric of co-activation, see Smith et al., 2011) across the time-series, with nodes that correlate more strongly with one another inferred to be more strongly connected (see Fig. 1). Here, we restrict our investigation to intracommunity connections, yielding eight subnetworks responsible for different cognitive processes (see Table 1).

Despite the central importance of these subnetworks, the principles that govern their organization remain understudied, with the majority of investigations focusing on whole-brain as opposed to subnetwork organizational principles (Bullmore and Sporns, 2009, 2012). We refer collectively to the organizational principles that describe the patterns by which small collections of nodes connect to one another as *activation motifs*. Current techniques typically quantify a single activation motif in isolation (Rubinov and Sporns, 2010). For example, the degree to which a node's neighbors are neighbors with one another is referred to as clustering (thought to quantify intra-community information integration, Bullmore and Sporns, 2009). Researchers can use these activation motifs to make inferences about network structure. For instance, using these techniques, research has found that the whole-brain follows a "small-world" structure – a configuration that consists of densely intra-connected communities connected via hubs (i.e., nodes that have disproportionately high number of connections) that maximizes local and global network efficiency while minimizing wiring costs (Bullmore and Sporns, 2012; Bassett and Bullmore, 2017). It remains unclear, however, whether subnetworks – which are organized to address more specialized functions compared to the more integrative requirements of the whole-brain – show similar or different organization to the whole-brain.

Additionally, while these approaches are useful for investigating networks across task, time, and individual, their isolated application carries two major drawbacks. First, activation motifs are often highly correlated with one another, and may have unexpected interactions that are only apparent when considering multiple motifs simultaneously. Second, current techniques are unable to separate the influence of activation motifs from the influence of *anatomical properties* – the anatomical context of the brain that influences connection strength. For example, the distance between nodes is often strongly related to connection strength (Honey et al., 2009), and may confound attempts to quantify activation motifs. Further, many existing techniques provide only a single point estimate of a given statistic, making it difficult to compare that value against chance. Finally, although weighted metrics are becoming more readily available (e.g., Sizemore et al., 2016), many approaches require thresholding of data (i.e., setting weak connections to 0), thus discarding potentially important information.

Table 1
Brief description of each network modeled in the present paper.

Network	Function
Auditory	Hearing
Subcortical	Subcortical structures such as the thalamus, putaman, and brainstem
Dorsal Attention	Top-down influences on attention
Ventral Attention	Attention switching
Salience	Directing attention to motivationally and emotionally relevant external stimuli
Cingulo-Opercular Task Control	Initiating and sustaining task behavior
Fronto-Parietal Task Control	Top-down exertion of control and executive function
Default Mode	Internally generated cognition (though precise function still debated)

To more comprehensively characterize subnetwork organization, we use the cGERGM – a new modeling framework that simultaneously quantifies multiple elements of network architecture in a way that separates the influence of activation motifs and anatomical properties (Stillman et al., 2017). Using the cGERGM, we attempt to answer two related questions. First, we investigated the organizing principles (both activation motifs and anatomical properties) driving subnetwork organization. Of particular interest to the present investigation are two activation motifs – clustering and preferential attachment. Clustering refers to the tendency for nodes' neighbors to connect with one another, thereby forming connected triads. Extensive prior work has shown that the brain consists of densely clustered communities (i.e., the subnetworks that are the focus of the present paper Power et al., 2011), and further evidence suggest this is hierarchical, with the subnetworks themselves comprised of multiple smaller densely interacting clusters of regions (Meunier et al., 2010). Based on these findings, we expect to find significantly greater clustering than would be predicted by chance. Preferential attachment refers to the tendency of nodes to connect to other strongly connected nodes across the network (i.e., across the network, are nodes preferentially attaching to other nodes that have greater connections). Preferential attachment quantifies the extent to which hubs are a structural feature of the network and occur greater than would be expected by chance. While extensive past work has documented the importance of hubs in the brain (e.g., van den Heuvel and Sporns, 2011, 2013; Andrews-Hanna, 2012), past work has generally viewed hubs as a property of a node – if a given node has a large number of connections compared to the other nodes in the network, that node is considered a hub. The cGERGM, however, allows us to ask a different question – is the structure of the network such that hubs are more

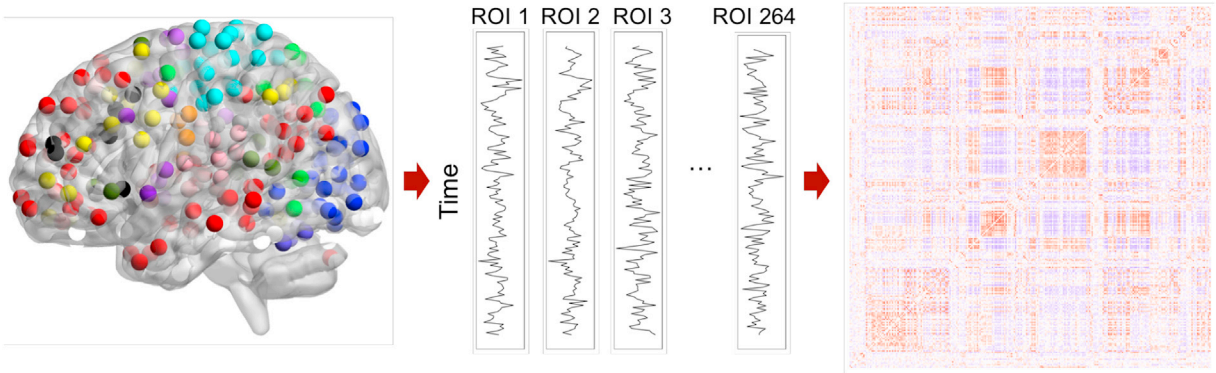

Fig. 1. Visual depiction of network construction in the context of resting state functional connectivity. Left: the brain is divided up into nodes. Here, we use the 264 node atlas described by Power et al. (2011), which further provides subnetwork membership of each node, indicated here by their color. In addition to this, we use a 20-node atlas of the default mode network described by Bellana et al. (2016a, b). Center: once nodes have been established, we extract the time series for each node (averaged across all voxels within the node), yielding a time series for each node. Right: we then correlate these time series to yield the weighted, undirected correlation network.

prevalent than would be expected by chance? Some initial evidence suggest that preferential attachment may not, in fact, be a property of subnetworks. For instance, our past work using the cGERGM to investigate the default mode network (Stillman et al., 2017) found significantly greater clustering than would be expected by chance, but significantly *less* preferential attachment than would be expected by chance. This suggests that the default mode network displays "segregated highway" – a structure organized to promote functional specialization in which communities build their own direct channels of communication rather than relying on hubs.

Beyond investigating the organization of each individual subnetwork, we can investigate similarities and differences across subnetworks. In particular, it is possible that each subnetwork is organized differently to address the different challenges of the cognitive process that network controls. On the other hand, it is also possible that a consistent organizational framework is present across subnetworks, possibly suggesting that a certain communication structure is optimal for coordination amongst brain regions to give rise to complex cognitive processes. By investigating the similarities and differences across subnetworks, we hope to gain a clearer understanding of how complex cognitive processes can emerge from the coordination of different brain regions.

1. Methods

1.1. Participants

For the present investigation, we use resting state data of 21 unrelated healthy young adults (ages 22–35, 10 Female) publicly available from the Human Connectome Project (HCP) S1200 release, "100 unrelated subjects" list (https://www.humanconnectome.org/, Van Essen et al., 2013).[1]

1.2. Scanning parameters

Each participant completed a number of structural, functional, and diffusion scans. We focus on a single resting state scan (REST1-LR), with full details of the other scans available on HCP. Acquisition was conducted using a 3-T scanner. Each resting state scan consisted of 1200 functional T2*-weighted blood-oxygen level-dependent (BOLD) multiband EPI (Moeller et al., 2010; Feinberg et al., 2010; Setsompop et al., 2012; Xu et al., 2012) images, with TR of 720 ms, TE of 33.1 ms, slice thickness of 2 mm, field of view $= 208 \times 180$ mm, matrix $= 104 \times 90$, and flip angle of 52°.

1.3. Data preprocessing

We use the preprocessed data provided by the HCP. These data have been put through the minimal data preprocessing pipeline (Glasser et al., 2013). The data is processed for spatial artifacts (spatial distortion correction, head motion correction, and corrected for B_0 distortion), registered to 2 mm MNI space, and subjected to global intensity normalization (Glasser et al., 2013; Smith et al., 2013). Following this, these data were subjected to ICA denoising via ICA-FIX (Salimi-Khorshidi et al., 2014) – an automated ICA based approach to removing noise, with particular utility for removing motion-induced noise (see Ciric et al., 2017).

1.4. Subnetwork specification and network creation

We use a previously validated atlas which divides the brain into 264 spherical ROIs that served as our nodes (Power et al., 2011). For each ROI and participant, we extracted all voxels within that ROI and averaged them, yielding one time-series per ROI per participant. We then regressed out from these time-series' the average white-matter time-series, the average cerebrospinal fluid time-series, and a time-series corresponding to the mean global signal (Power et al., 2012; Ciric et al., 2017). To gauge connectivity across these 264 regions for each participant, we then took the pairwise correlation across all regions, yielding a 264×264 correlation matrix corresponding to the whole brain network. Finally, to capture subnetwork structure, we divide the correlation matrix into 12 functional subnetworks delineated by the atlas (Power et al., 2011), yielding a correlation matrix for each subnetwork. Of these 12, 2 are too small to possess interesting network structure and also to recover reliable cGERGM parameters – Cerebellar (4 nodes) and memory/retrieval (5 nodes). As such, we omitted them from the analysis. As the cGERGM requires large computational resources as the networks increase in size, for the default mode network we use a 20-node partition (Bellana et al., 2016a, b) used in our prior work (Stillman et al., 2017) rather than the 58 node network given defined by Power and Colleagues.[2] We further omit here the results of the sensory/somatomotor network (35 nodes, 11 out of 21 participants converged) and visual network (31 nodes, 18 out of 21 participants converged), as convergence was inconsistent on a reasonable time-line. We note, however, that the results for these models that successfully converged are consistent with the results presented here (and are presented in the supplement). Finally, we omitted all nodes for which network assignment was uncertain. This left us with 8 functional subnetworks.

1.5. The correlation generalized exponential random graph model (cGERGM)

We model each of these 8 functional subnetworks separately with the cGERGM across all 21 adults. The cGERGM takes as input a weighted correlation network (meaning it does not require thresholding of the data) and models the network as a function of multiple (user-specified) activation motifs and/or anatomical properties. Here, we consider five such features – edges, two-stars, triads, Euclidean distance, and hemisphere – however we note one strength of the cGERGM is its ability to specify nearly any feature the researcher believes may matter in accounting for functional connectivity. The effects of specified anatomical properties are modeled using a Beta regression of the mean of each partial correlation weight on the properties of interest (equation (A.3) in the Appendix). The activation motifs of the normalized partial correlation network are modeled using an exponential family of probability distributions that parameterize the effects of motifs on the weight of the edges within the network. The resulting model is a joint probability distribution on the correlation network with specified motifs and anatomical properties as covariates (equation (A.4)).

Activation motifs quantify relational patterns that specify how nodes interact with one another. Here, we discuss three such activation motifs (see Fig. 2). The first is network density (i.e., "edges"), which measures the average strength of connections between regions in the network, and serves as an intercept for the model. The second is the tendency for nodes to show "preferential attachment," a phenomenon in which nodes are more likely to be connected to nodes that are highly connected

[1] Subject numbers: 100307, 100408, 101107, 101309, 103111, 103818, 106016, 108828, 110411, 111312, 111716, 113619, 113922, 114419, 115320, 116524, 499566, 654754, 672756, 751348, 899885. These 21 participants were selected without regard to demographic or anatomical differences. Our sample size was selected to balance computational time considerations with having a sufficiently large sample to draw conclusion about the nature of network organization across participants.

[2] This specification of the DMN appears to be consistent (though smaller in scope) to the specification defined by Power and Colleagues – our 20 node network overlapped at least partially with 24 nodes of the 58-node power atlas. We further note that there were minor overlaps with one node each from the visual, salience, and ventral attention networks. However, these overlaps were fairly minor, constituting only 10, 6, and 2 voxels (respectively).

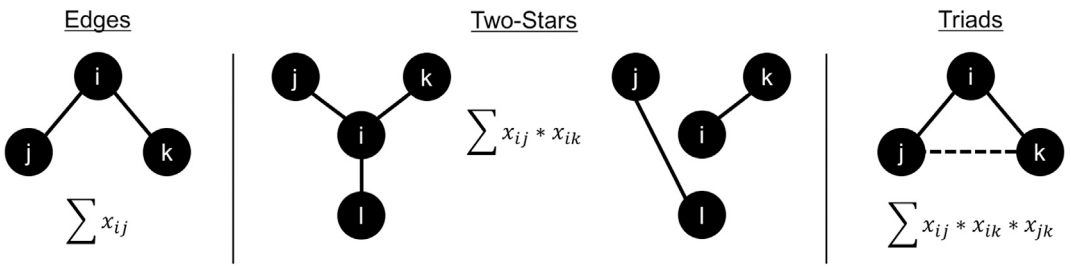

Fig. 2. Activation motifs modeled by the cGERGM for a network x. Left: the edges statistic simply models the average connectivity across the network, and is equivalent to network density. Middle: two-stars capture preferential attachment – the degree to which nodes are likely to be connected to other nodes that have many connections. This is quantified by, for each trio of nodes, taking the product of the edges connected to the focal node. In the figure, the graph on the left has high two-stars (4), compared to the graph on the right (0). This metric will therefore be higher when more nodes are more strongly connected to the same node, and graphs with high two-stars will thus have more centralized hubs. Right: triads models the degree to which nodes cluster together into tight-knit communities. It is quantified by, for each trio of nodes, taking the product of all the edges between the nodes.

themselves. Networks that demonstrate preferential attachment have non-flat degree distribution – certain nodes that have many connections, and most others have few connections – and are often described as being reliant on hubs to integrate information. This is quantified via a metric called "two-stars", calculated by, for each trio of nodes, taking the product of the edges with the focal node. The third motif quantifies the tendency for two nodes connected to a third to be connected to one another. This is called "triadic closure" or "clustering," and is found in many networks across disciplines and application areas.

Anatomical properties quantify how other attributes of the nodes or relationships between them might influence connection strength. Here, we consider two such features. First, we consider the spatial Euclidean distance between two brain regions. In past studies, the Euclidean distance has been shown to be strongly related to the connection strength between two regions (Honey et al., 2009), though recent literature suggests that these relationships may be a consequence of motion-induced noise (Power et al., 2012; Satterthwaite et al., 2012). We can thus include a matrix that contains the distances between each pair of nodes in the cGERGM, and use that to predict how strongly related two nodes are to one another. We also consider a second anatomical property, which measures whether or not two regions reside in the same hemisphere, with the prediction that intrahemispheric connections will be stronger than interhemispheric connections.

Rather than modeling each feature individually, the cGERGM characterizes the individual effect of each feature through a joint probability distribution that quantifies the likelihood of the given network as a function of both activation motifs and anatomical properties. Thus the network under study is modeled via a parametric probability distribution, where parameters quantify the unique effect of features specified in the model. In addition to yielding more informative estimates of network structure, these distributions can further be used to statistically assess the degree to which a given activation motif or anatomical property influences the structure of the network against chance. The cGERGM therefore allows for testing of the unique contribution of each feature of interest, controlling for the others, while accounting for all of the observed data within the correlation matrix.

Typically, models that include both activation motifs and anatomical properties produce better fit than models that include just one or the other (Stillman et al., 2017). Therefore, for each network of each participant, we use the full model with all five features (21 participants $\times$ 8 subnetworks $=$ 168 models total).

1.6. cGERGM model fitting overview

The specification of the cGERGM requires two steps (see Appendix A for a more in-depth description of the mathematical specification of the cGERGM). First, the edges of the observed correlation network, which naturally take values on [-1, 1], are normalized as partial correlations that lie on [0, 1] via a one-to-one multivariate transformation. This is done to enable inference as the (unique) partial correlations associated with a correlation matrix do not suffer from the mathematical constraints on the correlation matrix itself, such as the positive semidefinite constraint. The effects of specified anatomical properties are modeled using a Beta-distribution regression of the mean of each partial correlation weight on the properties of interest, while the activation motifs of the normalized partial correlation network are modeled using an exponential family of probability distributions that parameterize the effects of motifs on the weight of the edges within the network. The resulting model is a joint probability distribution on the correlation network whose summary statistics are given by the specified motifs and anatomical properties (equation (A.4)).

To quantify the effects of both the anatomical properties and the activation motifs, the joint likelihood in equation (A.4) is estimated via maximum likelihood. As the probability distribution of a cGERGM has an intractable normalizing constant, Markov Chain Monte Carlo (MCMC) simulation is utilized to carry out the estimation. For any model in the cGERGM family, estimation proceeds as follows. First, the estimation algorithm makes an initial seed for all model parameters. These parameters are then used to simulate a large sample of networks from the model generative process. These networks are used to approximate the log likelihood of the model for the current set of model parameters. Parameter estimates are updated according to a greedy update that yields a higher log-likelihood value given the observed data. This iterative process is repeated until the parameter estimates converge. Detailed information about fitting the cGERGM is provided in the Appendix.

1.7. Data statement

All models in this paper were fit using the publicly available R *GERGM* software (Denny et al., 2016). The scripts used to generate the present results, along with each of the eight correlation networks, are available at the Open Science Framework: osf.io/3fmvj.

2. Results

2.1. Model fit

Before interpreting model coefficients, we first investigate model fit. This is important because model coefficients are uninterpretable if fit is poor. To assess fit of each model, we examine whether our model produces simulated networks that have the same basic properties as the observed network. To this end, we simulate 500 networks from the estimated model parameters (Hunter et al., 2008). From these simulated networks, we can compare whether various structural features of the network (both those included in the model as well as those that are not) are similar for the observed and simulated networks. Here, we use five structural features of the network – triads, two-stars, edges, network intensity, and degree distribution – to gauge how well our model is

recreating the observed data. To compare the simulated and observed values, we generate box-plots (for triads, two-stars, edges, and intensity) and density plots (for degree distribution) reflecting the values of each respective statistic over the 500 simulated networks, and compare those distributions to our observed values. Good fit is achieved when there is a strong correspondence between the inter-quartile range of the boxplots and the observed value, and a similar degree distribution for the observed and simulated networks. We plot all boxplots/density plots for all networks of a single representative participant in Figs. 3 and 4 (plots for the other participants can be generated via the code available on OSF). This participant appears to be achieving good model fit, as the large majority of the observed values are within the IQR of the simulated boxplots, and the degree distributions for observed and simulated roughly match one another. Across all participants, we see similar patterns, with fit summaries across participants given in Table 2. Given these fit results, the parameters recovered by the model can generally be thought to be unbiased and to reflect the data generating process for the subnetworks reasonably well.

As an additional way of visualizing the success of our models, we can plot the observed networks and compare them to the simulated networks. For plotting purposes, we find a simulated network that is highly likely given the estimated model, and then visually compare how our simulated networks appear relative to the observed network. To find a network that is highly likely given the model, we again use the 500 simulated networks. We then evaluate the approximate likelihood function, given in Equation (A.5), with each simulated network. We plot the network that scores the highest on the likelihood function (i.e., the simulated network most likely to be observed under the estimated model). As can be seen in Fig. 5, we get decent (though imperfect) recreation of the networks of interest for a representative participant.

2.2. Parameter results

For each subnetwork, we investigated the parameter outputs generated by the cGERGM. For each feature specified, the cGERGM provides (for each participant) both a point estimate and standard error corresponding to the unique influence of that statistic on network organization (parameter estimates and standard errors are given in Table 3). This thus allows for inferential testing of the influence of these features relative to chance. For all features, significant positive parameters correspond to significantly greater influence of that feature in shaping connection strength within the network. Similarly, significant negative parameters correspond to less influence of that feature than would be expected by chance. Results of the edges term are plotted in Fig. 6, with results for the two-stars, and triads parameter estimates plotted in Fig. 7, and the results for the anatomical properties given in Fig. 8. As there was (in general) a high degree of consistency across networks, we discuss the results of each statistic in turn.

Edges. Across all subnetworks, we find a highly consistent pattern such that the edges term is almost uniformly positive, but with a high degree of variability. The positive edges effect was significant for the vast majority of participants and subnetworks, with all participants displaying a significant edges parameter for every network except for Auditory (43% non-significant), Ventral Attention (33% non-significant), and Subcortical (5% non-significant). This suggests that, overwhelmingly, these eight subnetworks are more densely connected than would be expected by chance. This finding is perhaps unsurprising given that these subnetworks were defined by partition methods that maximize intra-community connection while minimizing intercommunity connection.

Triads. As can be seen in Fig. 7, across all eight subnetworks, we find a highly consistent pattern such that almost every network is demonstrating a high degree of clustering (i.e., triads parameter greater than 0). For the larger subnetworks (Cingulo-opercular, Salience, Default Mode, and Fronto-parietal), every single participants' triads parameter estimate was significantly different from 0. The other networks each had some participants whose triads coefficient was not significantly different than 0 (across all participants, none had significantly negative triads estimates). To investigate overall patterns, we conducted a one-sample *t*-test testing whether the mean triads coefficient was different than zero. Consistent with the results of the larger networks, all four networks were significantly different from zero, Ventral attention: $t(20) = 8.05$, $p < .001$, Dorsal attention: $t(20) = 13.50$, $p < .001$, Auditory: $t(20) = 26.45$, $p < .001$, Sub-cortical: $t(20) = 6.33$, $p < .001$. Together, these results provide strong evidence that, in general, the present subnetworks are displaying higher clustering than would be expected by chance. This is consistent with extant theorizing on the hierarchical community organization of the brain. Specifically, past work has suggested that the brain consists of tightly clustered communities (subnetworks) that themselves consist of tightly clustered subsubnetworks (Meunier et al., 2010).

Two-Stars. Across all of the subnetworks, we again find a highly consistent pattern across participants. However, as can be seen in Fig. 7, this pattern indicated a *negative* two-stars parameter across subnetworks (i.e., lower hub use). As with the triads parameter, this pattern was significant in every single participant for the larger subnetworks (Cingulo-opercular, Salience, Default Mode, and Fronto-parietal). As above, to investigate the significance of the smaller subnetworks, we conducted a one-sample *t*-test testing whether the mean two-stars coefficient was

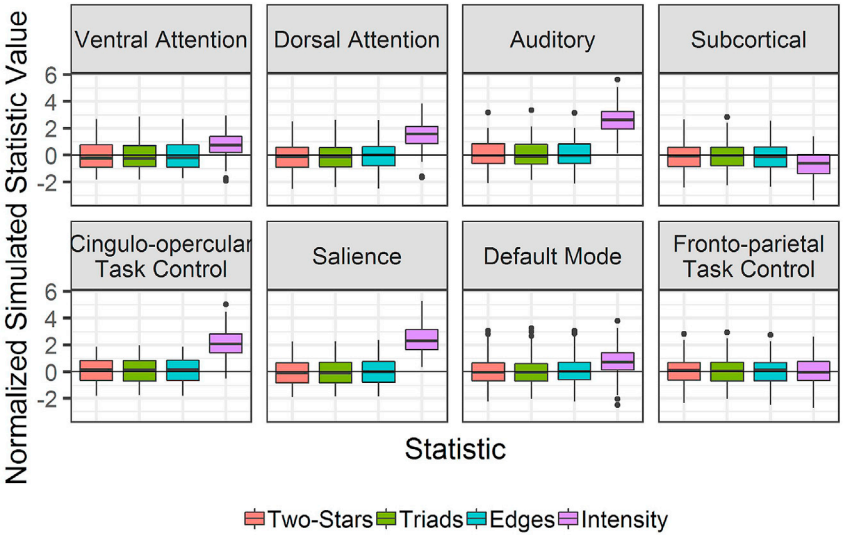

Fig. 3. GOF plots of all networks for a representative participant (participant 100307). For each network, 500 networks are simulated from the output parameters. For each of these simulated networks, we calculate the four statistics of two-stars, triads, edges, and intensity to compare to the observed values. For plotting purposes, we normalize these simulated values, and then offset them from the observed value. In other words, in these plots, 0 represents the observed value, and good fit corresponds to low deviance from 0 – typically, when the observed value is within the IQR of the simulated value, it is considered good fit.

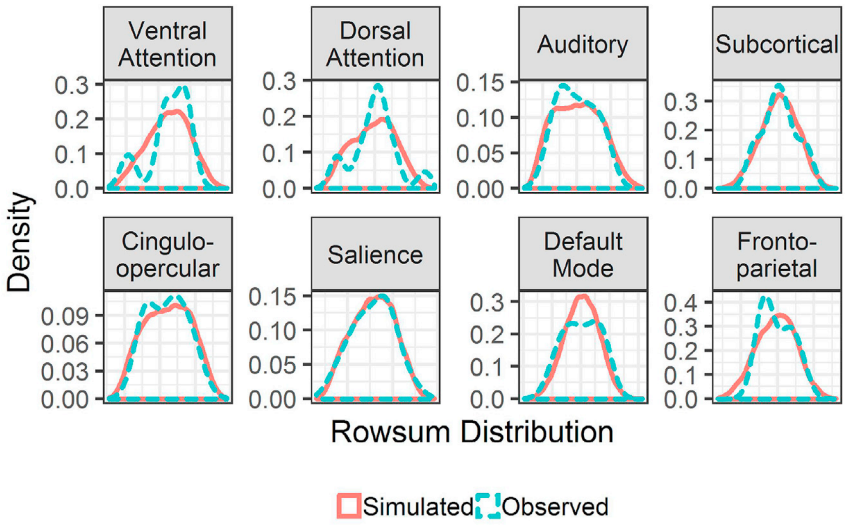

Fig. 4. Density plots of observed and simulated connection values for all networks of a representative participant (participant 100307). Values are obtained by summing the connection value of each node (i.e., the rowsums of a given network). This is a valuable additional metric on which to assess model fit because it captures not only aggregate patterns in connectivity, but the distributions over connections to individual nodes. Here again, we see a good model fit for most subnetworks.

Table 2
Goodness of fit across all participants for the four parameters of interest for each network. The value in the cells corresponds to the average absolute deviation between the observed value and the median of the simulated value. The percentage in parentheses corresponds to the percentage of participants whose observed value was within the IQR.

Network	Two-Stars	Triads	Edges	Intensity
Ventral Attention	0.25 (100%)	0.33 (100%)	0.56 (100%)	0.02 (57%)
Dorsal Attention	0.39 (100%)	0.64 (100%)	1.07 (100%)	0.02 (38%)
Auditory	0.67 (100%)	1 (100%)	1.89 (100%)	0.02 (14%)
Subcortical	0.2 (100%)	0.3 (100%)	0.53 (100%)	0.01 (57%)
Cingulo-opercular Task Control	0.93 (100%)	1.58 (100%)	1.98 (100%)	0.02 (24%)
Salience	1.24 (100%)	1.77 (100%)	3.47 (100%)	0.03 (5%)
Default Mode	1.98 (95%)	2.65 (95%)	5.46 (95%)	0.04 (0%)
Fronto-parietal Task Control	2.69 (95%)	4.16 (95%)	8.26 (95%)	0.02 (10%)

different than zero. Consistent with the results of the larger networks, all four networks displayed significantly negative parameters across participants, Ventral attention: $t(20) = -8.40$, $p < .001$, Dorsal attention: $t(20) = -14.20$, $p < .001$, Auditory: $t(20) = -27.30$, $p < .001$, Subcortical: $t(20) = -6.46$, $p < .001$ Overall, this suggests that the networks are demonstrating anti-preferential attachment, meaning they are less likely to use hubs than would be expected by chance alone. Unlike the above findings, the lack of preferential attachment is somewhat surprising given previous work and research emphasizing the importance of hubs, at least at the whole brain level. As noted above, this discrepancy may result from the fact that the present investigation tests whether a network structurally demonstrates preferential attachment, rather than whether certain nodes have disproportionate strength of connections. Our results are consistent, however, with our past work on the default mode network, and suggest that subnetworks more generally may be demonstrating segregated highway architecture, a point we return to in the discussion.

Anatomical properties – Euclidean distance and hemisphere. Unlike the results for activation motifs – which are largely consistent across subnetwork and participant – results for the Euclidean distance and hemisphere parameters are inconsistent both across subnetwork and across participants (see Fig. 8). For hemisphere, there appears to be a consistent

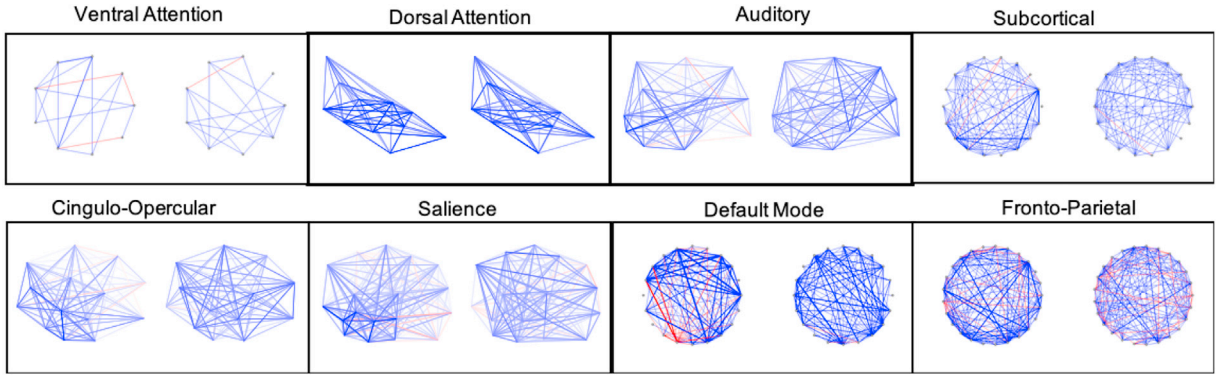

Fig. 5. Observed networks (left) against simulated networks (right) for all networks for a representative participant. To determine the positions of each node, the observed networks are thresholded at 1/2 of the maximum absolute correlation value and node positions are then determined using the Fruchterman-Reingold (FR) force-directed graph layout algorithm (Fruchterman and Reingold, 1991). The same positions are then used for the highest likelihood simulated network. The network layouts for ventral attention, subcorical, default mode, and fronto-parietal networks appeared very stretched out along one dimension when using FR, so for these networks we instead use a circle layout.

Table 3
Average parameter estimates, average standard errors, and the percentage of participants for whom the parameter was significantly different from 0 for each of the five features modeled.

Network	Edges	Triads	Two-Stars	Hemisphere	Distance
Ventral Attention	0.5(0.13) – 67%	5.07(2.84) – 57%	−7.82(4.15) – 57%	−0.04(0.07) – 29%	−0.07(0.04) – 38%
Dorsal Attention	0.7(0.11) – 100%	6.29(2.28) – 81%	−9.54(3.35) – 81%	−0.04(0.04) – 14%	−0.08(0.03) – 57%
Auditory	0.36(0.15) – 57%	8.16(1.59) – 90%	−12.36(2.31) – 90%	0.09(0.08) – 29%	0.04(0.04) – 29%
Subcortical	0.21(0.03) – 95%	4.65(2.45) – 57%	−7.05(3.65) – 57%	−0.02(0.01) – 33%	−0.01(0.01) – 52%
Cingulo-opercular Task Control	0.56(0.06) – 100%	8.88(1.32) – 100%	−13.4(1.93) – 100%	−0.01(0.02) – 19%	−0.02(0.01) – 33%
Salience	0.52(0.07) – 100%	9.01(1.18) – 100%	−13.62(1.71) – 100%	−0.01(0.02) – 14%	−0.03(0.01) – 62%
Default Mode	0.65(0.08) – 100%	9.6(0.91) – 100%	−14.49(1.29) – 100%	0.07(0.03) – 71%	−0.01(0.01) – 19%
Fronto-parietal Task Control	0.55(0.07) – 100%	10.16(0.77) – 100%	−15.31(1.13) – 100%	0.05(0.02) – 52%	−0.02(0.01) – 62%

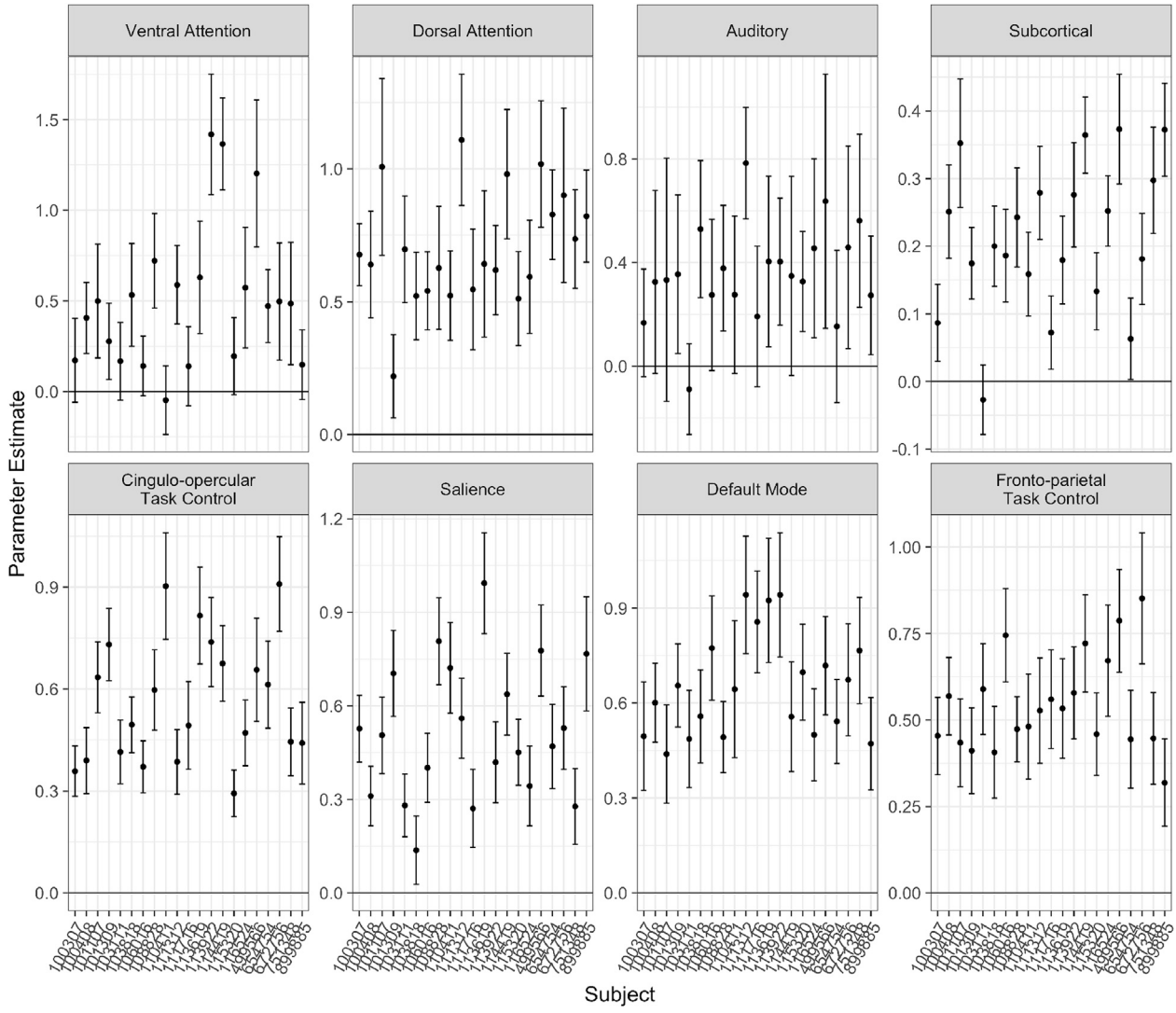

Fig. 6. Edges (density) parameter estimates for each subject for each of the eight different subnetworks investigated. Error bars correspond to 95% confidence intervals.

pattern such that there is a significantly positive effect of hemisphere (i.e., nodes are more strongly connected when in the same hemisphere) for the fronto-parietal and default mode networks, but the remaining networks were inconsistent, with participants displaying significantly positive, significantly negative, and non-significant effects of hemisphere on connectivity. If we conduct one-tailed t-tests compared to 0, we find a significantly positive effect of hemisphere for the Default Mode ($t(20) = 6.74$, $p < .001$), Fronto-Parietal ($t(20) = 5.87$, $p < .001$), and Auditory ($t(20) = 3.79$, $p = .001$) subnetworks, a significantly negative effect for the Dorsal Attention ($t(20) = -3.74$, $p < .001$), and Subcortical ($t(20) = -3.61$, $p = .002$), networks, and non-significant effects for the Cingulo-Opercular, Salience, and Ventral Attention networks ($ps \geq .15$). Thus, it appears the influence of hemisphere is dependent on the subnetwork of interest.

Distance was similarly inconsistent across individual and subnetwork, with all networks having between only 20% and 62% of individual subject parameters significant. Notably, and in contrast to hemisphere, if we once again conduct t-tests against 0, we find a consistent finding across all subnetworks except for auditory such that there is a significantly negative influence of distance – in other words, the further apart

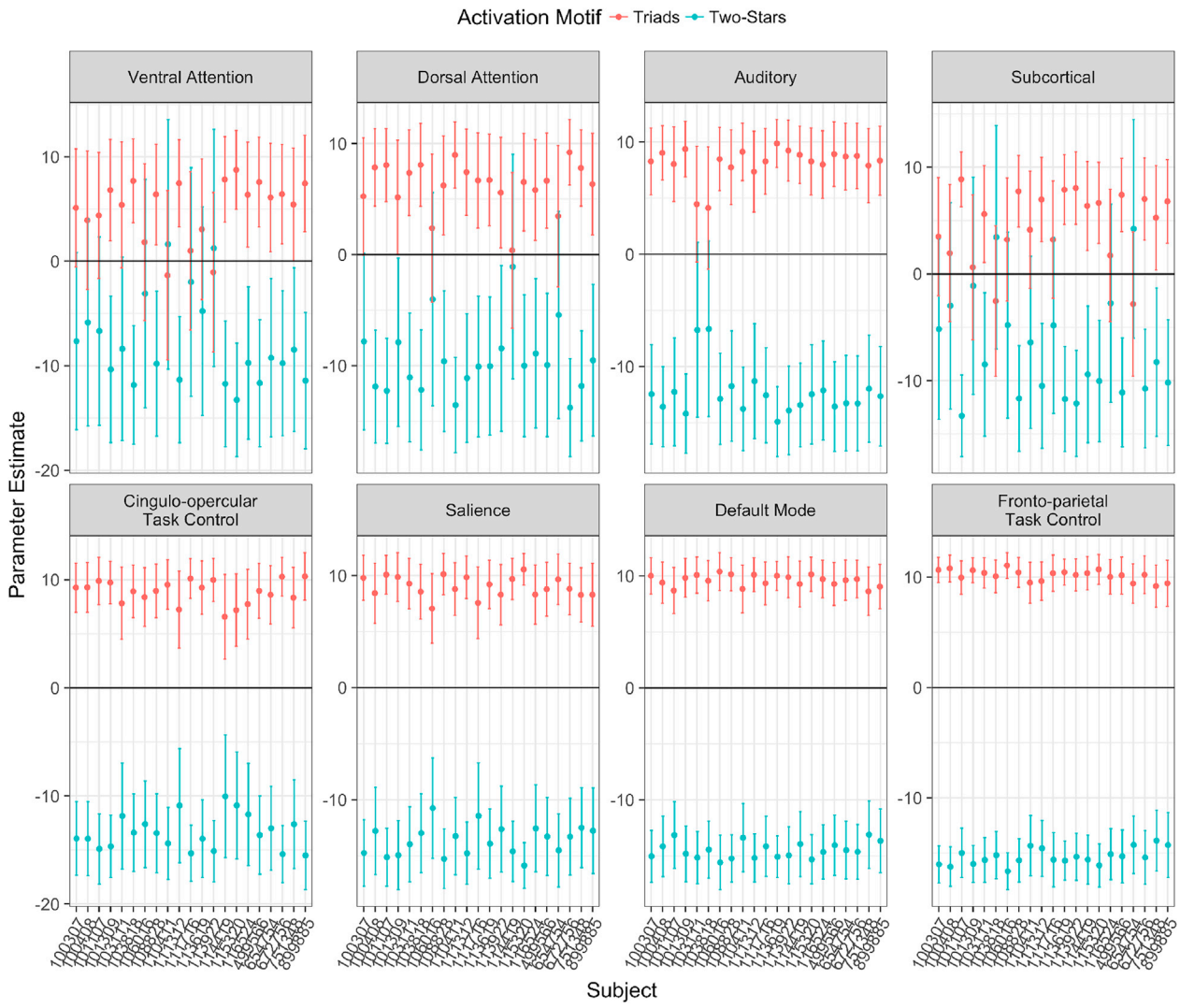

Fig. 7. Two-stars (preferential attachment/hub use, blue, negative points) and triads (clustering, red, positive points) parameter estimates for each subject for each of the eight different subnetworks investigated. Error bars correspond to 95% confidence intervals.

two nodes, the less strongly connected they are (Ventral Attention, $t(20) = -3.18$, $p = .005$; Dorsal Attention, $t(20) = -5.36$, $p < .001$; Subcortical, $t(20) = -4.80$, $p < .001$; Cingulo-Opercular Task Control, $t(20) = -3.07$, $p = .006$; Salience, $t(20) = -6.80$, $p < .001$; Default Mode, $t(20) = -2.13$, $p = .05$, Fronto-Parietal Task Control, $t(20) = -6.63$, $p < .001$. The auditory network, in contrast, had a significantly positive distance parameter, $t(20) = 2.52$, $p = .02$.

3. Discussion

Quantifying the organization of the many functional subnetworks of the human brain is critical to understanding how regions within these subnetworks organize to give rise to emergent cognitive processes. We used the correlation generalized exponential random graph model (cGERGM) to simultaneously model organizing principles for eight functional subnetworks of the human brain. Across all eight functional subnetworks, we found remarkably consistent influences of activation motifs, such that all subnetworks displayed greater clustering than would be predicted by chance, and less preferential attachment (i.e., hub usage) than would be expected by chance. Nearly all individual subjects' networks display this pattern, suggesting this network configuration is relatively robust. Additionally, across all networks we find significantly greater edge strength (i.e., density) than would be expected by chance. Further, we found weak but suggestive evidence that the greater the Euclidean distance between nodes, the weaker their connection, though we note that this effect only reached significance when averaged across subjects. Interestingly, there were no consistent effect for hemisphere, with certain networks showing stronger intrahemispheric connections, others showing stronger interhemispheric connections, and some showing no significant differences.

3.1. Using cGERGM to probe network structure

One notable finding is the high degree of consistency we find in the activation motifs across all of our subjects and subnetworks. This consistency suggests that heightened clustering and reduced preferential attachment may be optimized for subnetworks to coordinate to give rise to specific cognitive processes. In practice, such an organization might resemble densely intraconnected communities that have built their own communication networks between themselves, and may be optimized for function-specific coordination between brain regions (Braunstein et al., 2007; Stam et al., 2014; Tewarie et al., 2015).

The present results further demonstrate how the cGERGM can be used to quantify the unique contributions of different influences of network structure. One natural question is to ask how cGERGM parameters correspond to existing descriptive metrics commonly used to quantify

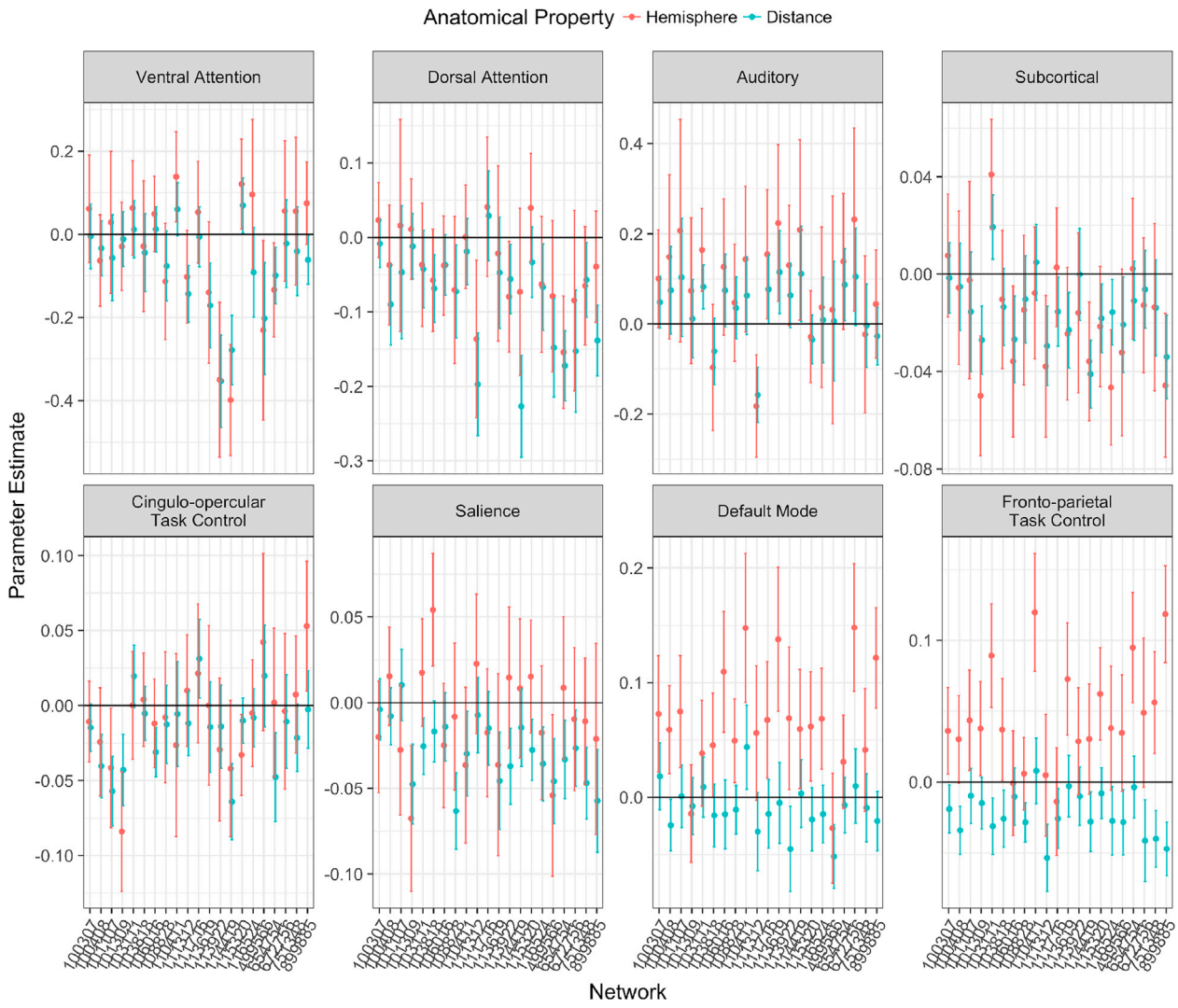

Fig. 8. Parameter estimates for the impact of hemisphere (red) and spatial distance (blue) on each of the eight subnetworks tested. Error bars correspond to 95% confidence intervals.

brain networks. While future research should substantiate the link between the cGERGM and existing techniques, we urge caution when comparing cGERGM parameters to descriptive statistics. For instance, it may seem natural to compare the triads parameter in the cGERGM of a network with a descriptive statistic like the number of triads or the clustering coefficient in the same network. However, the cGERGM parameter value quantifies the effect of the variable on the generation of the network, whereas descriptive statistics summarize features of a network. Furthermore, the parameter estimates take into account variability in the observed network as well as interactions among other variables in the model.

Additionally, two unanswered questions are whether these subnetwork properties – high clustering and low preferential attachment – are distinct from those of the whole brain, and whether they generalize to other levels of coarseness or granularity. Regarding the first question: one possibility is that the whole brain, due to the relative importance of integration, would show greater preferential attachment than we see at the subnetwork level. The sparse connections between subnetworks of the brain likely plays an important role in its overall topology, and it leaves open the possibility that the whole-brain may show markedly different network structure than that of the subnetworks. However, it is also possible that the whole-brain mirrors subnetwork structure – in other words, although the whole-brain makes use of hub regions, it may not show preferential attachment at the level of nodal organization. While the cGERGM is not yet equipped to run networks of the size required for full-brain investigations, we are currently working on extending the capabilities of cGERGM so that we can address these questions directly.

Second, whether our results generalize to more sparse or granular subnetwork partitions will similarly be more answerable in time. Studying organizational principles at different (spatial) scales has recently been recognized as crucial for a complete understanding of the structure of brain networks (Betzel and Bassett, 2017), and initial findings suggest both consistencies and differences when investigating network structure at different scales. Therefore, one possibility is that the same general structure appears across spatial resolution of subnetworks. For instance, past work has suggested that the brain demonstrates hierarchical modularity (Meunier et al., 2010) – the brain is divided into densely connected subnetworks, which themselves contain densely connected subnetworks. We might thus predict high clustering regardless of the level of granularity of the subnetworks. On the other hand, it is also possible that, as specifications become more granular, the operation of the subnetwork may more closely reflect the whole-brain in requiring integration across more specialized sub-communities. As we note above, it remains to be seen whether the whole-brain indeed shows network-level hub use (i.e., preferential attachment), but if so, we might

expect more granular subnetwork specifications to similarly show heightened (as opposed to decreased) preferential attachment. Finally, one strength of the cGERGM is that it can be adapted to hierarchical implementations, meaning future research may be able to simultaneously model a given network at multiple levels of resolution. We encourage future research using the cGERGM across multiple levels of spatial resolution.

3.2. The role of hubs

While previous investigations have emphasized the importance of hubs for network organization in the brain, the present work suggests that, at least for functional subnetworks, hub use is actually occurring less frequently than would be expected by chance. One possible explanation for this discrepancy may be the level of analysis that researchers studying hubs have focused on. In particular, past work has often looked at hubs at the level of the node – for instance, by identifying a node as a hub based on degree (Buckner et al., 2009; Cole et al., 2010; Andrews-Hanna et al., 2010), in which hubs are identified based on the number or strength of connections relative to other nodes (for alternative approaches, see Power et al., 2013). In contrast, our approach investigates the propensity of nodes across the network to be connected to more popular (i.e., hub) nodes – in other words, as a network property rather than as a property of a single node. This distinction may be important going forward: hub nodes are clearly important within the brain, but their presence does not necessarily imply the network is organized to use hubs.

3.3. The impact of distance

While initial investigations found strong relationships between Euclidean distance and the strength of functional connectivity (Honey et al., 2009), more recent evidence suggest those may be primarily driven by motion-related artifacts (Power et al., 2012, 2015). Indeed, an early version of the present investigation that did not take stringent approaches to deal with motion found much stronger effects of distance on connectivity. We note, however, that our results can be interpreted as weak evidence for an effect of distance on connectivity. Specifically, although the distance parameter was not consistently statistically significant for any subnetwork at an individual level, collapsing across participants revealed a clear pattern of distance influencing connection strength for every network except the auditory network. While much of the relationship between distance and connectivity may be due to motion artifacts, the present data provides some initial evidence that the distance–connectivity link may still exist. Future research should investigate this possibility using more comprehensive approaches to accounting for noise, such as making sure nodes are not closer than 20 mm apart, as these regions can have artificially inflated functional connectivity due to factors such as physics and vasculature (Power et al., 2011).

3.4. Using cGERGM to inform computational models of distributed cognition

One implication of the present study is to provide tools for neuroscientists to more deeply understand how collections of nodes interact in order to give rise to complex cognitive processes. In particular, by isolating the impact of specific activation motifs, researchers can test computational frameworks that attempt to understand cognitive processes via how brain regions organize (both in subnetworks and the whole brain). For instance, cGERGM allows for testing between models that emphasize either complex feedback loops, feedforward systems, or systems that integrate information at centralized hubs. Further, beyond understanding how different activation motifs describe connectivity, researchers can use cGERGM to test specific hypotheses about how different regions interact. Specifically, researchers can develop their own anatomical properties that reflect expected communication patterns between regions, and model those while accounting for other anatomical properties as well as activation motifs. Overall, we believe cGERGM opens the door for researchers developing models of specific cognitive phenomena to be informed and constrained by how communication is actually occurring across these regions, thus increasing the biological plausibility of these models. We invite future research using cGERGM in computational neuroscience.

Data and code availability statement

All data used in the present study were from the publicly available Human Connectome Project (HCP) S1200 release, available at https://www.humanconnectome.org/. We use the following subjects from the S1200 release: 100307, 100408, 101107, 101309, 103111, 103818, 106016, 108828, 110411, 111312, 111716, 113619, 113922, 114419, 115320, 116524, 499566, 654754, 672756, 751348, 899885. Data and scripts are available at the Open Science Framework, osf.io/3fmvj.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2019.03.036.

Appendix A. Technical Details on the cGERGM

We now provide mathematical details of the cGERGM. First, we explain the model in general and how to fit the model using maximum likelihood estimation. We next explicitly describe the models that were fitted in this manuscript thereby linking the mathematical notation with the variables utilized in our study.

The Model

Formally, the cGERGM is a joint probability distribution that characterizes the network and exogeneous structure of an $n \times n$ correlation matrix $\rho = (\rho_{i,j} : i < j \in 1,\ldots,n)$. Entries $\rho_{ij} \in [-1,1]$ and $\rho_{ij} = \rho_{ji}$ for all i and j. The primary difficulty of directly analyzing ρ lies in the fact that correlation matrices must be non-negative definite. Without accounting for this mathematical constraint, models can lead to inferential false positives due to artifacts in the data (Pourahmadi, 2011). Thus, for valid inference of ρ, we would like a model that ensures that estimates, say $\hat{\rho}$, are also non-negative definite. The cGERGM provides a family of models for a random matrix $\rho \in [-1,1]^{n \times n}$ that (i) upholds the non-negative definite constraint of ρ, (ii) describes the relational (network) structure of ρ, and (iii) describes the relationship between the entries of ρ and a collection of potentially useful predictors $z_1,\ldots,z_q$.

Let $m = n(n-1)/2$ be the total number of unique entries in ρ. The formulation of the cGERGM requires three steps, described next. First, the relational features of ρ are represented by a network on the unit interval $x \in [0,1]^m$. The user specifies a vector of features $\boldsymbol{h}(x)$ of length $p < m$ that describe the structural patterns of the graph x. Once the features have been specified, the bounded network x is modeled by the joint density function

$$f_X(x|\boldsymbol{\theta}) = \frac{\exp\left(\boldsymbol{\theta}^T \mathrm{h}(x)\right)}{\int_{[0,1]^m} \exp\left(\boldsymbol{\theta}^T \mathrm{h}(z)\right) dz}, \tag{A.1}$$

where $\boldsymbol{\theta}$ is a p - dimensional vector of parameters that quantify the effects of the network features $\boldsymbol{h}(x)$ on the likelihood of the graph x.

In the second step, the vector of unconstrained partial correlations $\phi = (\phi_{ij} : i < j \in [n]) \in [-1,1]^m$ is used to capture the exogeneous structure of ρ. The matrix x is mapped to the partial correlations ϕ via the transformation $x_{ij} = T_{ij}((\phi_{ij}+1)/2 \big| \mu_{ij}, \alpha)$, where $T_{ij}(\cdot \big| \mu_{ij}, \alpha)$ is the cumulative distribution function of a Beta distribution with mean μ_{ij} and scale parameter α. Let $T : \mathbb{R}^m \to [0,1]^m$ be defined as the m-dimensional vector $T(\phi|\boldsymbol{\mu},\alpha) = (T_{ij}(\phi_{ij} \big| \mu_{ij}, \alpha), i < j \in [n])$. Then ϕ is a random vector from the probability density function $f_\Phi(\phi|\boldsymbol{\theta},\boldsymbol{\mu},\alpha)$, given by

$$f_\Phi(\phi|\boldsymbol{\theta},\boldsymbol{\mu},\alpha) = f_X\left(T\left(\frac{\phi+1}{2}\right)\Big|\boldsymbol{\theta}\right)\prod_{ij} t_{ij}\left(\frac{\phi+1}{2}\Big|\mu_{ij},\alpha\right)$$

$$t_{ij}\left(w\big|\mu_{ij},\alpha\right) = \frac{\Gamma(\alpha)\left(1-w_{ij}\right)^{\left(1-\mu_{ij}\right)\alpha-1}}{\Gamma\left(\mu_{ij}\alpha\right)\Gamma\left(\left(1-\mu_{ij}\right)\alpha\right)} w_{ij}^{\mu_{ij}\alpha-1}, \tag{A.2}$$

where $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx$ is the gamma function defined for $t > 1$ and $w_{ij} \in (-1,1)$. A notable property of (A.2) is that when the partial correlation network ϕ does not contain any network structure, i.e. when $\boldsymbol{\theta} = 0$, then the m components of $(\phi+1)/2$ are independent samples from a Beta(μ_{ij},α) distribution, where $\mu_{ij} \in (0,1)$ is the expected value of $(\phi_{ij}+1)/2$ and α is the precision of the distribution.

Let z_ℓ be a vector of dyadic observations $z_\ell = (z_{ij}(\ell) : i < j \in [n])$ for $\ell = 1,\ldots,q$. The effect of each predictor on ϕ is quantified using a generalized linear model for the mean $\boldsymbol{\mu}$ of ϕ:

$$\mathrm{logit}\left(\mu_{ij}\right) = \beta_0 + \sum_{\ell=1}^{q} \beta_\ell z_{ij}(\ell) \tag{A.3}$$

The regression model in (A.3), together with the Beta density $t_{ij}(\cdot|\mu_{ij},\alpha)$ is well-studied and generally referred to as Beta regression (Cribari-Neto and Zeileis, 2010). The final step in defining the cGERGM relies on the relationship between the partial correlation matrix ϕ and its associated correlation matrix ρ for the collection of n brain regions. Let $\phi_{j\,j+k}$ denote the partial correlation between region j and $j+k$ given regions $j+1,\ldots,j+k-1$ for $k \geq 2$. A well-known result in multivariate statistics is that there exists a one-to-one mapping between ρ and ϕ (Anderson, 1962). For brevity, we do not write out the details of this transformation; and simply represent this mapping from ρ to ϕ by $\rho = g(\phi)$. Then the Jacobian of this transformation, as shown by Joe (2006), is given by

$$\left|J(c)\right| = \left[\prod_{i=1}^{n-1}\left(1-c_{i\,i+1}^2\right)^{n-1}\prod_{k=2}^{n-2}\prod_{i=1}^{n-k}\left(1-c_{i\,i+k}^2\right)^{n-k-1}\right]^{-1/2}$$

Applying the inverse probability transform to the density in (A.2), we can generate a *non-negative definite* correlation network ρ from the density:

$$f_R(\rho|\boldsymbol{\theta},\boldsymbol{\mu},\alpha) = f_\Phi\left(g^{-1}(\rho)\big|\boldsymbol{\theta},\boldsymbol{\mu},\alpha\right)\left|J\left(g^{-1}(\rho)\right)\right|,$$

In summary, for fixed parameters $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_p)$, $\boldsymbol{\beta} = (\beta_0,\ldots,\beta_q)$, and $\alpha > 0$ the cGERGM model for a correlation network ρ on n brain regions is described by the following generative process

$$\rho \sim f_R(\,\cdot|\boldsymbol{\theta},\boldsymbol{\mu},\alpha) \tag{A.4}$$

$$\mathrm{logit}\left(\mu_{ij}\right) = \beta_0 + \beta_1 z_{ij}(1) + \ldots + \beta_q z_{ij}(q).$$

Maximum Likelihood Estimation of the cGERGM

Given an observed correlation network ρ, a collection of exogeneous predictors $z_1,\ldots,z_q$ and a p-dimensional vector of network summary statistics $\boldsymbol{h}(x)$, one can fit the cGERGM in model (A.4) via maximum likelihood estimation of the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$. To do so, the observed correlation network ρ is first transformed to its representative partial correlation matrix ϕ via the one-to-one transformation $\phi = g^{-1}(\rho)$ described above. The partial correlations are then scaled to the unit interval. Let ϕ^* represent these scaled values. Then, $\phi_{ij}^* = \frac{1}{2}(\phi_{ij}+1)$. Given ϕ^* the maximum likelihood estimators (MLEs) of the unknown parameters $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, denoted by $\hat{\theta}$ and $\hat{\beta}$, respectively, are the values that maximize the joint log likelihood arising from (A.2) and (A.3):

$$\ell(\boldsymbol{\theta},\boldsymbol{\beta} \mid \phi^*) = \boldsymbol{\theta}^T \mathrm{h}(T(\phi^*|\boldsymbol{\beta})) - \log C(\boldsymbol{\theta}) + \sum_{ij} \log t_{ij}(\phi^*|\boldsymbol{\beta}), \tag{A.5}$$

where

$$C(\boldsymbol{\theta}) = \int_{[0,1]^m} \exp(\boldsymbol{\theta}^T \mathrm{h}(z)) dz.$$

In equation (A.5), note that $t_{ij}(\phi^*|\boldsymbol{\beta})$ is the Beta probability density function evaluated at the scaled value ϕ^*_{ij} given parameters μ_{ij} and α (expressed in (A), and $T(\phi^*|\boldsymbol{\beta})$ is the joint cumulative distribution of a Beta distribution with mean μ_{ij}. Here, α is a nuisance parameter that is a function of μ_{ij}, and μ_{ij} is related to the regression coefficients $\boldsymbol{\beta}$ through the Beta regression model in (A.3).

The maximization of (A.5) can be achieved through alternate maximization of $\boldsymbol{\beta}|\boldsymbol{\theta}$ and $\boldsymbol{\theta}|\boldsymbol{\beta}$. One first provides an initial value $\boldsymbol{\theta}^{(0)}$. These values are initialized following the method of Hummel et al. (2012), where an initial guess is refined by slowly changing the objective to be closer to the observed network, allowing the optimizer to function more effectively on the highly uneven likelihood surface. Then the maximum likelihood estimators are calculated by iterating between the following two steps until convergence.

Maximum Likelihood Estimation of the cGERGM

Given: Scaled partial correlations ϕ^*; network statistics $\mathbf{h}(x)$; exogenous regressors $z_1, \ldots, z_q$; initial value $\boldsymbol{\theta}^{(0)}$.

For $r \geq 1$, iterate until convergence:

1. Given $\boldsymbol{\theta}^{(r)}$, estimate $\boldsymbol{\beta}^{(r)}|\boldsymbol{\theta}^{(r)}$:

$$\boldsymbol{\beta}^{(r)} = \text{argmax}_{\beta}\left(\boldsymbol{\theta}^{(r)}\mathbf{h}(T(\phi^*|\boldsymbol{\beta})) + \sum_{ij}\log t_{ij}(\phi^*|\boldsymbol{\beta})\right). \tag{A.6}$$

2. Set $\widehat{y} = T(\phi^*|\boldsymbol{\beta}^{(r)})$. Then estimate $\boldsymbol{\theta}^{(r+1)}|\boldsymbol{\beta}^{(r)}$:

$$\boldsymbol{\theta}^{(r+1)} = \text{argmax}_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}^T\text{h}(\widehat{y}) - \log C(\boldsymbol{\theta})\right). \tag{A.7}$$

The maximization in (A.6) is accomplished numerically using gradient descent, though any hill climbing algorithm will suffice. The maximization in (A.7) is challenging due to the computational intractability of the normalization constant $C(\boldsymbol{\theta})$. The Metropolis-Hastings (M-H) simulation procedure from Wilson et al. (2017) is used to approximate this constant. For iteration r, the M-H procedure simulates networks based on the current parameter estimates $(\boldsymbol{\theta}^{(r)}, \boldsymbol{\beta}^{(r)})$ via an acceptance-rejection algorithm based on a truncated normal proposal distribution. Networks continue to be simulated until network statistics of the simulated networks is detected are deemed stationary. The constant $C(\boldsymbol{\theta})$ is approximated by averaging over the simulated networks.

Once the parameter updates from (A.6) and (A.7) converge within an acceptable tolerance, the maximum likelihood estimators θ and β are set to the last iteration's values. Standard errors are determined through M-H simulation of correlation networks under the fitted values. Furthermore, correlation networks can be readily simulated from the joint distribution in (A.4) using M-H given the fitted parameters θ and β. Maximum likelihood estimation and inference of the cGERGM can be done in the publicly available R package *GERGM* (Denny et al., 2016).

Model Specification in this Study

For each adult that was analyzed, we fit the cGERGM to the correlation matrix arising from resting state fMRI. In the context of our mathematical notation above, we fit the cGERGM separately to 21 adult correlation matrices $\rho_1, \ldots \rho_{21}$. For each adult, the cGERGM was fit with five activation motifs, three of which were endogenous network-based predictors, and the remaining two were exogeneous predictors. The endogeneous predictors, $h_1(x)$, $h_2(x)$, and $h_3(x)$, were the total density of edges, total density of two stars, and the total density of triads, respectively. The exogeneous predictors, z_1, z_2, were the Euclidean distance between pairs of regions and the hemisphere location of each region, respectively. Fitting the cGERGM with these motifs specified provided estimators for their parameters in the model, $\theta_1, \theta_2, \theta_3, \beta_1$, and β_2, which are shown in Figs. 6–8.

References

Anderson, T.W., 1962. An Introduction to Multivariate Statistical Analysis. Wiley, New York.

Andrews-Hanna, J.R., 2012. The brains default network and its adaptive role in internal mentation. Neuroscientist 18, 251–270.

Andrews-Hanna, J.R., Reidler, J.S., Sepulcre, J., Poulin, R., Buckner, R.L., 2010. Functional-anatomic fractionation of the brain's default network. Neuron 65, 550–562.

Bassett, D.S., Bullmore, E.T., 2017. Small-world brain networks revisited. Neuroscientist 23, 499–516.

Bassett, D.S., Sporns, O., 2017. Network neuroscience. Nat. Neurosci. 20, 353–364.

Bellana, B., Liu, Z., Anderson, J.A., Moscovitch, M., Grady, C.L., 2016a. Laterality effects in functional connectivity of the angular gyrus during rest and episodic retrieval. Neuropsychologia 80, 24–34.

Bellana, B., Liu, Z., Diamond, N., Grady, C.L., Moscovitch, M., 2016b. Similarities and differences in the default mode network across rest, retrieval, and future imagining. Hum. Brain Mapp. 38, 1155–1171.

Betzel, R.F., Bassett, D.S., 2017. Multi-scale brain networks. Neuroimage 160, 73–83.

Braunstein, L.A., Wu, Z., Chen, Y., Buldyrev, S.V., Kalisky, T., Sreenivasan, S., Cohen, R., Lopez, E., Havlin, S., Stanley, H.E., 2007. Optimal path and minimal spanning trees in random weighted networks. Int. J. Bifurc. Chaos 17, 2215–2255.

Bressler, S.L., Menon, V., 2010. Large-scale brain networks in cognition: emerging methods and principles. Trends Cognit. Sci. 14, 277–290.

Buckner, R.L., Sepulcre, J., Talukdar, T., Krienen, F.M., Liu, H., Hedden, T., Andrews-Hanna, J.R., Sperling, R.A., Johnson, K.A., 2009. Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to alzheimer's disease. J. Neurosci. 29, 1860–1873.

Bullmore, E., Sporns, O., 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. Nat. Rev. Neurosci. 10, 186–198.

Bullmore, E., Sporns, O., 2012. The economy of brain network organization. Nat. Rev. Neurosci. 13, 336–349.

Ciric, R., Wolf, D.H., Power, J.D., Roalf, D.R., Baum, G.L., Ruparel, K., Shinohara, R.T., Elliott, M.A., Eickhoff, S.B., Davatzikos, C., et al., 2017. Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. Neuroimage 154, 174–187.

Cole, M.W., Pathak, S., Schneider, W., 2010. Identifying the brain's most globally connected regions. Neuroimage 49, 3132–3148.

Cribari-Neto, F., Zeileis, A., 2010. Beta regression in r. J. Stat. Softw. 34.

Denny, M., Wilson, J., Cranmer, S., Desmarais, B., Bhamidi, S., 2016. Gergm: Estimation And Fit Diagnostics For Generalized Exponential Random Graph Models. R Package Version 0.10. 0. Available at: (Accessed 29 March 2017) https://CRAN.R-project.org/package=GERGM.

Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., Consortium, W.-M.H., et al., 2013. The Wu-minn human connectome project: an overview. Neuroimage 80, 62–79.

Feinberg, D.A., Moeller, S., Smith, S.M., Auerbach, E., Ramanna, S., Glasser, M.F., Miller, K.L., Ugurbil, K., Yacoub, E., 2010. Multiplexed echo planar imaging for sub-second whole brain fmri and fast diffusion imaging. PLoS One 5 e15710.

Fruchterman, T.M., Reingold, E.M., 1991. Graph drawing by forcedirected placement. Software Pract. Ex. 21, 1129–1164.

Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., et al., 2013. The minimal preprocessing pipelines for the human connectome project. Neuroimage 80, 105–124.

van den Heuvel, M.P., Sporns, O., 2011. Rich-club organization of the human connectome. J. Neurosci. 31, 15775–15786.

van den Heuvel, M.P., Sporns, O., 2013. Network hubs in the human brain. Trends Cognit. Sci. 17, 683–696.

Honey, C., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J.-P., Meuli, R., Hagmann, P., 2009. Predicting human resting-state functional connectivity from structural connectivity. Proc. Natl. Acad. Sci. Unit. States Am. 106, 2035–2040.
Hummel, R.M., Hunter, D.R., Handcock, M.S., 2012. Improving simulation-based algorithms for fitting ERGMs. J. Comput. Graph. Stat. 21, 920–939.
Hunter, D.R., Goodreau, S.M., Handcock, M.S., 2008. Goodness of fit of social network models. J. Am. Stat. Assoc. 103, 248–258.
Joe, H., 2006. Generating random correlation matrices based on partial correlations. J. Multivar. Anal. 97, 2177–2189.
Medaglia, J.D., Lynall, M.-E., Bassett, D.S., 2015. Cognitive network neuroscience. J. Cogn. Neurosci. 27, 1471–1491.
Meunier, D., Lambiotte, R., Bullmore, E.T., 2010. Modular and hierarchically modular organization of brain networks. Front. Neurosci. 4, 200.
Moeller, S., Yacoub, E., Olman, C.A., Auerbach, E., Strupp, J., Harel, N., Uğurbil, K., 2010. Multiband multislice ge-epi at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fmri. Magn. Reson. Med. 63, 1144–1153.
Pourahmadi, M., 2011. Covariance estimation: the glm and regularization perspectives. Stat. Sci. 369–387.
Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., Vogel, A.C., Laumann, T.O., Miezin, F.M., Schlaggar, B.L., et al., 2011. Functional network organization of the human brain. Neuron 72, 665–678.
Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2012. Spurious but systematic correlations in functional connectivity mri networks arise from subject motion. Neuroimage 59, 2142–2154.
Power, J.D., Schlaggar, B.L., Lessov-Schlaggar, C.N., Petersen, S.E., 2013. Evidence for hubs in human functional brain networks. Neuron 79, 798–813.
Power, J.D., Schlaggar, B.L., Petersen, S.E., 2015. Recent progress and outstanding issues in motion correction in resting state fmri. Neuroimage 105, 536–551.
Rubinov, M., Sporns, O., 2010. Complex network measures of brain connectivity: uses and interpretations. Neuroimage 52, 1059–1069.
Salimi-Khorshidi, G., Douaud, G., Beckmann, C.F., Glasser, M.F., Griffanti, L., Smith, S.M., 2014. Automatic denoising of functional mri data: combining independent component analysis and hierarchical fusion of classifiers. Neuroimage 90, 449–468.
Satterthwaite, T.D., Wolf, D.H., Loughead, J., Ruparel, K., Elliott, M.A., Hakonarson, H., Gur, R.C., Gur, R.E., 2012. Impact of in-scanner head motion on multiple measures of functional connectivity: relevance for studies of neurodevelopment in youth. Neuroimage 60, 623–632.
Setsompop, K., Gagoski, B.A., Polimeni, J.R., Witzel, T., Wedeen, V.J., Wald, L.L., 2012. Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. Magn. Reson. Med. 67, 1210–1224.
Sizemore, A., Giusti, C., Bassett, D.S., 2016. Classification of weighted networks through mesoscale homological features. J. Complex Netw. 5, 245–273.
Smith, S.M., Miller, K.L., Salimi-Khorshidi, G., Webster, M., Beckmann, C.F., Nichols, T.E., Ramsey, J.D., Woolrich, M.W., 2011. Network modelling methods for fmri. Neuroimage 54, 875–891.
Smith, S.M., Beckmann, C.F., Andersson, J., Auerbach, E.J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D.A., Griffanti, L., Harms, M.P., et al., 2013. Resting-state fmri in the human connectome project. Neuroimage 80, 144–168.
Sporns, O., 2011. Networks of the Brain. MIT press.
Stam, C., Tewarie, P., Van Dellen, E., Van Straaten, E., Hillebrand, A., Van Mieghem, P., 2014. The trees and the forest: characterization of complex brain networks with minimum spanning trees. Int. J. Psychophysiol. 92, 129–138.
Stillman, P.E., Wilson, J.D., Denny, M.J., Desmarais, B.A., Bhamidi, S., Cranmer, S.J., Lu, Z.-L., 2017. Statistical modeling of the default mode brain network reveals a segregated highway structure. Sci. Rep. 7, 11694.
Tewarie, P., van Dellen, E., Hillebrand, A., Stam, C.J., 2015. The minimum spanning tree: an unbiased method for brain network analysis. Neuroimage 104, 177–188.
Wilson, J.D., Denny, M.J., Bhamidi, S., Cranmer, S.J., Desmarais, B.A., 2017. Stochastic weighted graphs: flexible model specification and simulation. Soc. Network. 49, 37–47.
Xu, J., Moeller, S., Strupp, J., Auerbach, E., Chen, L., Feinberg, D., Ugurbil, K., Yacoub, E., 2012. Highly accelerated whole brain imaging using aligned-blipped-controlled-aliasing multiband epi. In: Proceedings of the 20th Annual Meeting of ISMRM, vol 2306.
Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, L., Polimeni, J.R., et al., 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. 106, 1125–1165.