ASSESSING STATISTICAL SIGNIFICANCE WHEN PARTITIONING LARGE-SCALE BRAIN NETWORKS

Yu-Teng Chang[⋆] *Dimitrios Pantazis*[†] *Richard M. Leahy*[⋆]

[⋆] Signal and Image Processing Institute, University of Southern California, Los Angeles, CA 90089
[†] McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139

ABSTRACT

Multivariate analysis of structural and functional brain imaging data can be used to produce network models of interaction or similarity between different brain structures. Graph partitioning methods can then be used to identify distinct subnetworks that may provide insight into the organization of the human brain. Although several efficient partitioning algorithms have been proposed, and their properties studied thoroughly, there has been limited work addressing the statistical significance of the resulting partitions. We present a new method to estimate the statistical significance of a network structure based on modularity. We derive a numerical approximation of the distribution of modularity on random graphs, and use this distribution to calculate a threshold that controls the type I error rate in partitioning graphs. We demonstrate the technique in application to brain subnetworks identified from diffusion-based fiber tracking data and from resting state fMRI data.

Index Terms— community structure, modularity, graph partitioning, significance testing

1. INTRODUCTION

The study of community structure of graphs has revealed interesting patterns in the structure and function of the human brain [1] with features that vary across neurological diseases, aging, and cognitive processes [2, 3]. A large number of methods have been proposed to identify natural divisions of networks into groups. Perhaps the most popular is modularity [4], which compares the network against a null model and favors within module connections when edges are stronger than their expected values. Divisions that increase modularity are preferred because they lead to modules with high community structure.

Despite the popularity of modularity methods, the statistical significance of network partitions has received less attention. Random networks can exhibit high modularity because of incidental concentration of edges, even though they have no underlying organizational structure [5]. This is even more evident in large networks where the number of possible divisions increases rapidly with the network size [6]. Therefore, significant divisions of a network should have higher modularity than random graphs [5, 7]. An alternative approach was proposed in [6] to evaluate the significance of a partition: a community structure should be robust against small perturbations of the network.

In this paper, we propose a statistical procedure to test the significance of a community structure based on its modularity value. As a surrogate of modularity, we use the largest eigenvalue of the difference between the affinity matrices of the network and its null model. Based on the previous work on null models [8], we show that the distribution of the largest eigenvalue can be well approximated with a Gamma distribution. We derive an empirical formula for the parameters of the Gamma distribution with respect to the size of the network and the variance of its edges. Based on this distribution we compute a p-value for the community structure, which can be used as a threshold criterion when partitioning a graph. We demonstrate our method with simulated and real brain networks.

2. METHOD

2.1. Modularity and Null Models

Consider an undirected weighted graph with N nodes, adjacency matrix $\mathbf{A}$, and connection strength between nodes i and j denoted as A_{ij}. As an example, the nodes of the network may represent regions of parcelated cerebral cortex with the connection strengths equal to the correlation or partial correlation between them as computed from resting state fMRI data. Alternatively, the connections may reflect the degree of connectivity between the two cortical regions as determined by fiber-tracking from diffusion tensor data. The degree vector $\mathbf{k}$ has elements $k_i = \sum_{j=1}^{N} A_{ij}$, indicating the sum of all edge strengths associated with node i. Define the edge vector $\mathbf{x}$ containing all possible edges in $\mathbf{A}$, according to the mapping $A_{ij} = x_l$, where $l = \frac{(2N-j)(j-1)}{2} + (i-j)$ ($\forall 1 \leq j < i \leq N$). In [8] we derived the expected network conditioned on the degree vector $\mathbf{k}$:

$$E(\mathbf{x}|\mathbf{k}) = \boldsymbol{\mu}_{\mathbf{x}|\mathbf{k}} = \boldsymbol{\mu}_{\mathbf{x}} + \boldsymbol{\Sigma}_{\mathbf{xk}}\boldsymbol{\Sigma}_{\mathbf{k}}^{-1}(\mathbf{k} - \boldsymbol{\mu}_{\mathbf{k}}) \qquad (1)$$

This work supported by the National Institutes of Health under grants 5R01EB000473, P41 RR013642 and R01 EB002010, and the National Science Foundation under grant BCS-1134780.

with the conditional covariance matrix:

$$\mathbf{\Sigma_{x|k}} = \mathbf{\Sigma_x} - \mathbf{\Sigma_{xk}}\mathbf{\Sigma_k}^{-1}\mathbf{\Sigma_{kx}} \quad (2)$$

As indicated in [8], the conditional expected network satisfies the configuration model criteria (same number of nodes and equal degrees), while at the same time preserving network topology and allowing negative connections. Furthermore, it is in general the best linear unbiased estimator given the node degrees.

Define $\mathbf{B}$ to be the difference matrix between the original and the expected null network, with elements $B_{ij} = A_{ij} - E(A_{ij}|\mathbf{k}) = A_{ij} - E(x_l|\mathbf{k})$. Modularity Q [4] is then computed as:

$$Q = \frac{1}{2m}\sum_{i,j} B_{ij}\delta(C_i, C_j) \quad (3)$$

where C_i indicates group membership of node i, $\delta(C_i, C_j) = 1$ only when node i and j are clustered within the same group, and m is the total sum of the weights of all edges in the network. The best bi-partition of the graph that maximizes modularity can be described by an indicator vector $\mathbf{s}$, having values 1 and -1 for the two subnetworks, respectively:

$$\hat{\mathbf{s}} = \max_{\mathbf{s}} Q = \max_{\mathbf{s}} \{\mathbf{s}^T\mathbf{B}\mathbf{s}\} \quad (4)$$

Spectral graph theory solves equation (4) in the continuous domain while constraining the norm of $\mathbf{s}$. Maximization is achieved by selecting the eigenvector corresponding to the largest eigenvalue λ_1 of $\mathbf{B}$. The elements of $\mathbf{s}$ are then discretized to $-1, 1$ by setting a zero threshold. Because of this discretization, further fine tuning is necessary to locally maximize Q, which can be done using, for example, the Kernighan-Lin algorithm. Given that the final maximum value of modularity Q is close to λ_1, we can use the latter as a surrogate for modularity.

2.2. Distribution of Largest Eigenvalue

We consider a graph partition statistically significant if modularity Q is substantially higher than the one achieved by partitioning random graphs. Therefore, we seek the distribution of modularity for random graphs; hypothesis testing would then proceed by selecting a 5% threshold in the right tail of this distribution. Instead of using modularity, we use the largest eigenvalue of $\mathbf{B}$ as a surrogate. Here we assume the network edges follow a jointly Gaussian distribution with mean $\mu\mathbf{1}$ and variance $\sigma^2\mathbf{I}$. Although the largest eigenvalue distribution of simple Gaussian random networks (GRN) has been studied, there is no closed form solution for the case of correlated edges caused by conditioning on $\mathbf{k}$. We therefore follow a Monte Carlo approach.

For a network size $N = 20$, we generate 100 Gaussian random networks $\mathbf{A}$ for given values of μ and σ^2. For each network, instead of randomly permuting network edges, we generate 10^6 random networks $\mathbf{A}^{\text{Random}}$, by sampling from equations (1,2), and then compute the largest eigenvalue of $\mathbf{B}^{\text{Random}} = \mathbf{A}^{\text{Random}} - E(\mathbf{A}^{\text{Random}}|\mathbf{k})$. The largest eigenvalue distribution is shown in Figure 1 for several values of mean μ and variance σ^2, and it is evident that it only depends on the value of σ^2.

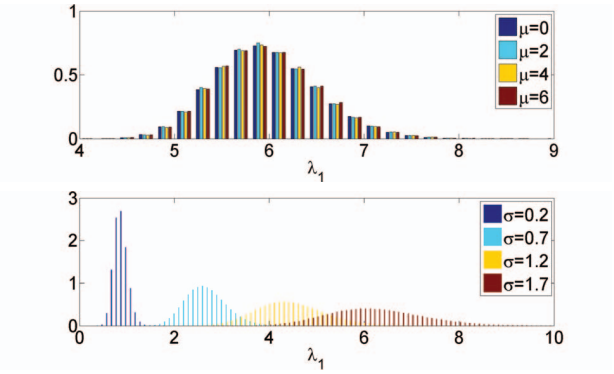


Fig. 1. Top: Distribution of the largest eigenvalue of $\mathbf{B}$ for several values of mean μ and fixed $\sigma = 0.8$. Bottom: Distribution of the largest eigenvalue of $\mathbf{B}$ for several values of standard deviation σ.

2.3. Approximation with Gamma Distribution Family

The empirical distribution of the largest eigenvalue in Figure 1 is skewed to the left for all values of σ^2. This distribution can be accurately approximated by a Gamma distribution:

$$f(x|\alpha, \beta) = \frac{x^{\alpha-1}e^{\frac{x}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)} \quad (5)$$

where α determines the shape and β the scale of the distribution. $\Gamma(\alpha)$ is the Gamma function with parameter α. An example of the approximation is shown in Figure 2.

Therefore, the empirical distribution of the largest eigenvalue can be well represented by a pair of parameters (α, β). These parameters are functions of the network size N and variance of the edges σ^2, and in the next section we identify the form of this functional relationship.

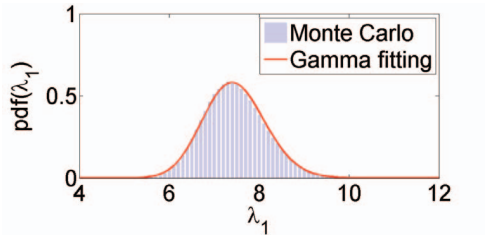


Fig. 2. Monte Carlo distribution of λ_1 for $\sigma = 1$ and its best fit with a Gamma distribution. The maximum likelihood estimation of the Gamma distribution parameters was $\hat{\alpha} = 117.99$ and $\hat{\beta} = 0.06321$ with estimator variance $\sigma^2_{\hat{\alpha}} = 0.0277$ and $\sigma^2_{\hat{\beta}} = 8 \times 10^{-9}$

2.4. Fitting with different network sizes

We repeated the above Monte Carlo procedure and estimation of the Gamma distribution parameters (α, β) for different values of network size N and edge standard deviation σ. Figure 3 shows that α does not depend on σ, whereas β linearly increases with σ. The network size N does not change the nature of these relationships.

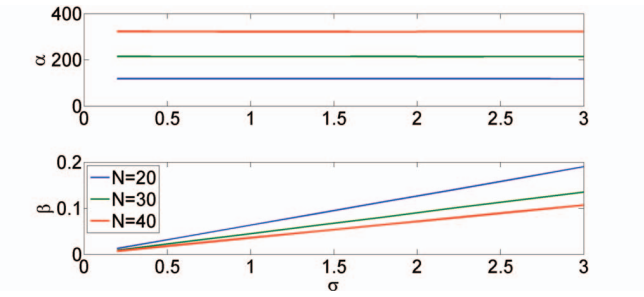


Fig. 3. Fitted parameters $\hat{\alpha}$ and $\hat{\beta}$ in the Gamma distribution as a function of standard deviation σ for network size $N = 20$, $N = 30$, and $N = 40$.

Figure 4 plots the value of α and the ratio β/σ for different values of network size N. Simple curve fitting methods produced the following formula for the parameters of the Gamma distribution:

$$\hat{\alpha}(N) = 2.459N^{1.335} - 16.453 \tag{6}$$

$$\hat{\beta}(N, \sigma) = \sigma\big(0.84N^{-0.87} + 0.0015\big) \tag{7}$$

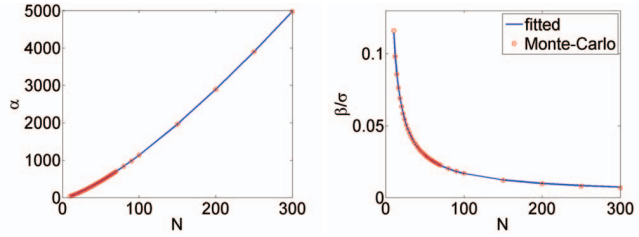


Fig. 4. Left: $\hat{\alpha}$ as a function of network size N. Right: ratio of $\hat{\beta}$ with respect to σ as a function of network size N.

For every bi-partition of a graph, we can determine the significance of the cut by comparing the largest eigenvalue of $\mathbf{B}$ against the distribution of eigenvalues of $\mathbf{B}^{\text{Random}}$. So rather than accepting a partition if we find an increase in modularity, instead we test whether the increase is statistically significantly above a 5% threshold. The procedure is given in the following steps. For a network $\mathbf{A}$, estimate edge variance σ^2. Then apply spectral graph partitioning and compute the largest eigenvalue of matrix $\mathbf{B}$ [8]. Use equation (6) and (7) to estimate the Gamma distribution parameters and then compare the largest eigenvalue against the Gamma distribution and accept the partition if it is above the 5% threshold in the right tail of the distribution.

3. RESULTS

To test whether the above formula holds for all networks irrespective of their community structure, we generated a number of different networks whose edges were i.i.d Gaussian random variables with mean 4 and variance 1. For each network we simulated two clusters with variable size $N_1 \geq N_2$, such that $N_1 + N_2 = N$, where $N = 20$ is the total number of nodes. This community structure was enforced by randomly allocating the stronger values of the Gaussian distribution as intra-cluster edges and the weaker values as inter-cluster edges. In all cases, we repeated the procedure in Section 2 to generate random graphs $\mathbf{A}^{\text{Random}}$ and re-estimate the parameters of the Gamma distribution. The results are shown in Figure 5 and are in close agreement with equations (6) and (7). This result agrees with our expectations; irrespective of the community structure of the original networks, $\mathbf{A}^{\text{Random}}$ is a randomized network with no community structure to affect the distribution of the largest eigenvalue.

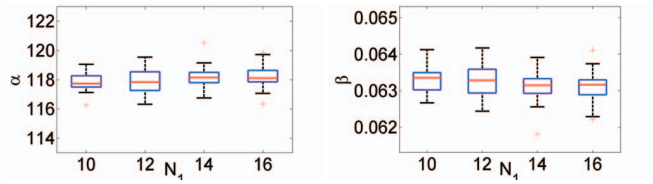


Fig. 5. Estimation of Gamma distribution parameters when the original networks used to create $\mathbf{A}^{\text{Random}}$ networks have a community structure of two clusters with N_1 and N_2 nodes such that $N_1 + N_2 = 20$. For each N_1 value, we generated 100 original networks which further produced 10^4 random networks each. The Gamma parameters estimated from the histogram of those random networks are in close agreement with those predicted by equations (6) and (7).

Figure 6a shows a Gaussian network with mean 1, variance 4, and a community structure of two 10-node clusters. To blur its community structure, we randomly permuted a portion of within and between edges producing the network on Figure 6b. The black lines in the adjacency matrices represent the clustering results by modularity-based partitioning. The computed modularity values are $Q_a = 0.098$ and $Q_b = 0.038$, indicating that the original unperturbed network has the stronger community structure. Our statistical procedure produced p-values $p_a \approx 0$ and $p_b = 0.73$, indicating only the first network has a significant community structure.

We applied significance-based modularity partitioning to the structural brain network reported in [9]. The network consists of 66 nodes representing FreeSurfer parcellated cortical regions and edges representing the neuronal fiber densities between pairs of regions. In [8] we reported partition results of this network, and using the method described above we confirm that all bi-partitions leading to 5 clusters are significant. The first bipartition, separating the two hemispheres

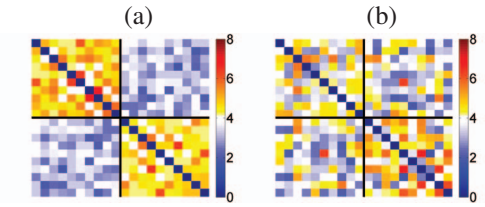


Fig. 6. The bi-partitioning results for two Gaussian networks; (a) shows a clear network structure while (b) is less structured.

(medial subnetwork goes with the right hemisphere), had $p \approx 0$, and the subsequent bipartitions were $p = 1.33 \times 10^{-6}$, $p = 2.17 \times 10^{-9}$, and $p = 8.96 \times 10^{-5}$. The structural data reveals subnetworks that largely follow the classical lobe-based partitioning of cerebral cortex.

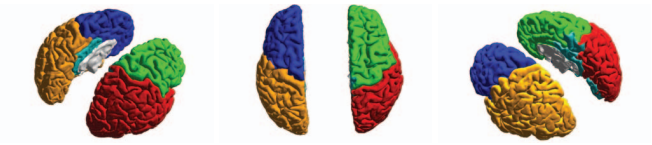


Fig. 7. Modularity-based partitioning of the structural brain network in [9]. Each color represents a different subnetwork. All clusters are statistically significant.

We also investigated subnetwork partitioning using networks extracted from resting state fMRI data [10]. The data consists of 191 subjects from the Beijing data set of the 1000 Functional Connectomes Project in NITRC [11]. The 96 nodes represent ROIs defined on the Harvard-Oxford atlas and edges indicate the correlation coefficient between fMRI signals for each pair of ROIs. Correlations were computed after bandpass filtering in the range 0.005-0.1Hz. We detected 3 clusters: the default mode network, a motor-sensory subnetwork, and a visual-related subnetwork. All bipartitions had p-values $p \approx 0$. In contrast to the structural subnetworks, the functional subnetworks involve larger scale interactions and follow well known functional subdivisions.

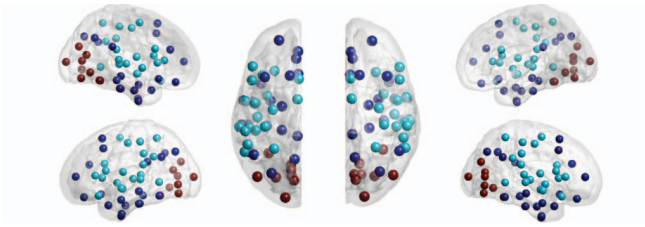


Fig. 8. Partition results of a resting state fMRI network. All 3 clusters are significant.

Finally, we partitioned the widely-studied Karate Club network in [12] into four clusters, as in Figure 9. Only the first bipartition (separating groups 1,2 against 3,4) is significant with p-value $p = 2.12 \times 10^{-12}$. This bipartition exactly predicts the subsequent split of the Karate club into 2 groups. The remaining two bipartitions are insignificant: $p = 0.067$ for Groups 1 - 2 and $p = 0.486$ for Groups 3 - 4.

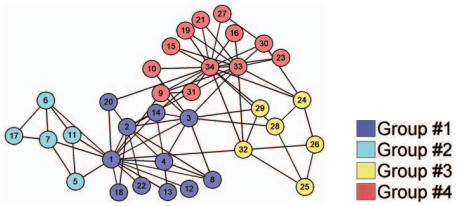


Fig. 9. Partition of the Karate Club network [12].

4. CONCLUSIONS

Even though graph partition methods are becoming increasingly popular, assessment of the statistical significance of the resulting partitions has drawn little attention. We believe this will change in the future in the same way that statistical thresholding procedures are paramount in analyzing brain imaging studies.

Using Monte Carlo procedures, we have derived equations that can control the type I error rate in bipartitions of graphs. It is therefore no longer necessary to perform computationally intensive permutations of edges and randomization of graphs every time a new graph is to be partitioned. The Gamma distribution provides an excellent fit to the empirical data so that a parametric approach based on these distributions should be sufficient for assessing subnetwork significance.

5. REFERENCES

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

[2] D. Meunier, S. Achard, A. Morcom, and E. Bullmore, "Age-related changes in modular organization of human brain functional networks," *Neuroimage*, vol. 4, pp. 715–723, 2009.

[3] A.F. Alexander-Bloch, N. Gogtay, D. Meunier, R. Birn, L. Clasen, F. Lalonde, R. Lenroot, J. Giedd, and E. Bullmore, "Disrupted modularity and local connectivity of brain functional networks in childhood-onset schizophrenia.," *Frontier Neuroinformatics*, vol. 4, pp. doi:10.3389, 2009.

[4] M.E.J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E*, vol. 69, no. 2, pp. 026113, Feb 2004.

[5] R. Guimerá, M. Sales-Pardo, and L.A.N. Amaral, "Modularity from fluctuations in random graphs and complex networks," *Phys. Rev. E*, vol. 70, no. 2, pp. 025101, 2004.

[6] B. Karrer, E. Levina, and M.E.J. Newman, "Robustness of community structure in networks," *Phys. Rev. E*, vol. 77, no. 4, pp. 046119, 2008.

[7] J. Reichardt and S. Bornholdt, "Partitioning and modularity of graphs with arbitrary degree distribution," *Phys. Rev. E*, vol. 76, no. 1, pp. 015102, 2007.

[8] Y.T. Chang, R.M. Leahy, and D. Pantazis, "Modularity-based graph partitioning using conditional expected models," *Physical Review E*, vol. 85, no. 1, pp. 016109, 2012.

[9] P. Hagmann, L. Cammoun, X. Gigandet, R. Meuli, C.J. Honey, V.J. Wedeen, and O. Sporns, "Mapping the structural core of human cerebral cortex," *PLoS Biol*, vol. 6, no. 7, pp. e159, Jul 2008.

[10] M.D. Fox, A.Z. Snyder, J.L. Vincent, M. Corbetta, D.C. Van Essen, and M.E. Raichle, "The human brain is intrinsically organized into dynamic, anticorrelated functional networks," *PNAS*, vol. 22, no. 27, pp. 9673–9678, Jul 2005.

[11] NITRC, "1000 functional connectomes project," `http://www.nitrc.org/projects/fcon_1000`, 2011.

[12] W.W. Zachary, "An information flow model for conflict and fission in small groups," *J. Anthropol Res.*, vol. 33, pp. 452–473, 1977.