

## Two test statistics for cross-modal graph community significance

Jonas Richiardi<sup>\*†</sup>, André Altmann<sup>\*</sup>, Michael Greicius<sup>\*</sup>

<sup>\*</sup>Functional Imaging in Neuropsychiatric disorders laboratory, Stanford University, USA

<sup>†</sup>Laboratory for Neurology and Imaging of Cognition, University of Geneva, Switzerland

**Abstract**—Comparing and combining data from different brain imaging and non-imaging modalities is challenging, in particular due to the different dimensionalities and resolutions of the modalities. Using an abstract and expressive enough representation for the data, such as graphs, enables gainful inference of relationship between biological scales and mechanisms. Here, we propose a test for the significance of groups of graph vertices in a modality when the grouping is defined in another modality. We define test statistics that can be used to explore subgraphs of interest, and a permutation-based test. We evaluate sensitivity and specificity on synthetic graphs and a co-authorship graph. We then report neuroimaging results on functional, structural, and morphological connectivity graphs, by testing whether a gross anatomical partition yields significant communities. We also exemplify a hypothesis-driven use of the method by showing that elements of the visual system likely covary in cortical thickness and are well connected structurally.

**Keywords**-systems neuroscience; network analysis; multimodal neuroimaging; brain graphs; brain connectivity

### I. INTRODUCTION

In the neuroimaging community, it is now common to refer to “brain networks” or “connectivity graphs”, where the terms apply to functional MRI (fMRI) connectivity of the brain, diffusion weighted imaging (DWI) - based measures, or even covariation of regional thickness or surface area. Multimodal datasets, reflecting various aspects of the idea of connectivity, are increasingly being acquired. Given a suitable modelling framework, the biological underpinnings of connectivity can be explored and may reveal interesting relationships between very different biological scales and processes.

We believe that graph-based methods offer the right amount of abstraction and modelling expressivity to establish such meaningful relationships. Here, we focus on *communities* in graphs, broadly defined as a set of vertices that have (relative) large edge weights between them but small edge weights with the rest of the graph. By assessing how community-like a set of graph vertices in one modality is, when grouped together by information in another modality, we propose a principled way to relate different connectivities with gross anatomical information, and enable hypothesis-driven testing.

While numerous methods exist to find communities in networks [1], comparatively few techniques allow assessing the statistical significance of the community structure of a whole network [2], and fewer still that of individual communities [3]. Further, objective evaluation of community scoring functions has only recently been proposed [4]. This paper will focus on the case where a graph is computed in one modality (e.g., a graph established from DWI data), and partitioned into communities based on another modality or knowledge source (e.g. anatomical division into lobes). It is then desirable to establish whether such partitions result in communities that are significant.

### II. TESTING FOR COMMUNITY SIGNIFICANCE

#### A. Network communities and measuring modularity

Given a simple undirected labelled graph  $g = (V, E, \alpha, \beta)$ , where  $V$  is the set of vertices,  $E$  is the set of edges, and  $\alpha$  is the set of vertex labels, the set of edge labels  $\beta : \mathbb{R}_+$  can be represented as a symmetric adjacency matrix  $\mathbf{A} \in \mathbb{R}_+^{|V| \times |V|}$ . We can compute a vector of vertex strengths (vertex volumes)  $\mathbf{s} \in \mathbb{R}_+^{|V|}$ , where  $s_v = \sum_{v'=1}^{|V|} \mathbf{A}_{v,v'}$ . Then, the total edge strength of the graph is given by  $w \in \mathbb{R}_+ = \mathbf{1}^T \mathbf{s}$ . In addition, a partition vector  $\mathbf{p} \in \mathbb{N}^{|V|}$ , defining the assignment of each vertex to a community, is given. Here, the graph and its adjacency matrix are defined on a modality, while the partition vector is defined in another modality.

A null model, encoding a random graph with the same expected strength distribution as the graph with adjacency matrix  $\mathbf{A}$ , can be obtained as  $\mathbf{N} = \frac{1}{2w} \mathbf{s} \mathbf{s}^T$  (called the configuration null model, or Newman-Girvan null model (NG)). Any other appropriate null model may be used to define the null adjacency matrix  $\mathbf{N}$ .

Finally, the modularity matrix [5] is defined as  $\mathbf{B} = \mathbf{A} - \mathbf{N}$ . The modularity  $Q$  is defined as a sum over  $C$  communities  $Q = \frac{1}{2w} \sum_{c=1}^C \sum_{i,j} B_{ij} \delta_{p_i, p_j}$ , where  $\delta$  is the Kronecker delta and  $p_i$  is an element of the partition vector. This quantity is evaluated for the whole graph, and does not tell us whether a particular community is significant. This, however, is of particular interest. Indeed it is likely that only *some* of the communities in one modality have a clear relationship with the organisation in another modality. We therefore propose tests for individual communities.

#### B. Test statistics for individual communities

Each community  $c$  defines a subgraph  $g_c = (V_c, E_c, \alpha_c, \beta_c)$  within the whole graph, where  $V_c \subset V, E_c \subset E, \alpha_c \subset \alpha, \beta_c \subset \beta$ . Intuitively, if the summed weight of all edges within vertices in a community is large relative to the summed weight of edges between vertices in this community and the rest of the graph, then the community is more likely to be significant than if it is around the same magnitude. We define *normalised community strength* as a plausible quantity in this regard:

$$S_c = \frac{\sum_{i \in V_c, j \in V_c} A_{ij}}{\sum_{i \in V_c, i \sim j} A_{ij}}, \quad (1)$$

where  $\sim$  represents an edge between vertex  $i$  and  $j$  ( $(v_i, v_j) \in E$ ). The denominator is the volume of the vertices in the subgraph (sum of weights inside the subgraph plus sum of weights of edges crossing the community boundary, that is,  $\sum_{i \in V_c, j \notin V_c, i \sim j} A_{ij}$ ).

This quantity is different from the *local modularity* of [6], which counts edges linking vertices where at least one (instead of both in Equation 1) is in the subgraph. It is equal to a weighted version of the conductance  $\Phi(c)$  of a subgraph  $c$  in an unweighted graph (“fraction of total edge volume that points outside the cluster” [4]),

if  $\sum_{i \in V_c, j \in V_c} A_{ij} = \frac{\sqrt{2}}{2} \sum_{i \in V_c, j \notin V_c, i \sim j} A_{ij}$ , namely, if the weight inside the subgraph is exactly  $\frac{\sqrt{2}}{2}$  smaller than the boundary volume. Thus, we have high correlation between  $S_c$  and  $1 - \Phi(c)$ .  $S_c$  is bounded as  $0 \leq S_c \leq 1$  and offers easy interpretation: 0 indicates a pathological community with no edge weight for any vertex in the community, and 1 indicates a community that has no edge to the rest of the graph (0 boundary weight). A value of 0.5 indicates that the within-community weight and the boundary weights are equal. A value  $0.5 < S_c < 1$  corresponds to the “weak sense” definition of community [7]: a community is defined as a subgraph where the within-community weight is above the boundary weight. Figure 1 plots  $S_c$  values as a function of within-community weight and boundary weight.

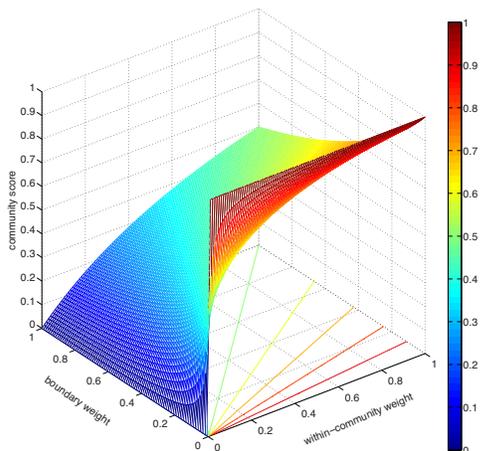


Figure 1. Normalised community strength values as a function of within-community weight and boundary weight.

We can further relate this quantity to graph modularity by using the null model in the denominator. This encodes the local excess (or lack) of edge strength within the community compared to the expected value in the null model. Thus, a valid test statistic corresponding to Equation 1 obtains from the modularity matrix and the adjacency matrix and is given by Equation 2, which we call *normalised modularity contribution*:

$$Q_c = \frac{\sum_{i \in V_c, j \in V_c} B_{ij}}{\sum_{i \in V_c, i \sim j} A_{ij}} \quad (2)$$

This statistic will be high if the vertices within the community have on average edge weights above what would be expected under the null model. The  $Q_c$  statistic is bounded above as follows:  $Q_c \leq S_c \leq 1$ . The lower bound depends on the null model, and may be a negative value. Summing the numerator of  $Q_c$  over communities and dividing by  $2w$  recovers the standard definition of modularity.

### C. Non-parametric testing procedure

We propose a permutation test to estimate the distribution of these two statistics under the null hypothesis “the value of the statistic is the same as if vertices of the community were chosen at random”. Communities are entirely defined by the partition vector  $\mathbf{p}$ , thus it is sufficient to randomly permute its components to obtain random assignment of vertices to communities. Each community

will keep the same number of vertices. Randomly permuting the partition vector a large number of times, and computing the test statistics of Equations 1 or 2 at each iteration is a fast procedure to generate the distribution of the test statistic under the null, as no graph has to be regenerated. Computing a normalised community strength has complexity  $\mathcal{O}(|V_c||V|)$ , and is linear in the number of permutations. The permutations are embarrassingly parallel. The normalised modularity contribution requires a null model, which can add a computational cost of up to  $\mathcal{O}(|V|^2)$  in a naïve implementation, but need only be computed once. These costs are not an issue on modern hardware for graphs with vertex set cardinality in the low hundreds as commonly used in neuroimaging.

Note that a community can be declared significant according to this test even though the corresponding normalised community strength is below 0.5, a value indicating that the subgraph of interest does not meet the definition of weak-sense community. In this case, the interpretation is that the vertices in the subgraph have more within-subgraph strength than expected by chance (as represented by the permutation null model), but that the subgraph is not well-separated from the rest of the graph.

### D. Related approaches

Community significance is closely related to problem of cluster validity in pattern recognition. The use of a null model in this context is well-studied, including the principle of generating a random graph (or dissimilarity matrix). Indices of clustering validity (e.g. the Jaccard coefficient) have been the focus of most research and assess whole clusterings rather than a particular cluster. Likewise, Chang et al’s Monte Carlo fitting approach [2] focuses on evaluating the significance of a whole graph partition, rather than of individual communities, and is therefore not directly comparable to our approach.

The OSLOM algorithm [3], in particular the B-score which yields individual communities significance, can accommodate overlapping communities and hierarchical modules, which our algorithm cannot do. It is based on the combinatorial computation of the probability that a vertex has a certain number of neighbours in a subgraph. This can then be compared with the expected number of neighbours under a null hypothesis. Computing the OSLOM B-score for a subgraph  $c$  has complexity  $\mathcal{O}(|V_c|^2)$ , lower than our approach. For the kind of graphs and partition vectors we are interested in here, the main difference is that OSLOM is expected to have more specificity than our proposed method (since it is a design goal to avoid finding clusters in random graphs), but at the expense of sensitivity for subgraphs that do not meet the weak-sense definition of community. One benefit from our approach is easy interpretability of  $S_c$  (Eq. 1) and the direct relationship of  $Q_c$  to modularity.

## III. EXPERIMENTS

### A. Datasets and data processing

1) *Graph data with known communities*: We generated a synthetic adjacency matrix for 10 communities, comprising 5 vertices. The first 5 communities were “true” communities, whereas the last 5 were not supported by the data. Edge strengths were generated at random, using Pearson correlation of a random multivariate time series. The expected value of the edge weights outside communities and in false communities was proportional but smaller than the

expected values of edges within true communities. Edge weights smaller than 1/3 of the median were thresholded and replaced by zeros. The adjacency matrix for a high ratio of within-community-weight-to-boundary-weight expected values is shown in Fig 2(a).

We also used a subgraph of the binary DBLP collaboration network [4], randomly selecting 500 communities from the top 5000, and generating an equal number of fake communities of the same size by drawing at random from vertices not in the top 5000 communities. The sizes of communities ranged from 6 to 7556 vertices (median: 8). The total graph contained 31284 vertices.

2) *Functional connectivity graph from fMRI data:* We use a random subset of 8 subjects from a previous study [8], where subjects go through alternating epochs of rest and movie watching. The data was acquired using a Siemens 3T scanner, with  $3.75 \times 4.2$  mm voxels and a TR of 1.1 s.

The dataset was processed with an SPM8 pipeline [8], regionally averaged in 88 AAL atlas regions (excluding bilateral globus pallidus due to signal dropout) and wavelet-transformed in the 0.06–0.11 Hz subband. Blocks of rest and movie-watching were not separated. A functional connectivity graph was computed with edge weights from Pearson correlation. No regularisation was used due to the large number of time points.

3) *Structural connectivity graph from DWI data:* A pre-computed structural connectivity graph was obtained from the UCLA multimodal connectivity database (UCMD) [9] (ICBM study). The data was acquired using a Siemens 1.5T magnet, with 30 diffusion directions, and is an average of the structural connectivity of 138 ICBM normal subjects from the 68-regions parcellation of the Killian-Desikan atlas.

4) *Morphological connectivity graph data from structural MRI data:* A pre-computed cortical thickness correlation graph was also obtained from the UCMD’s ICBM study. This is computed across the same subjects as the DWI data, and uses 1mm isotropic voxels. The same atlas was used as for the DWI data.

## B. Results

1) *Synthetic and DBLP data graph, ground truth partition vector:* Using normalised modularity contribution with a NG null model on the synthetic graph shown in Figure 2(a), five out of five real communities were significant (1000 permutations, Bonferroni-corrected for a critical value of 0.05, all p-values 0.001), and all five wrong communities were non-significant (min p-value 0.035, max p-value 0.88). The Area Under Curve (AUC) was 1 (thus all significant communities had p-values below those of non-significant communities) and the Equal Error Rate (EER) was 0% (thus at an appropriate threshold all communities could be identified correctly). The performance was the same for normalised community strength. Violin plot for the permuted and non-permuted value of the statistic for all 10 communities are shown in Figure 2(b). The OSLOM B-score declared all communities non-significant, and thus achieved an AUC of 0.5 and EER of 50%. Since the normalised community strengths are between 0.16 and 0.19 for the ‘real’ communities, these do not satisfy the weak-sense definition of a community, and it is likely that the B-score is penalising them for this reason. Nevertheless a subgraph structure is visible and it is of interest to examine it, albeit with caution.

For the 31284-vertices DBLP graph, the AUC was 0.999, the EER was 0.1%, and setting a threshold of 0.5 on the p-value

yielded a sensitivity of 100% and a specificity of 99.4% when using normalised community strength. Using OSLOM B-score on the same data yielded an AUC of 0.924, an EER of 7.6%, a sensitivity of 84.8% and a specificity of 100%. Since the average subgraph strength in the 500 real communities is 0.76 (standard deviation 0.25), they satisfy the weak sense definition of community, and we trust that these are not artefactual communities.

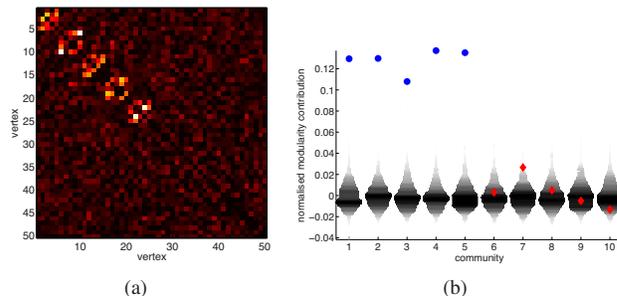


Figure 2. (a) Synthetic adjacency matrix with 10 communities of varying number of vertices. The first 5 (until vertex 50) are supported by the data, the last 5 are not. Edge weights set by correlation. (b) Permuted distribution (in black) and original value (dots) of the normalised modularity contribution statistic on the synthetic graph. The real communities (first 5) are significant (blue dots), and the wrongly assumed communities are not (last 5, red diamonds)

2) *DWI data graph, anatomical partition vector:* With DWI data, the graph where we wish to determine community significance is derived from tractography, and the partition vector defining communities is from gross anatomy. Each of the 68 vertices (region) was assigned to one of 6 anatomical ‘lobes’ (communities): Frontal, Temporal, Occipital, Parietal, Cingulate, or Insula, using the rough mapping in <http://surfer.nmr.mgh.harvard.edu/fswiki/CorticalParcellation>. This tests whether intra-lobe structural connectivity is particularly strong.

In the test, all communities except the Insula are declared significant at  $p=0.001$ , both for the normalised community strength and the normalised modularity contribution statistic with the NG null model. The frontal lobe has the highest  $S_c$  value, 0.46, and indeed looks to have relatively few connections to other lobes. The Insula community consists only of the left and right insular cortex, and they respectively exhibit much larger structural connectivity with ipsilateral brain regions (in particular with frontal (pars opercularis, pars triangularis) and temporal (superior temporal) regions) than with the contralateral insula. These regions have been reported as having important afferent and efferent connections to and from the insula in primates [10].

For the OSLOM B-score, only the Cingulate community is significant, all others have  $p=0.99$ .

3) *Structural MRI data graph, anatomical partition vectors:* First, we wish to test whether cortical thickness correlation is significantly higher within lobes. The same gross anatomical partition vector as for the DWI data was used.

Both  $S_c$  and  $Q_c$  statistics showed that the Occipital, Parietal, and Temporal communities were significant at 0.05 Bonferroni-corrected. Cingulate and Frontal communities failed the corrected significance test with  $S_c$  ( $p=0.019$  and  $0.012$  respectively), but passed with  $Q_c$ . The Insula community failed both significance

tests, although marginally ( $p=0.009$  for both  $S_c$  and  $Q_c$ ). All communities failed the OSLOM B-score test ( $p=0.99$ ). Here, The highest  $S_c$  is for the Parietal community (0.30), on par with its value on DWI data. Other communities all have weaker  $S_c$ , indicating overall that communities in morphological connectivity graphs may be more weakly apparent than in DWI graphs.

We can use the same graph data, but a different partition vector, to test a structural hypothesis about the visual system: covariation of volumes of components of the visual system could indicate of their interdependence in development [11]. Here, we can group regions into a functional “visual system” community to test whether the cortical thickness correlation is remarkable. Our community comprises bilateral fusiform (temporal lobe), bilateral pericalcarine (occipital lobe), and bilateral lingual gyrus (occipital lobe). Running the test with the  $Q_c$  statistic shows that this community is significant at  $p=0.001$ . Furthermore, it also seems to be subtended by strong structural connectivity, as the same community defined in the DWI data graph also yielded a significant result at  $p=0.003$ . However, the  $S_c$  value in the morphological connectivity graph is rather low (0.10), and slightly higher in the structural connectivity graph (0.20), and thus the effect is probably not very strong. The OSLOM B-score deems the “visual system” community insignificant.

4) *fMRI data graph, anatomical partition vector*: Here, the partition vector is based on the identity of brain lobes to which the fMRI connectivity graph regions belong. The null model used was the NG null model. All tests were Bonferroni-corrected at the 5% significance level.

The “central” community (bilateral Rolandic operculi) is non-significant in all subjects. Frontal, Limbic, and Parietal communities are significant for the majority of subjects using the  $Q_c$  statistic, but for 50% or less when using the  $S_c$  statistic. Occipital, Subcortical, and Temporal communities are significant for a majority of subjects using both  $Q_c$  and  $S_c$  statistics. OSLOM B-score found no significant community in any subject. The Occipital community had an average  $S_c$  of 0.15 (standard deviation 0.02), while the lowest average was for the Temporal community ( $S_c$  average 0.09, standard deviation 0.02). Thus, it is expected that Occipital, subcortical, and temporal “lobes” have a significant excess of connectivity within them, but since they do not satisfy the definition of a weak-sense community, that they are also strongly functionally connected to other lobes.

#### IV. DISCUSSION AND LIMITATIONS

We have proposed a practical and interpretable approach to test the significance of communities defined across modalities, yielding plausible significant and insignificant communities in graphs computed from several modalities. We have shown that the approach has high sensitivity, perhaps at the expense of specificity, and suggested checking results based on the normalised community strength of communities. We believe that such an approach can be particularly interesting when testing hypotheses linking brain structure and function. In particular, the improved sensitivity may be required to detect structure in noisy data, where insisting on a strict definition of community may cause many false negatives.

We point out that choosing a null model appropriate for the data is not trivial, and a null model preserving degree distributions may not be the most appropriate. If Pearson correlation is used to set the

adjacency matrix  $\mathbf{A}$ , it has been posited that some null models, such as a random degree-preserving rewiring, lead to overestimation of the modularity [12]. We could use the Hirschberger-Qi-Steuer (HQS) algorithm to generate a null model from the covariance matrix corresponding to  $\mathbf{A}$ .

Our approach applies to graphs whose edge labels are positive. Edges with negative-valued labels are thresholded away, possibly losing information. Alternatives have been developed in the network science literature, with negative edge labels indicative of between-community links. However the role and significance of negative correlation in fMRI data and brain imaging is still being debated, and it is not clear these approaches would apply.

One pitfall is that for very small communities in small graphs, the number of distinct combinations of  $|V_c|$  vertices may be smaller than the number of permutation iterations, leading to artificially low p-values. This happens if  $\binom{|V|}{|V_c|}$  is smaller than the number of permutation iterations.

#### ACKNOWLEDGMENTS

This work is supported in part by the European Union under a Marie Curie International Outgoing Fellowship (FP7-PEOPLE-2011 IOF #299500), and in part by NIH grant #NS073498. The authors thank the anonymous reviewers for their helpful suggestions.

#### REFERENCES

- [1] S. Fortunato, “Community detection in graphs,” *Physics Reports*, vol. 486, no. 3, pp. 75–174, 2010.
- [2] Y.-T. Chang, D. Pantazis, and R. M. Leahy, “Assessing statistical significance when partitioning large-scale brain networks,” in *Prof. IEEE Int. Symposium on Biomedical Imaging (ISBI)*, 2012, pp. 1759–1762.
- [3] A. Lancichinetti, F. Radicchi, J. J. Ramasco, and S. Fortunato, “Finding statistically significant communities in networks,” *PLoS one*, vol. 6, no. 4, p. e18961, 2011.
- [4] J. Yang and J. Leskovec, “Defining and evaluating network communities based on ground-truth,” in *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics*. ACM, 2012, p. 3.
- [5] M. E. Newman, “Modularity and community structure in networks,” *Proceedings of the National Academy of Sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.
- [6] A. Clauset, “Finding local community structure in networks,” *Physical Review E*, vol. 72, pp. 026132–1–6, 2005.
- [7] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, “Defining and identifying communities in networks,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 9, pp. 2658–2663, 2004.
- [8] J. Richiardi, H. Eryilmaz, S. Schwartz, P. Vuilleumier, and D. Van De Ville, “Decoding brain states from fMRI connectivity graphs,” *NeuroImage (Special Issue on Multivariate Decoding and Brain Reading)*, vol. 56, no. 2, pp. 616–626, May 2011.
- [9] J. A. Brown, J. D. Rudie, A. Bandrowski, J. D. V. Horn, and S. Y. Bookheimer, “The UCLA multimodal connectivity database: a web-based platform for brain connectivity matrix sharing and analysis,” *Frontiers in Neuroinformatics*, vol. 6, pp. 28–1–28–17, 2012.
- [10] J. R. Augustine, “Circuitry and functional aspects of the insular lobe in primates including humans,” *Brain Research Reviews*, vol. 22, no. 3, pp. 229–244, Oct 1996.
- [11] T. J. Andrews, S. D. Halpern, and D. Purves, “Correlated size variations in human visual cortex, lateral geniculate nucleus, and optic tract,” *J Neurosci*, vol. 17, no. 8, pp. 2859–2868, Apr 1997.
- [12] A. Zalesky, A. Fornito, and E. Bullmore, “On the use of correlation as a measure of network connectivity,” *Neuroimage*, vol. 60, no. 4, pp. 2096–2106, May 2012.