# 480

# Efficient distributed algorithms for pattern detection in graphs derived from fMRI measurements

*Guillermo A. Cecchi[1], A. Ravi Rao[1], Dante R. Chialvo[2], Avi Ma'ayan[3], Maria V. Centeno[2] and A. Vania Apkarian[2]*
[1]IBM T. J. Watson Research Center, Yorktown Heights, NY 10598, USA
[2]Dept. of Physiology, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611
[3]Dept. of Pharmacology and Biological Chemistry, Mount Sinai School of Medicine, New York, NY 10029, USA.

## Introduction

Recently, we introduced an approach to study topological properties of functional brain networks[1]. We demonstrated that the graphs determined by the structure of pairwise correlations between voxels display very robust topological statistical regularities, including power-law connectivity scaling and small-worldness, that are shared among other large-scale biological and technological networks.

One difficulty in extending this analysis, and applying it to obtain a useful discriminatory power between different brain states, is that the computations become intractable very easily as one moves up from two-point correlations. Here we present a novel approach that extends our previous findings to include *directional* links, and based on this analyze the presence and significance of higher-order correlation patterns.

In order to tackle the inherently costly computational demands, we developed a series of algorithms implemented on distributed platforms that render our approach feasible.

## Directional Links

In our previous study, we defined graphs embedded in the functional networks by thresholding the normalized covariance between voxels *i* and *j* to define the binary link $d_{ij}$:

$$c_{ij} = \left\langle (v_i(t) - \overline{v}_i)(v_j(t) - \overline{v}_j) \right\rangle_t \sigma_i^{-1} \sigma_j^{-1} \qquad \overline{v}_i = \left\langle v_i(t) \right\rangle_t, \sigma_i^2 = \left\langle (v_i(t) - \overline{v}_i)^2 \right\rangle$$

$$\text{if}: c_{ij} > C_r \text{ then } d_{ij} = 1, \text{else}: d_{ij} = 0$$

We define now a directional link in the spirit of Granger causality, such that the predictability of one voxel by the past of another one implies "causation" by the latter:

$$c_{ij}(\tau) = \left\langle (v_i(t-\tau) - \overline{v}_i)(v_j(t) - \overline{v}_j) \right\rangle_t \sigma_i^{-1} \sigma_j^{-1}{}_t$$

$$\text{if}: \quad c_{ij}(\tau = 0) > C_r \quad \text{then} \quad d_{ij} = 1, d_{ij} = 1$$

$$\text{else if}: c_{ij}(\tau > 0) > Cr \quad \text{then} \quad d_{ij} = 1, d_{ij} = 0$$

$$\text{otherwise}: \quad d_{ij} = 0$$

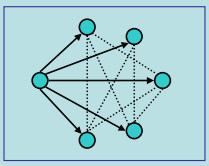


**Figure 1**

The limitation of this type of approach is the identification of many spurious correlations, as in common source-based correlations. We address some of these concerns by precisely eliminating the neutral links (ie. not directed) that can be explained by a common directed source, as shown in Figure 1. This procedure allows for a considerable reduction in the number of undirected (or neutral) links respect to similar studies[2], including our own.

## Topological Network Properties

We explored the potential explanatory power of this method by studying the topological properties of the derived directed networks, using as a test case three highly similar tasks: a self-paced finger-tapping tasks, cued by (**a**) a small visual cue, (**b**) and auditory cue, and (**c**) a large visual cue.

Interestingly, in 5 out of 6 subjects, the normalized diameter of the network, ie. the average number of "hops" from one node to any other, divided by the diameter of the corresponding random network, can separate between the visual and auditory cue tasks, as shown in Figure 2. The significance of this finding can also be inferred from the results shown in Figure 3, where pseudo-maps color-code the local connectivity of the neutral and directed connections. Observe the remarkable similarity of the neutral maps, and the dramatic disparity for the directed links pseudo-map.

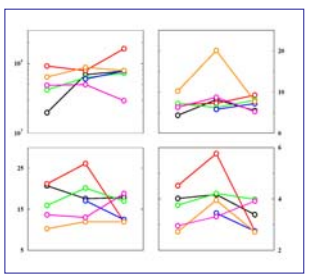


**Figure 2**

Top left: total number of nodes in the network. Top right: mean connectivity (all links). Bottom left: diameter of the network. Bottom right: normalized diameter. The colors indicate different subjects, and the data points correspond (from left to right) to tasks (**a**), (**b**) and (**c**). In all cases, the "giant" component of the network was considered, ie. the largest connected subnetwork, which is always at least one order of magnitude larger than the smaller components.

## Higher-order Patterns

Embedded in these networks, there is a potential wealth of higher-order patterns that may significantly relate to functional aspects, as it has been shown for other (but certainly very different) biological networks[3]. The correspondingly higher time complexity of the algorithms for pattern discovery limits the search space; here we focus on *topological cycles*, ie. the classes of irreducible topologies of cycles determined by neutral and directed links[4]. For simplicity, and as a first step, we classify these cycles based on the number of pass-through nodes they include, as represented in Figure 4.

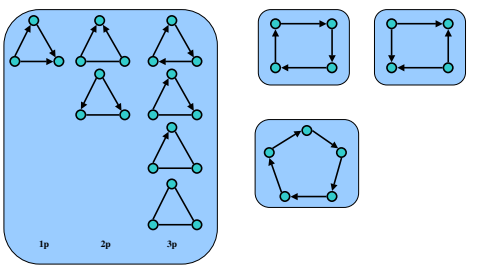
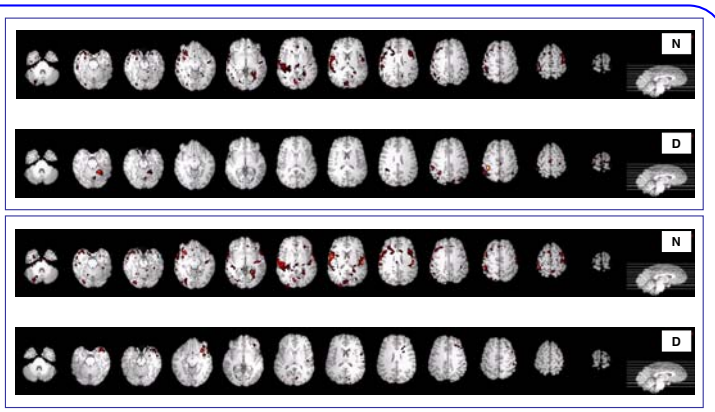



**Figure 4**

Left: all topological classes for cycle of length 3, where p stands for number of pass-through nodes. Top right: example of 4p and unique case of 0p for cycle-length 4. Bottom right: example of 5p for cycle-length 5.



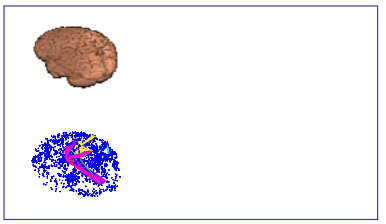



**Figure 3**

Top panel: pseudo-map for the small visual cue task; the top figure corresponds (labelled N) to neutral links, and the bottom one to directed links (D). Middle panel: same as above for the auditory cue task. Bottom: a subset of 9 on the most connected "sources" (ie. having outgoing links), for the small visual cue task. The color code preserves the identity of the source.

## Cycle Distribution

As an initial analysis, we computed the distribution of the different topological classes within each cycle length. In other words, we estimated the relative contribution of 0 p(ass-through), 1p, 2p, etc. to the total number of cycles present for lengths 3, 4 and 5. In five of the six subjects, we found that the distribution of topological classes follows a very consistent pattern in all cycle lengths across the three tasks, and unique for each individual, suggesting a "signature" for this kind of task. Remarkably, the total number of nodes within the same subject can vary dramatically. However, for the one subject that did not follow this pattern, we found that it was precisely the auditory task that consistently displayed a significant difference respect to the other two tasks.

To further investigate the possibility of a differentiation based on topological classes, we study another subject under more disparate tasks: (a) passively listening to music, (b) silently generating words, and (c) silently recalling words. We found that the music listening task consistently differed in its classes distribution, even when the network diameter and number of nodes did not.

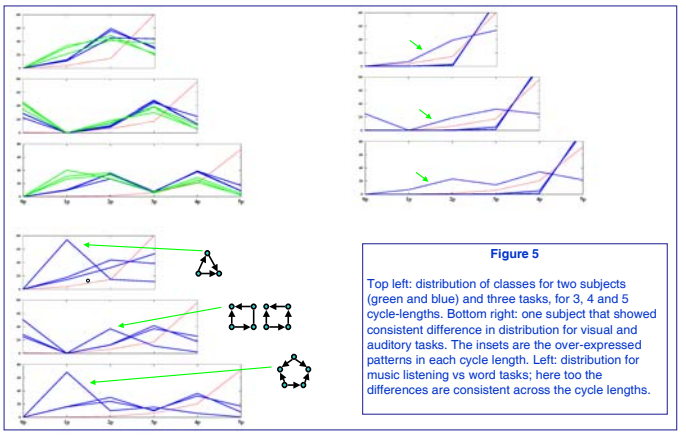


**Figure 5**

Top left: distribution of classes for two subjects (green and blue) and three tasks, for 3, 4 and 5 cycle-lengths. Bottom right: one subject that showed consistent difference in distribution for visual and auditory tasks. The insets are the over-expressed patterns in each cycle length. Left: distribution for music listening vs word tasks; here too the differences are consistent across the cycle lengths.

## Outlook and Conclusions

The algorithms used here were developed for a general purpose application to biological networks, and were optimized for distributed memory environments. Both the directional link determination and the topological classes computation are highly intensive, but feasible implementations were achieved using IBM's Blue Gene supercomputer. On a single rack (1024 processors), the link computation takes 8 minutes for 400 volumes scans, and a maximum of 45 minutes for the cycles at a sub-sampling rate of 1 in 4 nodes. The increasing accessibility to these type of high performance computing environment will render this and similar approaches a mainstream aspect of functional brain mapping.

We described a novel method to capture higher-order correlations in functional brain networks. The initial results presented here regarding the discriminatory power of the method to separate brain states, and the ability to identify areas contributing in qualitatively distinct fashions within a task, highlight the potential of this approach to uncover hidden structures of the functional brain.

## References

1. *Scale-Free Brain Functional Networks*, V.M. Eguiluz, D.R. Chialvo. G.A. Cecchi, M. Baliki & V.A. Apkarian, *Physical Review Letters* 94, 018102 (2005)
2. *A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs*, S. Achard, R. Salvador, B. Whitcher, J. Suckling & E. Bullmore, *Journal of Neuroscience* 26(1):63-72 (2006); *Functional connectivity by cross-correlation clustering*, S. Dodel, J.M. Herrmann & T. Geisel, *Neurocomputing* 44, 1065 (2002).
3. *Network Motifs: Simple Building Blocks of Complex Networks*, R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, *Science 298:824 – 827 (2002).*
4. *Characterization of large cycles in biological and engineered networks*, A. Ma'ayan, G.A. Cecchi, J. Wagner, A.R. Rao, A. Lipshtat, R. Iyengar and G. Stolovitzky, submitted (2006).