LFN Project - *Graph based analysis of the functional connectome in neuropsychiatric disorders*

Christy Jo Manthara[1] Francesco Pio Monaco[2] Michele Russo[3]
1 2080644, 2 2087638, 3 2087637

Abstract

Neuropsychiatric disorders are a family of diseases that have a variable impact on the lives of those who suffer from them, and insights are needed to understand the underlying mechanisms and plan treatments to maximise their quality of life. This project uses one of the main approaches to the study of brain mechanisms, to analyse such disorders and their impact on functional structure.

Introduction

Many biological, social, and technological systems can be represented as networks, consisting of multiple elements that interact with each other. A network perspective highlights that the behavior of a complex system is shaped by the interactions among its constituents. This approach offers the possibility to analyze systems of different natures within a unifying mathematical framework.

The brain can be viewed as a network of interconnected components that support the emergence of adaptive behaviour and cognition. Currently, various imaging techniques are used to peer inside the brain, such MRI, electroencephalogram (EEG), and positron emission computed tomography (PET) with the aim of detecting salient network elements that regard communication and signal transfer.

In particular, the combination of magnetic resonance imaging (MRI) and graph theory analysis has become a powerful approach to study large-scale networks of both structural and functional connectivity. This approach provides essential information that contributes to our understanding of the underlying neural mechanisms and finding techniques that better highlights brain functionalities is an increasingly active of topic research ([9]; [3]; [6]; [10]).

The aim of this project is to discuss how the brain changes in different diseases by representing the brain as a network and then applying meaningful metrics to report features that can underline the impact of the disease on the whole brain network. For this study we focused on neuropsychiatric disorders such as *bipolar disorder*, *ADHD*, and *schizophrenia*.

Data & Materials

BIDS dataset *ds000030*: 20 subjects for each category (control, bipolar, ADHD, schizophrenia);
nilearn is used to extract the timeseries from the fMRI data;
networkx is used to compute all the metrics and to perform community detection;
graph-tool is used for graphlets counting.

Creating the graphs

Subject graphs creation

Once extracted the pre-processed time-series we proceed in determining the power correlation matrix. Each subject has 1000 time-series. The power correlation matrix is calculated for all pairs of normalized time-series, using *Pearson correlation*, defined as:

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} \tag{1}$$

where: X, Y are normalized timeseries, *cov* is the covariance, σ_X is the standard deviation of X, and σ_Y is the standard deviation of Y. *Sparsification* was then obtained by using an adaptive threshold, inspired by [5], computed in the following way:

1. Flattening the power correlation matrix without the diagonal elements (self-correlation);

2. $t = (\sum_{i=0}^{n^2-n} x_i) * \frac{f}{100}$, $f \in [0,1]$;

3. All elements $s_{X,Y}$ (of the sparsified matrix) are computed as:

$$s_{X,Y} = \begin{cases} 0 \text{ if } \rho_{X,Y} \leq t \\ 1 \text{ otherwise} \end{cases} . \tag{2}$$

A way to obtain directly a *sparsified matrix* while also taking into account non-linearities in the functional connectivity is by using a *Graphical Lasso* (GL), the method did not work due to ill-conditioned system for the solver; the work of [5] shows that using a standard correlation technique with adaptive thresholding offers results that are very close to a GL while being less demanding in terms of computation.

The *sparsified matrix* is an adjacency representation of an unweighted, disconnected graph. Data about the number of edges and nodes are available in (Table 1).

Random graphs creation

To have a baseline for relevancy, random graphs with the same degree frequency are computed for each subject in our analysis (for a total of 80 random graphs) as done in [19], via the *networkx* function `configuration_model` that uses the *Newman model*.

Graph metrics

All the graphs are then processed to compute the subsequent global metrics: closeness, betweenness, node clustering, degree.
From there, the information returned is the mean on the whole graph and the mean of the top 5 nodes. Additional metrics include:

modularity, that measures the effectiveness of a network division, in the sense that there are many edges within communities (intra-edges) and only a few between them (inter-edges), it is defined as:

$$Q = \frac{1}{2m} \sum_{i,j} \left(A_{ij} - \gamma \frac{k_i k_j}{2m} \right) \delta(c_i, c_j) \tag{3}$$

where m is the number of edges (or sum of all edge weights), A is the adjacency matrix of G, k_i is the (weighted) degree of i, γ is the resolution parameter, and $\sigma(c_i, c_j)$ is 1 if i and j are in the same community else 0 [13].

Assortativity is the tendency of nodes with similar numbers of edges to connect.

$$r = \frac{\sum_{jk}(jk(e_{jk} - a_j b_k))}{\sigma_a * \sigma_b} \tag{4}$$

where: r is the assortativity coefficient, jk are pairs of connected nodes, e_{jk} is the fraction of edges that connect nodes with degrees j and k, a_j and b_k are the fractions of vertices with degrees j and k, respectively, in the end σ_a and σ_b are the variances of the degree distributions [7].

Transitivity measures the probability that the adjacent vertices of a vertex are connected, defined as [4]:

$$T = 3\frac{\#triangles}{\#triads} \tag{5}$$

Global efficiency measures how efficiently a network exchanges information. It is defined as [11]:

$$E(G) = \frac{1}{n(n-1)} \sum_{i \neq j \in V} \frac{1}{d(i,j)} \tag{6}$$

Local efficiency is the average efficiency of the local subgraphs:

$$E_{\text{loc}} = \frac{1}{N} \sum_{i \in V} E(G_i) \tag{7}$$

where G_i is the subgraph of the neighbours of i and V is the set of nodes of G. Since $i \in G_i$, the local efficiency E_{loc} tells how much the system is fault tolerant, thus how efficient is the communication between the neighbours of i when i is removed [11].

Results are available in (Table 1), the notebook offers a distribution plot for each metric to better highlight these results, the plots were omitted for space reasons.

Significance of the Features

A comparative analysis was performed between the features of our study and those of random graphs for each condition. The significance of these features was evaluated using p-values, with an alpha level that was adjusted for multiple comparisons using the Bernoulli correction method (for three comparisons). The results of this analysis revealed that all the features under consideration were statistically significant, full results can be obtained in the notebook.

Table 1. Metrics mean and std across different conditions.
top5 values are computed by averaging the 5 highest values for that metric.

Metric	Control	Schizophrenia	Bipolar	ADHD	Random
closeness	0.2009 ± 0.0140	0.1969 ± 0.0170	0.2083 ± 0.0177	0.2057 ± 0.0148	0.3284 ± 0.0102
betweenness	0.0047 ± 0.0005	0.0048 ± 0.0005	0.0046 ± 0.0005	0.0046 ± 0.0005	0.0027 ± 0.0001
degree	11.0967 ± 0.7213	11.4928 ± 1.2532	11.2157 ± 1.0959	10.8211 ± 0.8681	10.8349 ± 0.8877
clustering	0.3900 ± 0.0265	0.3937 ± 0.0297	0.3796 ± 0.0458	0.3810 ± 0.0435	0.0448 ± 0.0141
top5_close	0.2762 ± 0.0192	0.2740 ± 0.0201	0.2825 ± 0.0222	0.2782 ± 0.0169	0.4242 ± 0.0161
top5_betw	0.0578 ± 0.0150	0.0603 ± 0.0146	0.0564 ± 0.0153	0.0493 ± 0.0102	0.0277 ± 0.0040
top5_deg	54.2500 ± 7.8932	60.2667 ± 13.2132	55.1556 ± 6.3895	54.5300 ± 6.6495	52.3237 ± 7.3323
top5_clust	1 ± 0	1 ± 0	1 ± 0	1 ± 0	0.7861 ± 0.2023
modularity	7.0561 ± 2.6546	6.9279 ± 2.8339	6.2486 ± 1.9614	5.4550 ± 1.9286	0.4743 ± 0.2179
global_efficiency	0.2324 ± 0.0126	0.2293 ± 0.0158	0.2397 ± 0.0175	0.2359 ± 0.0139	0.3486 ± 0.0113
edges	4246 ± 0	4246 ± 0	4246 ± 0	4246 ± 0	4130.3816 ± 39.9887
nodes	768.3500 ± 49.9708	747.1111 ± 80.1586	764.1111 ± 75.5886	789.4000 ± 61.1361	767.8553 ± 67.5386
assortativity	0.3314 ± 0.0657	0.3626 ± 0.0981	0.3053 ± 0.1039	0.3639 ± 0.0708	-0.0148 ± 0.0174
transitivity	0.4291 ± 0.0297	0.4416 ± 0.0513	0.4182 ± 0.0415	0.4168 ± 0.0438	0.0516 ± 0.0164
top5_modular_close	0.1994 ± 0.07	0.2199 ± 0.07	0.2466 ± 0.09	0.2512 ± 0.08	1.4066 ± 0.1583
top5_modular_betw	149.04 ± 439	78.52 ± 79	48.48 ± 34	95.44 ± 83	3.05 ± 2.17

Table 2. p-values of t-tests. Results vs control group.

Metric	Schizophrenia	Bipolar	ADHD
closeness	0.4367	0.1608	0.2953
betweenness	0.5535	0.5887	0.5237
degree	0.2345	0.6921	0.2816
clustering	0.6856	0.3943	0.4369
top5_close	0.7346	0.3510	0.7245
top5_betw	0.6105	0.7723	0.0434
top5_deg	0.0932	0.7018	0.9041
modularity	0.8864	0.2979	0.0354
global_efficiency	0.5031	0.1451	0.4084
assortativity	0.2539	0.3556	0.1409
transitivity	0.3580	0.3561	0.3066
local_efficiency	0.9092	0.3967	0.5350
top5_modular_closeness	0.3871	0.0749	0.0378
top5_modular_betweenness	0.5067	0.3397	0.5948

Community detection in functional brain networks

Graph *communities* are subsets of nodes within the graph, such that connections between the nodes inside the community are denser than connections with the rest of the network.

Community detection divides the nodes of the preliminary network into several non-overlapping clusters by maximizing the module quality metric Q also called *modularity* [14]. A higher value of *modularity* represents that the detection approaches a more evident community structure.

The *community detection* algorithm takes as input a brain graph $G = (V, E)$ generated from *f-MRI* data. Since G is an un-connected graph, for each connected component we apply the *greedy_modularity_communities* function [4]. This function uses *Clauset-Newman-Moore greedy modularity maximization* to find the community partition with the **largest modularity**.

This method is a *hierarchical clustering* algorithm that starts by assigning all $v \in V$ into a sole community, then repeatedly joining together the two communities whose amalgamation produces the largest increase in Q. Then it returns the $\text{argmax}_i Q(C_i)$ where C_i is the resulting clustering at each iteration. The entire process can be represented as a tree whose leaves are the vertices of the original network and whose internal nodes correspond to the joins. This *dendrogram* represents a hierarchical decomposition of the network into communities at all levels [1].

We decided to use this algorithm since it is faster than other algorithms we tested, such as *spectral clustering* and *Louvain clustering*.
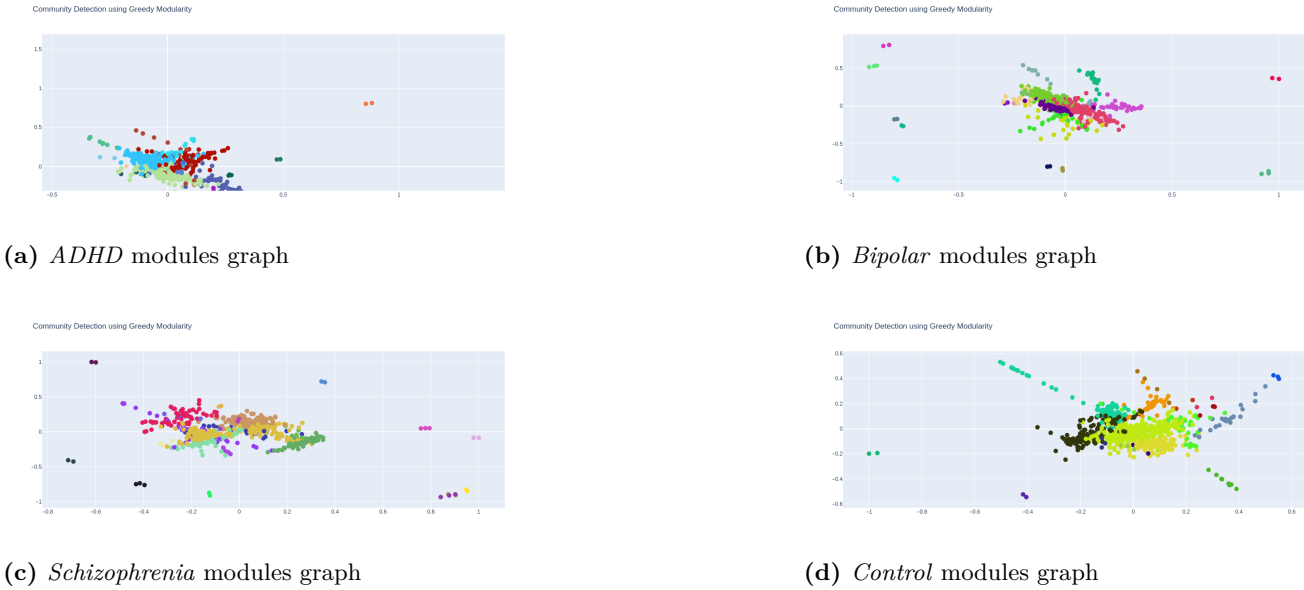

(a) *ADHD* modules graph

(b) *Bipolar* modules graph

(c) *Schizophrenia* modules graph

(d) *Control* modules graph

Figure 1. These images show the graph partitioning after using *greedy_modularity_communities*, each module has a unique color, nodes with different colors belongs to different modules.

Modular centralities

To identify influential nodes, it is important to consider the community structure. In modular networks, a node's influence can be divided into two types: local influence, which is linked to its community features, and global influence, which is related to its interactions with other communities. We implemented *modular centralities*, by starting from the definition of *closeness centrality* and *betweeness centrality*. We took inspiration from [8]; in their work, they illustrated the process to infer *modular centralities* from already existing centralities.

Let $G = (V, E)$ be a network with a non-overlapping community structure partitioned into m communities $C_1, C_2, ..., C_m$. The *local network* G_l is formed by uniting disjoint modules $G_l = \cup_{k=1}^m C_k$. Let C_k be a non-overlapping community, where V_k and E_k denote the set of vertices and edges within C_k, respectively. For a selected centrality measure β_L, the Local Centrality of node v_i^k, such that $v^i \in C_k$ is denoted as $\beta_L(v, C_k)$.

Consider a network G, where the global network G_g is the union of connected components $S = \{S_1, \ldots, S_p\}$ obtained after removing intra-community links. Let $\mathcal{S}$ be the set of revealed connected components ($|\mathcal{S}|$ is its size). The global network is defined as $G_g = \cup_{q=1}^p S_q$. For a selected centrality measure β_G, the *Global Centrality* of node

v_i^q, such that $v_i^q \in S_q$, is defined as $\beta_G(v, \mathcal{S}_q)$. Global centrality for isolated nodes is set to 0.

Input : Graph $G(V, E)$, Centrality measure β
Output: A map $M(node : centrality\ vector)$

1 Remove all the inter-community links from G to form the local network G_l
2 Remove all the intra-community links from G to form the global network G_g
3 Create and initialize an empty map $M(node : B_M)$
4 $B_M(v_i) = (\beta_L(v_i^k), \beta_G(v_i^q))$ represent the centrality vector, where each node v_i of the network should be associated with its Local and Global value according to the selected centrality measure β
5 **for** *each $C_k \subset G_l$, where $k \in \{1, \ldots, m\}$* **do**
6   **for** *each $v_i^k \in V_k$* **do**
7   Calculate $\beta_L(v_i^k)$
8       $B_M(v_i).add(\beta_L(v_i^k))$
9   **end**
10 **end**
11 **for** *each $S_q \subset G_g$, where $q \in \{1, \ldots, p\}$* **do**
12   **for** *each $v_i^q \in V_q$* **do**
13   Calculate $\beta_G(v_i^q)$
14       $B_M(v_i).add(\beta_G(v_i^q))$
15   **end**
16 **end**
17 **for** *each $v_i \in V$* **do**
18    $M.add(v_i, B_M(v_i))$
19 **end**
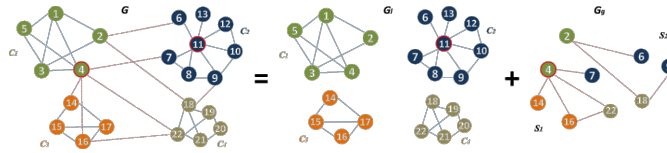20 Return the map M

Figure 2. Pseudo code [8].

Figure 3. A toy example representing the *Local network* (G_l) and the *Global network* (G_g) associated to a modular network (G) made of four non-overlapping communities [8].

Subsequently *global centrality* and *local centrality* are computed for each node in graph G, the 5 most central nodes are retrieved, using a ranking metric. The tangent of the argument is used, as in [8], because it has a higher range than the argument. It is defined by:

$$\tan(\varphi(v_i)) = \frac{\beta_G(v_i^q)}{\beta_L(v_i^k)} \quad k \in \{1, \ldots, m\} \text{ and } q \in \{1, \ldots, p\} \tag{8}$$

Graphlet analysis

Using the *ESU* algorithm [18] implementation provided by the library graph-tool we analyzed *graphlets* up to $k = 4$ (Figure 4). The results, available in (Table 3), confirm our initial assumption, backed by the literature [17], of finding a prevalence of *triangle-like* motifs w.r.t. random graphs where the most prominent ones are the *line-like*.

Table 3. Results of the graphlet counting. Mean on the condition.

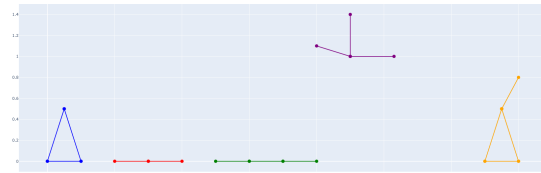| Motif	Control	ADHD	Bipolar	Schizophrenia	Random
3-line	13648	53661	52568	54598	81334
Triangle	13648	13743	13481	14236	1611
4-line	521043	538413	509534	567986	1460743
Tri-branch	236878	242328	230094	255204	703604
Branching Triangle	380948	387710	364426	433006	121449

Figure 4. From left to right: Triangle, 3-node line, 4-node line, Tri-branch, Branching Triangle

Results

The overall results showed in (Table 1) show slight changes in mean between the conditions and a very different setup for the random graphs.

To emphasize the differences among conditions a t-test was performed between the control group and the others.

The p-values shown in (Table 2) were computed on the t-statistic, these values show the likelihood of finding a mean difference by chance, considering that the graphs display the underlying structure of a brain's connectome. A lower value suggests that the observed differences are not due to a random behaviour but reflect a variation given by the neuropsychiatric disorder.

A notable value can be noticed in the *modularity*, as well as in *closeness*, *global efficiency* and *modular closeness*, for *ADHD* and *bipolar* disorder; this is in accordance with different studies in the literature that found a different modular organization in *ADHD*, where the brain has a large number of processor areas that are bottlenecked in sequential processing [2]; for bipolar disorder studies found more diffuse long ranging connections [16] [2], this could be explained by an higher *top_5_closeness* and *modular closeness* value whose p-values are still fairly low.

An interesting result comes from comparing the graphlets counts between the *control* group and the *schizophrenia* one, all graphlets have an higher mean, this could be related with the fact that this condition is often associated with increased connectivity [15].
It is also possible to notice a lower mean for the *triangle-like* graphlets in *bipolar* disorder w.r.t. controls, despite many works finding abnormal connectivity patterns and changes in functional connectivity (the one that was considered in creating the graphs) [2], in literature no graph analysis on motifs for *bipolar disorder* was available to validate the results.

Conclusions

In summary, our work found evidence of structural and modular modifications in the brain network of diseased subjects. Other than the results discussed previously and validated by low p-values, it was found that the networks of *schizophrenics*' brains show less small-worldness and abnormal modularity, as discussed in the literature [12], although our p-values do not find this results significant.

The pipeline developed for this project is easily adaptable for all f*MRI* datasets in *.nii.gz* format. It offers a basic validation of the results using random graphs while enabling a broad analysis of graph properties. Our results confirm that network models are well suited to analyze the integrity of the functional brain in *neuropsychiatric* disorders, while highlighting how different diseases are concerned to different changes picked up by different metrics, graph theory can aid practitioners both during the onset of the disorders than during treatment to track the development of the disorder.

Future Work

A future focus could be working with more *fMRIs* of the same subjects taken at different times to map the changes in the graphlets (since literature is scarce on this topic), other interesting paths would be following different subjects treated with different drugs to explore their impact in preserving the network or working with temporal graphs to make use of the temporal information extracted by the *fMRI*.

In conclusion, there is a growing recognition of the need for more detailed and nuanced datasets in the study of mental illness, particularly pertaining to the advancement of the patient's mental health state and patient's age. The utilization of more granular datasets can offer valuable insights into the intricacies of the progression of mental illnesses and their precise impact on the human brain.

Code Availability

All the code is available at this link, the Jupyter Notebook offers an interactive and fast analysis of the graphs using pre-computed data, but also offers the possibility to re-create the graphs from the fMRI data, all figures and tables are replicable from scratch.

Work contributions

Christy Jo Manthara: motif counting (cycle).
Francesco Pio Monaco: data download and time-series extraction, graph creation via correlation and adaptive threshold, computation of local and global graph metrics, significance of graph metrics (z-scores, p-values, t-tests), motif counting (cycle), code parallelization, analysis of the results.
Michele Russo: computation of local and global graph metrics, motif counting (plotting, cycle), detection of communities, analysis of each community, graph plots, analysis of the results, modular centralities, significance for modular centralities.

Fraction of work:
Christy Jo Manthara: 6%
Francesco Pio Monaco: 47%
Michele Russo: 47%

References

1. AARON CLAUSET, J.NEWMAN, C. M. Finding community structure in very large networks.

2. BARTTFELD, P., PETRONI, A., BÁEZ, S., URQUINA, H., SIGMAN, M., CETKOVICH, M., TORRALVA, T., TORRENTE, F., LISCHINSKY, A., CASTELLANOS, X., ET AL. Functional connectivity and temporal variability of brain connections in adults with attention deficit/hyperactivity disorder and bipolar disorder. *Neuropsychobiology 69*, 2 (2014), 65–75.

3. BASTIANI, M., SHAH, N. J., GOEBEL, R., AND ROEBROECK, A. Human cortical connectome reconstruction from diffusion weighted mri: The effect of tractography algorithm. *NeuroImage 62*, 3 (2012), 1732–1749.

4. DEVELOPERS, N. Networkx documentation. https://networkx.org/documentation/stable/reference/algorithms/generated/networkx. algorithms.community.modularity_max.greedy_modularity_communities.html#rce363827c0a4-2, 2023.

5. FATTAHI, S., AND SOJOUDI, S. Graphical lasso and thresholding: Equivalence and closed-form solutions. *Journal of machine learning research* (2019).

6. FORNITO, A., ZALESKY, A., PANTELIS, C., AND BULLMORE, E. T. Schizophrenia, neuroimaging and connectomics. *NeuroImage 62*, 4 (2012), 2296–2314. Connectivity.

7. FOSTER, J. G., FOSTER, D. V., GRASSBERGER, P., AND PACZUSKI, M. Edge direction and the structure of networks. *Proceedings of the National Academy of Sciences 107*, 24 (2010), 10815–10820.

8. GHALMANE, Z., EL HASSOUNI, M., CHERIFI, C., AND CHERIFI, H. Centrality in modular networks. *EPJ Data Science 8*, 15 (2019), 15. Open Access.

9. HAGMANN, P., THIRAN, J.-P., JONASSON, L., VANDERGHEYNST, P., CLARKE, S., MAEDER, P., AND MEULI, R. Dti mapping of human brain connectivity: statistical fibre tracking and virtual dissection. *NeuroImage 19*, 3 (2003), 545–554.

10. IRIMIA, A., AND VAN HORN, J. D. The structural, connectomic and network covariance of the human brain. *NeuroImage 66* (2013), 489–499.

11. LATORA, V., AND MARCHIORI, M. Efficient behavior of small-world networks. *Phys. Rev. Lett. 87* (Oct 2001), 198701.

12. MICHELOYANNIS, S. Graph-based network analysis in schizophrenia. *World J Psychiatry 2*, 1 (2012), 1–12.

13. NEWMAN, M. E. J. Networks: An introduction.

14. NEWMAN, M. E. J., AND GIRVAN, M. Finding and evaluating community structure in networks. *Phys. Rev. E* (2004).

15. OLEJARCZYK, E., AND JERNAJCZYK, W. Graph-based analysis of brain connectivity in schizophrenia. *PloS one 12*, 11 (2017), e0188629.

16. PALANIYAPPAN, L., AND COUSINS, D. A. Brain networks: foundations and futures in bipolar disorder. *Journal of Mental Health 19*, 2 (2010), 157–167. Accessed through SciHub.

17. SPORNS, O., AND KÖTTER, R. Motifs in brain networks. *PLoS biology 2*, 11 (2004), e369.

18. WERNICKE, S. Efficient detection of network motifs. *IEEE/ACM transactions on computational biology and bioinformatics 3*, 4 (2006), 347–359.

19. WU, T., WANG, L., CHEN, Y., ZHAO, C., LI, K., AND CHAN, P. Changes of functional connectivity of the motor network in the resting state in parkinson's disease. *Neuroscience Letters 460*, 1 (2009), 6–10.