

Network Science Project:

Replication of

An Edge-Centric Perspective on the Human Connectome:

Link Communities in the Brain,

by M. A. de Reus, V. M. Saenger, R. S. Kahn and M. P. van den Heuvel

(2014) Philosophical Transactions of the Royal Society B

DOI: 10.1098/rstb.2013.0527

Student: Yee Sern Tan

Semester: Fall, 2016

Introduction

To make this project of replication on a neuroscience worthwhile, it is appropriate to get the relevant background of connectomics, for which I turn to [1]. Therein mentioned is that the properties for an ideal brain graph to take account of:

- Spatial embedding
- Heterogeneity of node properties
- Weight of connectivity
- Directionality of connectivity
- Heterogeneity of edge types
- Dynamic changes in network organization over time

Each of these properties present added complexities. However, efforts to satisfy even several of these at the same time are daunting, and are mostly limited to small animals.

Analyzing at the macroscale appears at first to be a gross simplification, requiring parcellating the brain into disjoint regions. Nonetheless such simplification enables the mapping of connections across long distances, as is required for examining integrative brain features, and becomes practical with currently available imaging technologies, especially when non-invasive ones are required, either for ethical reasons or for longitudinal studies. The main technique for structural imaging at this scale is diffusion MRI, having the data modeled with methods such as diffusion tensor imaging and diffusion spectrum imaging. This having said, advances in larger, macroscales provide better direction for finer scales to proceed, and are in turn benefited by preliminary knowledge of microscale connections that form the substrate of macro-analyses. Further, the choice of macro scale yields more statistically averaged, stable results that is convenient for analysis, in contrast to microscale where changing forces such as plasticity and adaptability makes the graph dynamic. The choice of scale needs prominent mention for this paper, since the evaluation of the number of communities and the community-related metrics such as participation coefficient are closely tied to them.

The underlying parcellation is the Desikan-Killiany atlas (into 68 gyral-based neuroanatomical brain regions) [2], and the data is obtained with imaging for T1-weighted diffusion MRI. The cortical surface is then parcellated according to the atlas. The result is thresholded into a binary graph. The graph under study is obtained after averaging the results over 50 adult healthy participants, by including only edges that are present in 60% of participants. Since this is the only study known to have studied edge-centric properties of brains at this scale, it would be impossible for a cross-

validation study of edge-centric property of connectomes, for comparison with and validation of this brain atlas.

Note: the figures and illustrations for this project come embedded within the IPython Notebook file.

Edge-communities

Given that parts of the brain serve various functions, it is natural to consider them as participating in multiple communities. This paper works on community of edges, representing an edge-centric perspective by making each edge to exclusively belong to one community. To initiate community finding among edges, the line graph is first computed. The nodes of the line graph are the edges of the original graph, the edges in the line graph are present where two edges the original graph share a node. To give equal weighting for each stub in the original graph, the line graph has edges weighted, as follows: an edge, between node 1 and node 2 of the line graph, translating to edge 1 and edge 2 of the original graph, is weighted $1/(d-1)$, where d is the degree of the node at the intersection between edge 1 and edge 2 in the original graph.

Community detection is done by Rosvall and Bergstrom's Infomap algorithm [3]. The algorithm gives an information-theoretic minimization of entropy required in describing the edges of the network and their connectedness. Although the Infomap works with directed graphs as well, the underlying graph is undirected since directionality in structural brain graphs requires telling apart cell-body from synapse. The software for running this algorithm is obtained from Rosvall's website [4]. Considerable effort has been made to obtain the community that agrees with the paper. Several parameters have been experimented on, in order to identify the convergent community. The algorithm is probabilistic, so different runs can result in different results. It is found that convergence requires a number of iterations that is at the order of hundreds. A file merging program confirms this convergence [5]. The confirmation and identification of communities is done manually by computing the sizes of communities and of their pairwise intersections, with reference to the graphs in Figure 2 of the paper.

To replicate the result on consistency (of community assignments) in the paper will require the brain graphs of each of the 50 participants, but this could not be obtained due to data ownership reasons.

Robustness to edge removal

Robustness to removal is assessed with 5 measures: edge betweenness, path length, clustering coefficient, local communicability, and global communicability. The simulation for robustness is by removing single edges, that represent lesions on single axonal tracts. Edge betweenness, here, can be viewed as a baseline for other metrics. The resultant metric for all other measures are normalized thus:

$$\text{processed_result} = (\text{removal_metric} - \text{original_metric}) / \text{original_metric},$$

such that the average is 1. (Some simplification in computation is applied for all measures except local communicability in my code to take advantage that the original metric is always the same.) To clarify on some definitions, the global communicability metric is defined as the effect of removing an edge on the average of communicability over all node-pairs of the whole network, and the local communicability metric is defined as the effect of removing an edge on the communicability of the node-pair joined by the edge. Each metric for each edge is then aggregated to a statistic corresponding to its community, and the values are reported in a box plot.

Average shortest (or characteristic) path length maps the path for which neuronal signal transmits most rapidly, least metabolism-consuming, and with least noise interference. Implicit in the theory that neuronal signals travel by the shortest path is the assumption that they have global information, which is to-date unexplained, although the small-world network suggesting such a connection is observed in connectomes of *C. elegans* and other brain networks in all micro-, meso-, and macro-scales [6].

The measurement of edge betweenness gives equal share among all pairs of nodes, and calculates the proportion of shortest-path lengths that have the edge as a constituent part.

Clustering coefficient, the proportion of closed triangles among the total of both closed and open triangles, of particular nodes, is instrumental in the study of motifs. It is found that the apical (open-triangle) motif occurs higher than random. Further, for directed graphs, apical motifs where the apex shares a reciprocal connection with at least one other node, have been found to appear at frequencies above expected ones for random graphs. The presence of these motifs is suggested to facilitate zero-lag synchronizing of neuronal signals [7].

Communicability is a generalization of the shortest path, that takes into account the possibility of longer paths, but penalized with lower weights. It is consistent with the model of diffusion spreading, where each node spreads parallelly to multiple of its neighbors with equal (or proportionate, for the weighted case) probability, and having the signal strength decrease at the order of the factorial of time. Diffusion is interesting, partly because it explains functional connectivity with structural connectivity. The communicability matrix defines a communicability-weighted network. In [8], the accuracy of detecting lesions using communicability is evaluated.

Some minor discrepancies are present between the box plots obtained and that presented by the original authors in Figure 5 of the paper. This could be due to different standards of box plot, which is not mentioned in the paper. In spite of this, the major shape and trends do agree.

Participation number

In the context of communities, the 2D plots of z-score versus participation coefficient have been outlined by [9], in the analysis of a wide variety of graphs, to have 7 attractor basins. Since the paper I replicate focuses on edge-communities, a related metric, the participation number of a node, is defined as the number of edge-communities the node is adjacent to. The distribution of this participation number is reproduced from the paper. The brain's hub regions, putatively comprising of the left/right superior frontal, superior parietal, precuneus and insula, is also found to be richly connected to many communities. The plots of how many hubs a community is adjacent to, and of the participation number of hubs and non-hubs in relation to their degrees, have been successfully replicated.

The presence of hubs and communities together represents an interesting study, in that they represent the functional integration and specialization properties in the brain. Tononi, Sporns and Edelman [10] formulated the concept of degeneracy that models the complex behavior of brain dynamics. These forces are abstract generalizations, and can be measured with mathematical quantities such as the Jaccard coefficient or cosine similarity. A well-formed study will require further hypotheses testing, and the statistics obtained will need sound interpretation.

Extensions

Studies have differing opinion on whether the degree distribution of brain networks follow scale-free [11] or log-normal [12] distribution. Not much can be expected from obtaining the degree

distribution of this network, since there are only 68 nodes, and the connections heavily show spatially embedded property, obscuring the mechanisms that possibly can generate these two types of distributions.

A further extension of this study, as highlighted by the authors, is to work on a weighted graph. The authors point out there are various measures on the weights that can be applied, such as streamline count, without a dominant rule/consensus. The selection of a threshold to binarize a graph will inevitably result in loss of quality, as a tradeoff with simplicity. It has been reported that connection strengths (edges) detected between regions (nodes) can vary by five orders of magnitudes. High thresholds will result in more false-negatives, while low thresholds will result in more false-positives. For our brain graph, the average degree of nodes is $11.3 > 2$, therefore the resultant line-graph is much more complex than the original, resulting in the appearance of edge-communities.

Acknowledgments

I am indebted to the authors of the paper I replicate, for providing one such alternative view on connectomics, and for providing the brain graph on the Dutch Connectome Lab website [13]. Special thanks to M. A. De Reus who helped me in identifying the labels for Desikan-Killiany atlas. The availability of open-source code and software, and of freeware has made this replication project possible: MapEquation [4], WinMerge [5], GCC Runtime Library, MinGW, MSYS and SWIG.

References

1. A. Fornito, A. Zalesky, and E. Bullmore, Fundamentals of Brain Network Analysis, 1st Edition (2016, Elsevier), Chapters 1-9.
2. R. S. Desikan, F. Segonne, B. Fischl et al., An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral-based regions of interest, (2016) NeuroImage 31, 968-980, doi:10.1016/j.neuroimage.2006.01.021.
3. M. Rosvall and C. T. Bergstrom, Maps of random walks on complex networks reveal community structure, (2008) Proceedings of the National Academy of Sciences 105(4), 1118-1123, doi:10.1073/pnas.0706851105.
4. Website: <http://www.mapequation.org/code.html>.
5. Software: <http://winmerge.org/?lang=en>.
6. E. Bullmore and O. Sporns, The economy of brain network organization, (2012) Nature Reviews Neuroscience 13(5), 336-349, doi:10.1038/nrn3214.
7. R. Vicente, L. L. Gollo, C. R. Mirasso et al., Dynamical relaying can yield zero time lag neuronal synchrony despite long conduction delays, (2008) Proceedings of the National Academy of Sciences 105, 17157-17162, doi:10.1073/pnas.0809353105.
8. J. Andreotti, K. Jann, L. Melie-Garcia et al., Validation of network communicability metrics for the analysis of brain structural networks, (2014) PLoS One 9(12), e115503, doi:10.1371/journal.pone.0115503.
9. R. Guimera and L. N. Amaral, Functional cartography of complex metabolic networks, (2005) Nature 433, 895-900, doi:10.1038/nature03288.
10. G. Tononi, O. Sporns and G. M. Edelman, Measures of degeneracy and redundancy in biological networks, (1999) Proceedings of the National Academy of Sciences 96, 3257-3262, doi:10.1073/pnas.96.6.3257.
11. D. S. Bassett, D. L. Greenfield, A. Meyer-Lindenberg et al., Efficient physical embedding of topologically complex information processing networks in brains and computer circuits, (2010) PLoS Computational Biology 6, e1000748, doi:10.1371/journal.pcbi.1000748.

12. S. Achard, R. Salvador, B. Whitcher et al., A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs, (2006) Journal of Neuroscience 26,63-72, doi:10.1523/jneurosci.3874-05.2006.
13. Dutch Connectome Lab, data source: http://www.myconnectome.nl/wordpress/wp-content/uploads/grouped_umcu50_60percent.mat.