
REPRODUCIBILITY IN NETWORK NEUROSCIENCE: A CASE PRACTICE AND SOME THOUGHTS

NEUR 602 COURSE PROJECT PAPER

Zhen-Qi Liu
McGill University
zhenqi.liu@mail.mcgill.ca

December 4, 2018

ABSTRACT

Reproducibility in neuroscience has been put into spotlight in recent years, however, not all aspects are covered in literature. In this project, we tried to address some important aspects of reproducible research in network neuroscience through a hand-on practice. We reproduced some key findings in a published paper, recorded the challenges and problems we met, and proposed some solutions. The overall results matched original paper, with numeral derivations. Critical issues are raised about what should be done particularly in the context of network neuroscience. The results of this project will also be available online for reproducibility.

Introduction

Reproducibility in research is gaining increasing attention and has progressed dramatically in the last ten years, when modern technology enabled exponentially growth of opportunities and new approaches [1]. Reproducibility in neuroscience, in particular, has been a central topic of the community, attentions are drawn into general reproducibility issues, data publishing, data sharing, experiment pre-registration, quantitative analysis, statistical pitfalls and more. However, issues coming from all sub-disciplines deserve some examination. With its complexity in research paradigms, reproducibility and its importance in neuroscience need no more emphasis, only from existing studies has revealed enough problems that needs years of work to elucidate.

Among different sub-areas of neuroscience, neuroimaging mostly utilizes physical equipment and computational methods. The reproducibility of it are taken for granted to be better than traditional psychological experiments, however, numerous researches have pointed out that it is not the case [2, 3]. We still need to carefully examine every step that we have taken.

The context of network neuroscience [4] is more of a recent thing. Brains can be seen to function as an interconnected system spanning a range of scales. The mesoscale organization principles are not immediately available thorough either macroscopic characterization or microscopic biochemical reactions, and require dedicated theoretical tools [5]. Using notions from network science and complex systems to study the connectivity inside the brain [6], network neuroscience has generated particularly insightful findings [7].

For simplicity, this class paper did not elaborate on methods and definitions, which are available in the original paper, and the implementation is also available for download. It rather tries to show an aspect to key processes through the reproducing practice, and to propose some quick solutions.

Results

Choosing Project

The project starts with choosing a proper project for the course. Selection criterion include first filtering from a wide range of network neuroscience articles for suitable and interesting topic. Five articles [8, 9, 10, 11, 12] stand out because of the best practices used in a range of topics including rich club, motifs and general graph-theoretical analysis. Then we examined the data and code availability. Unsurprisingly, none of the five articles chosen from initial round clearly stated the source of implementations, which means all the procedures have to be re-implemented according to the methods sections. To the writer's knowledge, it is not common for network analysis researches to have their original analysis code available for reproducing. We checked for raw data as a vital step to decide the possibilities of reproducing the article. It is astonishing to find out that two out of the five articles did not linked to a open data. The other three articles show some extent of data sharing of either using open data, or attached connectivity matrices, but they either missed important part of data (distance matrix), or did not have a functioning data source. For [8], we accidentally found the original data on the [author's website](#), and finally decided to use it as primary target for this project. A final overview of the methods sections of the articles show generally good details, while there are still substantial differences between researches, with some of the articles elaborating on every detail and others not. The details are summarized in box 1.

Acquiring Data

Data downloaded from [this link](#) is described by the link text as 'Group connectivity of streamline density and lengths. File contains principal (1000+ resolution, n=40) and replication dataset (n=40) [Data from van den Heuvel MP, Kahn RS, Goni J, Sporns O | Proc Natl Acad Sci U S A. | 10-07-2012 | High-cost, high-capacity backbone for global brain communication]'. It contains 2 sets of matrices, each with connectivity matrix (shown in figure 2) and distance matrix (shown in figure 3) of 1170 dimensions. To our knowledge, there is no further metadata attached to this data release.

Implementation Options

As there are generally no available code of its kind, we reimplemented the procedures using some basic graph analysis functions from [Brainconn](#), which is a python version of Brain Connectivity Toolbox [13]. The implementation tries its best to follow descriptions in the methods section of the article, however, there could still be variations from basic algorithm and null model implementations.

Results I: Rich Club Detection

Rich club detection is the first and the most vital step of reproducing this research, as it is the foundation of the analyses afterwards, and decides the key parameter of rich club level. As shown in figure 4, major trend seems a perfectly fit between original figure and reproduced result, the peak at $k = 10$ and the plateau afterwards, which are not trivial, are captured precisely. The minor shifts in values are thought to be caused by differences in null models. The calculation of normalized rich club ratio involves comparing the metric from the original network to a set of random networks for significance. The number and randomness of the null models will always affect the actual result.

Results II: More Findings

Connection types grouped by length

Rich club connections tend to be long-ranged in physical distance, illustrated by network connection types grouped by different lengths shown in figure 5a. The results in box 5b are reproduced to the extent that main conclusions match. There are differences in values possibly as the result of implementations details, but quite hard to identify the exact step of cause.

Network density and communication cost

Rich club connections are low in density but more costly, showing importance in network communication represented by figure 6a. The result in box 6b is not reproduced original description from methods section, but similar results are acquired with a simplified version of interpretation. The difference may be caused by randomness from algorithm and implementation details.

Path motifs

Interesting results (shown in figure 7a) emerge from network path motifs, showing communication patterns of LFRFL, LFL, etc. In reproducing the results, we find that main conclusions match as in figure 7b, and there are shifts in rankings shown in figure 7c and 7d. The cause could be randomness from the shortest path algorithm and other implementation details.

Replication dataset

The article used a second dataset for validation of the results. We also did the same procedure on the replication dataset and found the same significant results. For results from the replication dataset, simply run the scripts.

Discussion

Without any doubt, [8] is a great paper in network neuroscience, elaborating on rich club detection and the importance of hub nodes in network communication, supporting the long-existing hypothesis of cost-efficiency trade-off in the brain. Still, questions can be raised to improve the reproducibility issues.

Challenges in reproducing the results

Data publishing The connectivity matrix often plays the central role for reproducing virtually any research in network neuroscience. One equally important fact is distance matrix, which is contributing to many vital conclusions, and even harder to acquire because sometimes it relies on human expertise to determine the coordinates.

Preprocessing methods Normally, the connectivity matrix and coordinates of the network are acquired through the MRI preprocessing stage. There are quite vulnerable to flexibility through the parameter settings, and currently there has not attracted enough attention. Although not involved in this project, the preprocessing procedures in network neuroscience studies are the most hard part for any reproducing attempt. Researches are describing ‘analytical flexibility’ in fMRI processing[2, 3]. There are also various thresholding techniques to ‘make’ the connectivity matrix. The connectivity matrix used in this project is more sparse than normal, and one possible cause is thresholding and group level selection of connections. It remains a doubt whether the results can be significantly replicated if these parameters are changed.

Proposals

- Standard description for data acquisition, data processing, etc.
- Replicate results with different data processing choices, if possible.
- All network articles should at least release dataset of connection matrix, distance matrix, coordinates and metadata related.
- Make code available in non-standard/complex/multi-stage/custom analyses

Data and Code Availability

This project tries to follow file organization from [TIER Protocol 3.0](#) to facilitate reproducing this research. Code and data are available on Github repository [reprocourse-final](#).

Acknowledgement

I would like to thank Professor Jean-Baptiste Poline for the great course and help with this project. I also want to thank Dr. Yasser Iturria-Medina for helpful discussion and valuable suggestions regarding this project.

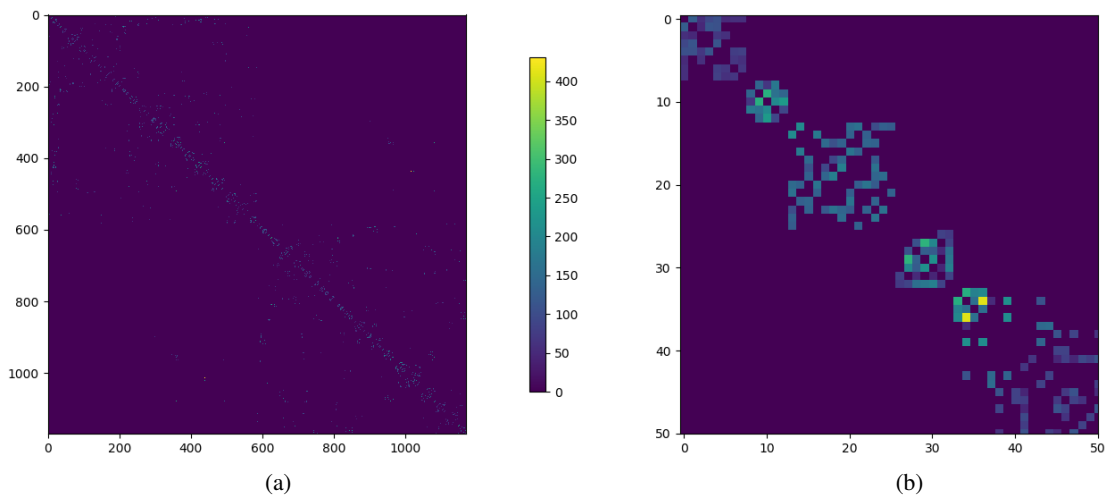
References

- [1] Keith A. Baggerly and Kevin R. Coombes. Deriving chemosensitivity from cell lines: Forensic bioinformatics and reproducible research in high-throughput biology. *The Annals of Applied Statistics*, 3(4):1309–1334, December 2009.
- [2] Joshua Carp. On the Plurality of (Methodological) Worlds: Estimating the Analytic Flexibility of fMRI Experiments. *Frontiers in Neuroscience*, 6, 2012.
- [3] Alexander Bowring, Camille Maumet, and Thomas Nichols. Exploring the Impact of Analysis Software on Task fMRI Results. *bioRxiv*, page 285585, September 2018.
- [4] Danielle S Bassett and Olaf Sporns. Network neuroscience. *Nature neuroscience*, 20(3):353–364, February 2017.
- [5] Richard F. Betzel, John D. Medaglia, and Danielle S. Bassett. Diversity of meso-scale architecture in human and non-human connectomes. *Nature Communications*, 9(1):346, January 2018.
- [6] Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52(3):1059–1069, September 2010.
- [7] Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, March 2009.
- [8] Martijn P. van den Heuvel, René S. Kahn, Joaquín Goñi, and Olaf Sporns. High-cost, high-capacity backbone for global brain communication. *Proceedings of the National Academy of Sciences*, 109(28):11372–11377, July 2012.
- [9] Martijn P. van den Heuvel and Olaf Sporns. Rich-Club Organization of the Human Connectome. *Journal of Neuroscience*, 31(44):15775–15786, November 2011.
- [10] Logan Harriger, Martijn P. van den Heuvel, and Olaf Sporns. Rich Club Organization of Macaque Cerebral Cortex and Its Role in Network Communication. *PLOS ONE*, 7(9):e46497, September 2012.
- [11] Mikail Rubinov, Rolf J. F. , Charles Watson, and Edward T. Bullmore. Wiring cost and topological participation of the mouse brain connectome. *Proceedings of the National Academy of Sciences*, 112(32):10032–10037, August 2015.
- [12] Kelly Shen, Gleb Bezgin, R. Matthew Hutchison, Joseph S. Gati, Ravi S. Menon, Stefan Everling, and Anthony R. McIntosh. Information Processing Architecture of Functionally Defined Clusters in the Macaque Cortex. *Journal of Neuroscience*, 32(48):17465–17476, November 2012.
- [13] Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069, 2010.

Article Info	Data format	Data Availability	Method Description	Code Availability	Possibility to reproduce
Heuvel and Sporns (2011) J Neurosci	MRI connectome	None	Good	None	No
Heuvel et al (2012) PNAS	sMRI connectome	Link in the article failed Found on the author's page	Okay	None	Yes
Harriger et al (2012) PLOS ONE	Tract tracing connectome	Open data (CoCoMac) Matrix attached (No distance)	Good	None	Partly
Rubinov et al (2015) PNAS	Tract tracing connectome	Open data (Allen) Matrix attached (No distance)	Great	None	Partly
Shen et al (2012) J Neurosci	fMRI connectome	None	Good	None	No

Figure 1: Choosing project

Tables and Figures

Figure 2: **Connectivity matrix:** (a) the original (b) the first 50 dimensions.

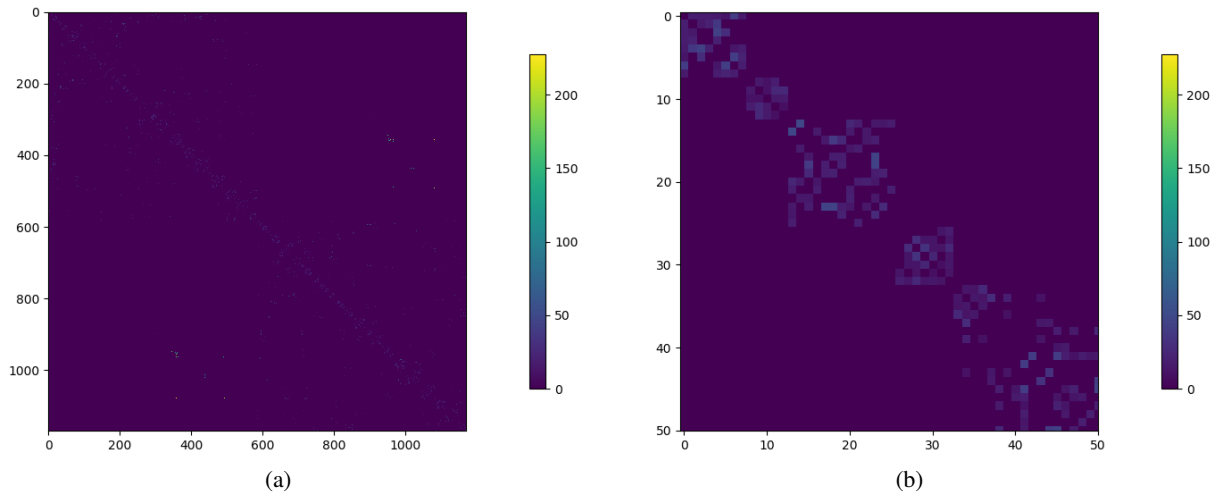


Figure 3: **Distance matrix**: (a) the original (b) the first 50 dimensions.

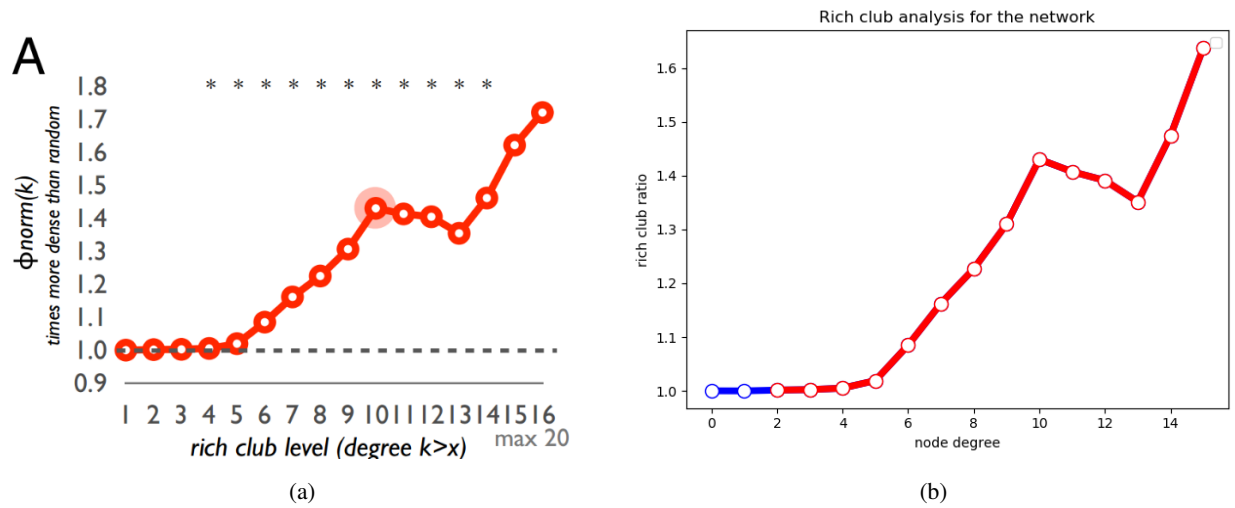
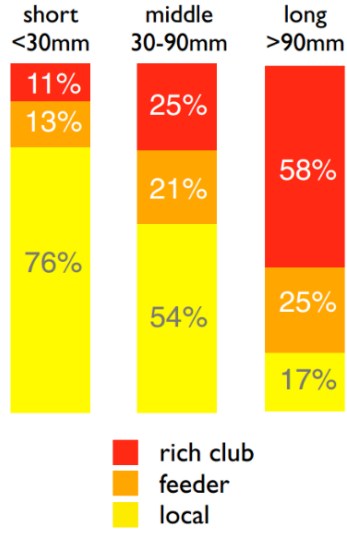


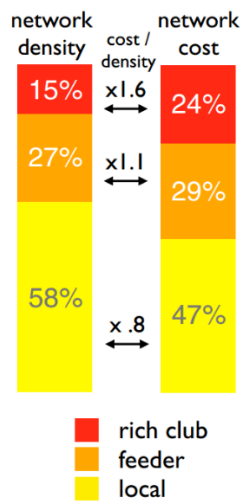
Figure 4: **Rich club detection**: (a) the original (b) the reproduced result.



(a)

Connection Length	Connection Type	Percentage
Short <30mm	local	67.022358
	feeder	23.348577
	rich club	9.629065
Middle 30-90mm	local	44.510740
	feeder	34.486874
	rich club	21.002387
Long >90mm	local	13.513514
	feeder	39.639640
	rich club	46.846847

(b)

Figure 5: **Connection types grouped by length:** (a) the original (b) the reproduced result.

(a)

Method 1	Network density	Network cost 1	Network cost 2
rich club	0.12425793	0.32317503	0.23628926
feeder	0.25629478	0.34364441	0.29132679
local	0.61944729	0.33318056	0.47238395

(b)

Figure 6: **Network density and communication cost:** (a) the original (b) the reproduced result.

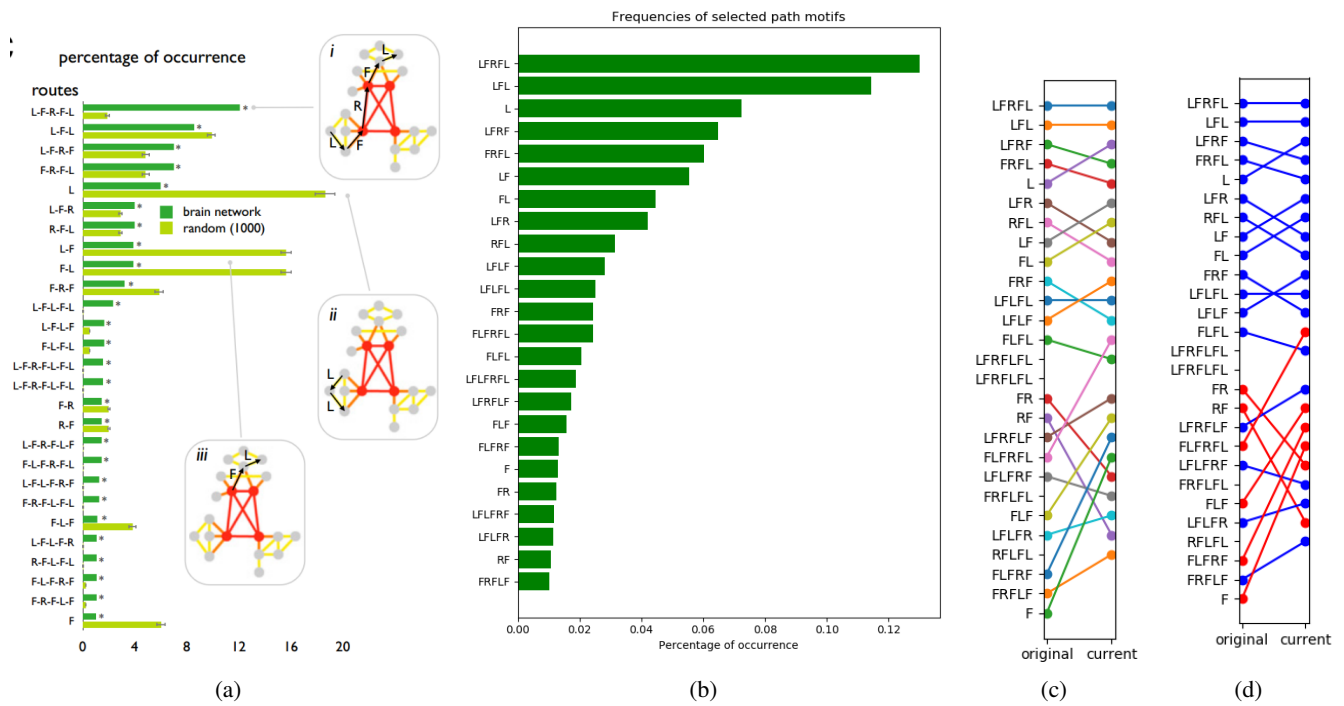


Figure 7: **Path motifs:** (a) the original (b) the reproduced result (c) comparison of ranking (d) comparison of ranking (changes less than three ranks shown in blue)