Reproducibility in Network Neuroscience: A case practice and some thoughts

Zhen-Qi Liu @ NEUR 602 Reprocourse

Overview

Reproducing a published paper !

Heuvel et al (2012) PNAS
High-cost, high-capacity backbone for global brain communication

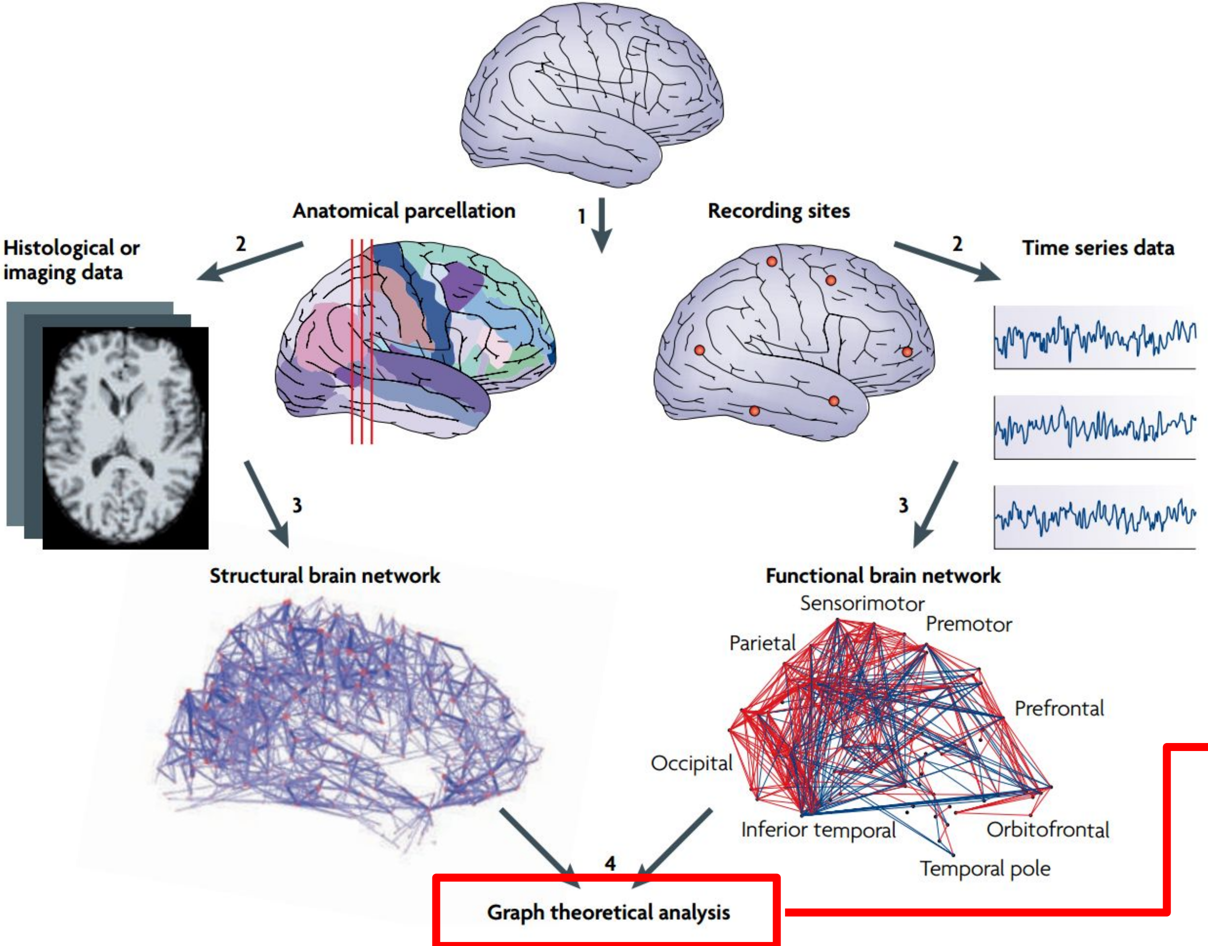
Process & Challenges

- Choosing project & Acquiring data
- Reproducing results
- Comparing & Evaluation

Conclusions & Proposals

- For network neuroscience - what’s special

Process - Network Neuroscience



Briefly described

Analytic flexibility exists!!

Starts from here

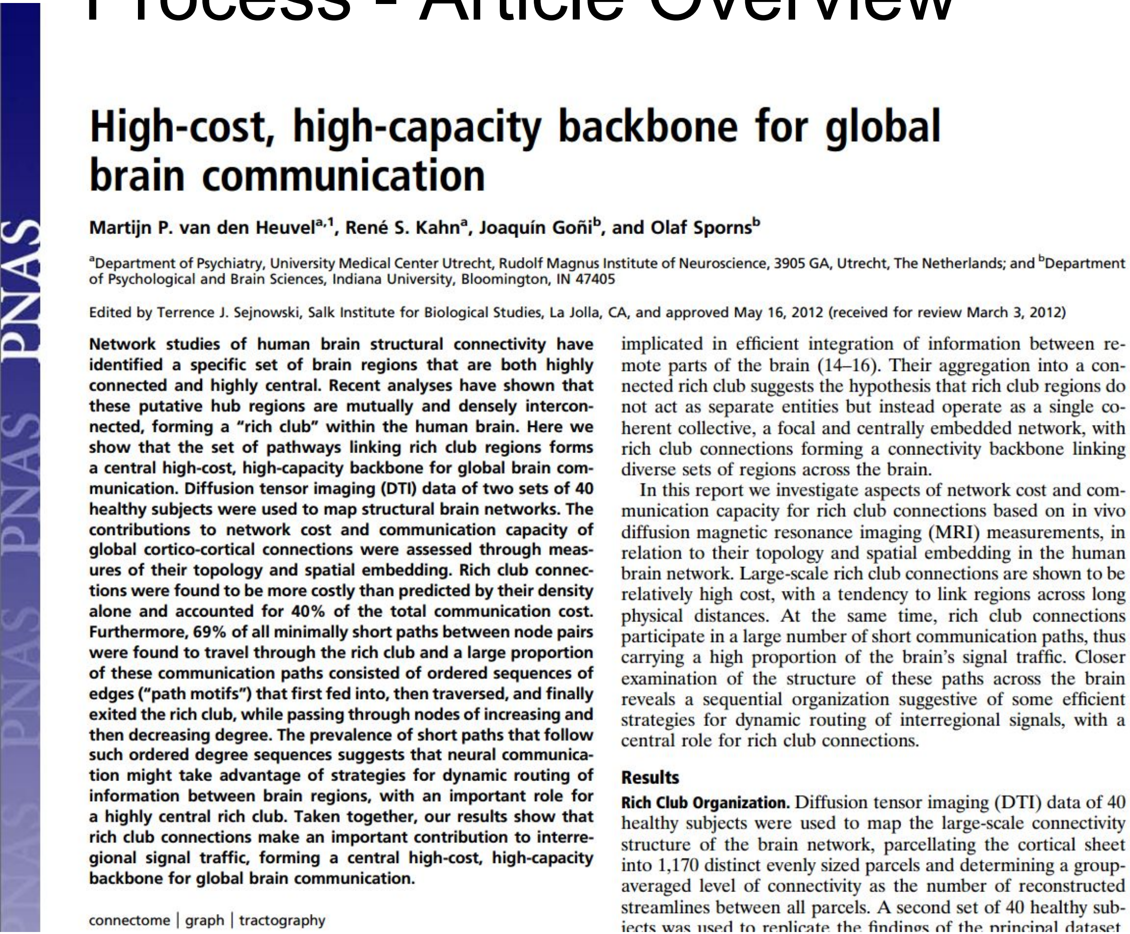
Bullmore and Sporns (2009) Nat. Rev. Neurosci.

Challenges - Choosing Project

Selection standards: topic interesting, data available, methods well-described, results straight-forward

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

Process - Article Overview



PNAS PNAS PNAS PNAS

High-cost, high-capacity backbone for global brain communication

Martijn P. van den Heuvel[a,1], René S. Kahn[a], Joaquín Goñi[b], and Olaf Sporns[b]

[a]Department of Psychiatry, University Medical Center Utrecht, Rudolf Magnus Institute of Neuroscience, 3905 GA, Utrecht, The Netherlands; and [b]Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405

Edited by Terrence J. Sejnowski, Salk Institute for Biological Studies, La Jolla, CA, and approved May 16, 2012 (received for review March 3, 2012)

Network studies of human brain structural connectivity have identified a specific set of brain regions that are both highly connected and highly central. Recent analyses have shown that these putative hub regions are mutually and densely interconnected, forming a "rich club" within the human brain. Here we show that the set of pathways linking rich club regions forms a central high-cost, high-capacity backbone for global brain communication. Diffusion tensor imaging (DTI) data of two sets of 40 healthy subjects were used to map structural brain networks. The contributions to network cost and communication capacity of global cortico-cortical connections were assessed through measures of their topology and spatial embedding. Rich club connections were found to be more costly than predicted by their density alone and accounted for 40% of the total communication cost. Furthermore, 69% of all minimally short paths between node pairs were found to travel through the rich club and a large proportion of these communication paths consisted of ordered sequences of edges ("path motifs") that first fed into, then traversed, and finally exited the rich club, while passing through nodes of increasing and then decreasing degree. The prevalence of short paths that follow such ordered degree sequences suggests that neural communication might take advantage of strategies for dynamic routing of information between brain regions, with an important role for a highly central rich club. Taken together, our results show that rich club connections make an important contribution to interregional signal traffic, forming a central high-cost, high-capacity backbone for global brain communication.

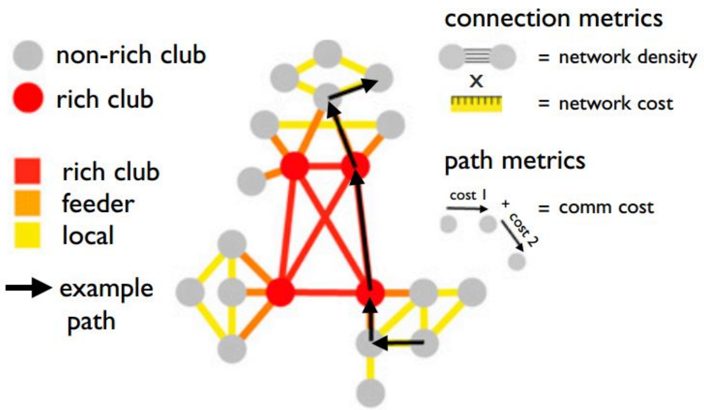
connectome | graph | tractography

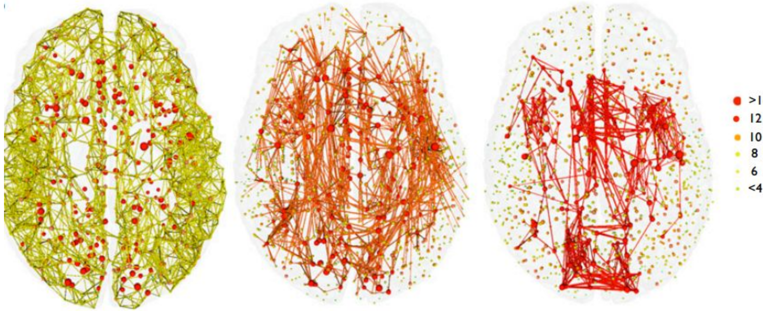
implicated in efficient integration of information between remote parts of the brain (14–16). Their aggregation into a connected rich club suggests the hypothesis that rich club regions do not act as separate entities but instead operate as a single coherent collective, a focal and centrally embedded network, with rich club connections forming a connectivity backbone linking diverse sets of regions across the brain.

In this report we investigate aspects of network cost and communication capacity for rich club connections based on in vivo diffusion magnetic resonance imaging (MRI) measurements, in relation to their topology and spatial embedding in the human brain network. Large-scale rich club connections are shown to be relatively high cost, with a tendency to link regions across long physical distances. At the same time, rich club connections participate in a large number of short communication paths, thus carrying a high proportion of the brain's signal traffic. Closer examination of the structure of these paths across the brain reveals a sequential organization suggestive of some efficient strategies for dynamic routing of interregional signals, with a central role for rich club connections.

Results

Rich Club Organization. Diffusion tensor imaging (DTI) data of 40 healthy subjects were used to map the large-scale connectivity structure of the brain network, parcellating the cortical sheet into 1,170 distinct evenly sized parcels and determining a group-averaged level of connectivity as the number of reconstructed streamlines between all parcels. A second set of 40 healthy subjects was used to replicate the findings of the principal dataset





Heuvel et al (2012) PNAS

Challenges - Acquiring Data

Media Downloads Internships Contact

Please let us know if you are looking for other resources or data.

- An MRI Von Economo - Koskinas Atlas
- Supplementary Table of our paper Bridging Cytoarchitectonics and Connectomics in Human Cerebral Cortex, J Neuroscience 2015
- Supplementary Materials of Linking Contemporary High Resolution Magnetic Resonance Imaging to the Von Economo legacy: A study on the comparison of MRI cortical thickness and histological measurements of cortical structure, Human Brain Mapping, 2015
- Right hemisphere anatomical DWI connectome data as used in Van den Heuvel and Sporns (An anatomical substrate for integration among functional networks in human cortex Martijn van den Heuvel, Olaf Sporns The Journal of Neuroscience 33 (36): 14489-14500) and by Avena-Koenigsberger et al. (Using Pareto optimality to explore the topology and dynamics of the human connectome Andrea Avena-Koenigsberger, Joaquin Goni, Richard F. Betzel, Martijn P. van den Heuvel, Alessandra Griffa, Patric Hagmann, Jean-Philippe Thiran, Olaf Sporns DOI: 10.1098/rstb.2013.0530).
- Send us an email if you are interested in the data of "Linking macroscale graph analytical organization to microscale neuroarchitectonics in the macaque connectome" Scholtens LH, Schmidt R, de Reus MA, van den Heuvel MP | J Neurosci. | 03-09-2014. [Respecting copyright reasons the data will be shared by email].
- Group-averaged binary connectivity matrix based on 50 healthy subjects. Data as used in the paper "Edge-centric perspective on the human connectome: link communities in the brain", De Reus et al. 2014, Philosophical Transactions of the Royal Society B. Oct 5;369(1653).
- Suggested further reading - Summerschool Donders Institute Lecture 15 August 2014
- 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]

http://www.dutchconnectomelab.nl/index.php#downloads

- Availability of data mentioned in the paper, but link is dead
- Found the data **accidentally** on the author's website
- No description provided, only a .mat file

Process - Data Overview

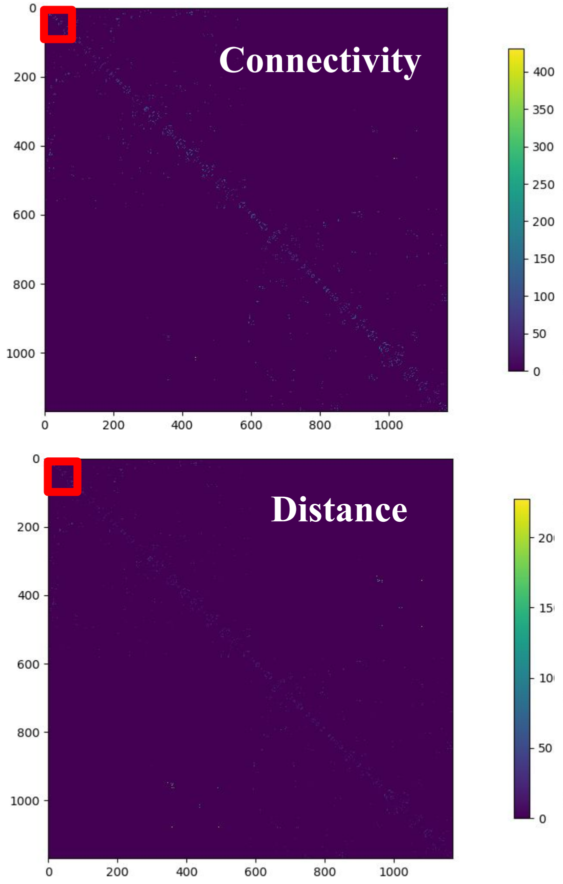
Connectivity matrix (from DTI data)
Distance matrix

1170 * 1170, **very sparse**

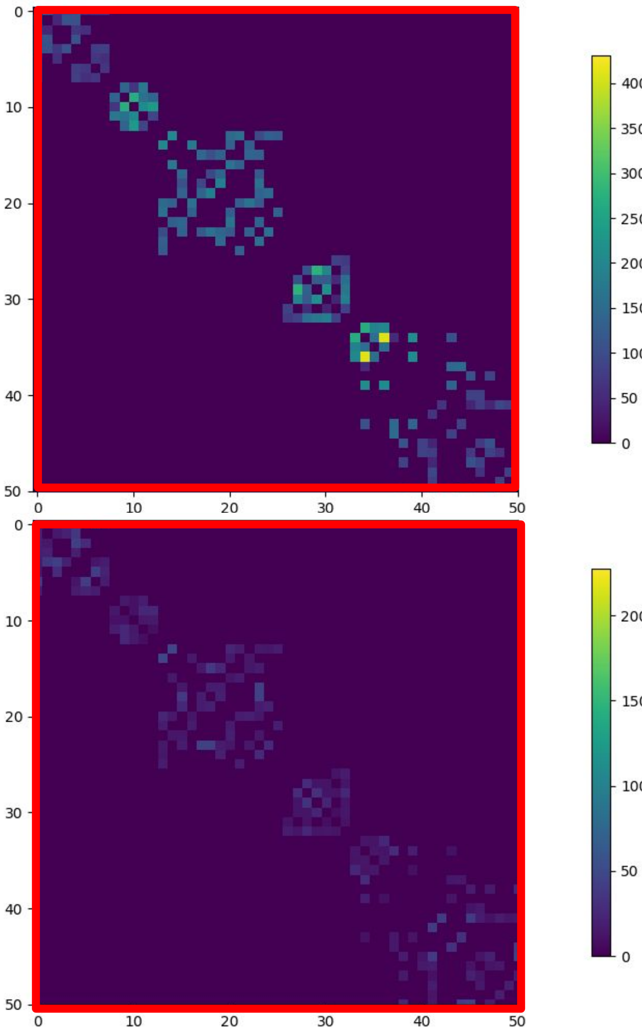
Principle and replication dataset

GR_Dataset_n2x40.mat

- **GROUP_MATRIX_HD_gr1**
- GROUP_MATRIX_HD_gr2
- **GROUP_MATRIX_HDlength_gr1**
- GROUP_MATRIX_HDlength_gr2



1170*1170 Matrix



First 50*50 dim zoomed-in

Challenges - Implementation details

Generally, there is no available code for an article of its kind.

What needs to be implemented?

- Rich club detection (key step)
- Other findings (main quantitative results)
    - Contrast network connection types grouped by different lengths
    - Compare network density and cost for connection types
    - Occurrence of path motifs

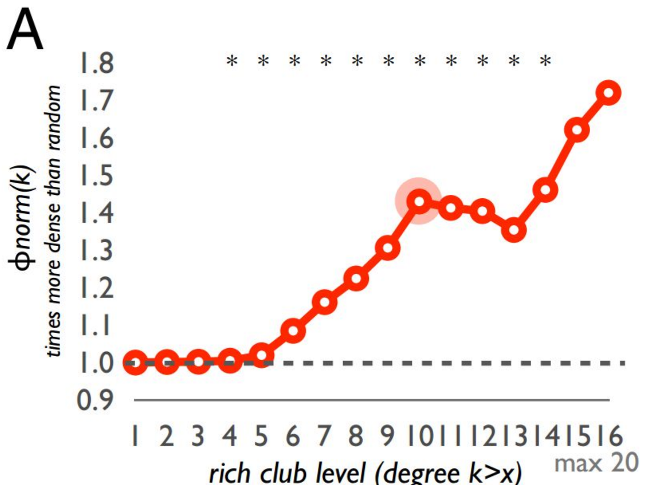
Fortunately, we have some open source implementations of basic functions.

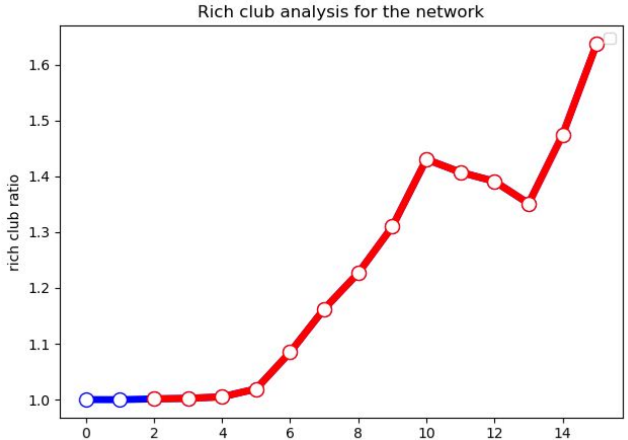
Brain Connectivity Toolbox (MATLAB)

bctpy Brain Connectivity Toolbox for Python

brainconn Brainconn (brain connectivity) library

Results (1) - Rich Club Detection





Results for rich club detection, illustrated as normalized rich club ratio to node degree.

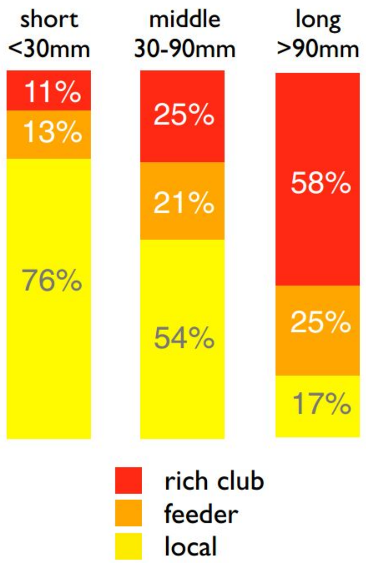
Reproduced?

- **Yes**, the results fit perfectly
- Minor shifts in node degree and ratio value

What's the cause?

- Randomness from null model

Results (2.1) - More findings



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

Rich club connections tend to be long-ranged in physical distance, illustrated by network connection types grouped by different lengths

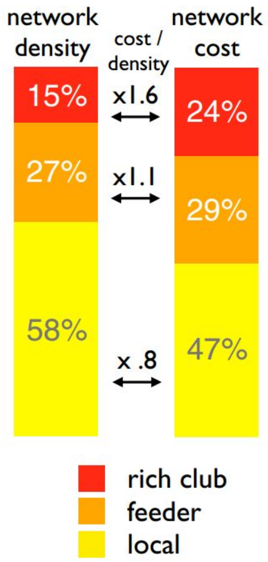
Reproduced?

- **Yes**, main conclusions match
- Minor shifts in values

What's the cause?

- Not sure, possibly implementation details

Results (2.2) - More findings



Rich club connections are low in density but more costly, showing importance in network communication.

Reproduced?

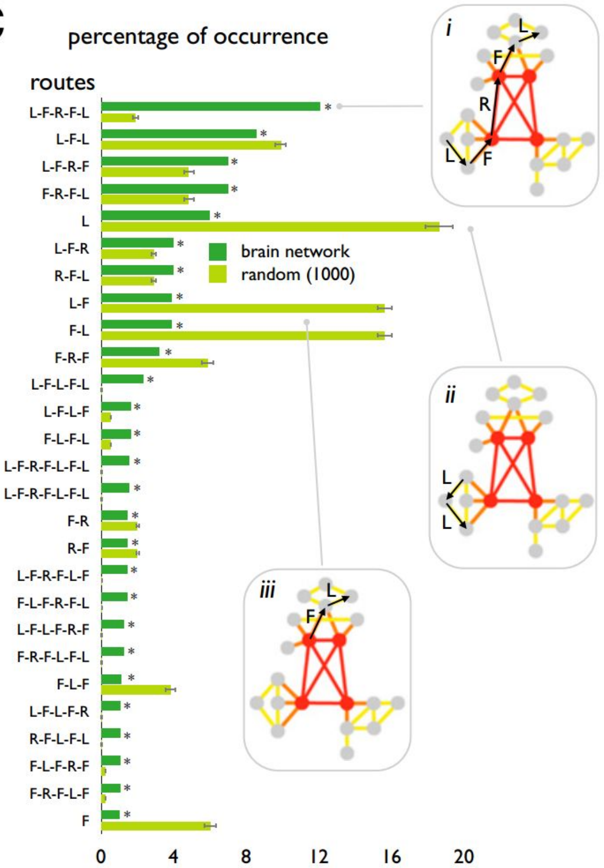
- **Not with original description** (Network cost 1)
- Similar results acquired (Network cost 2)

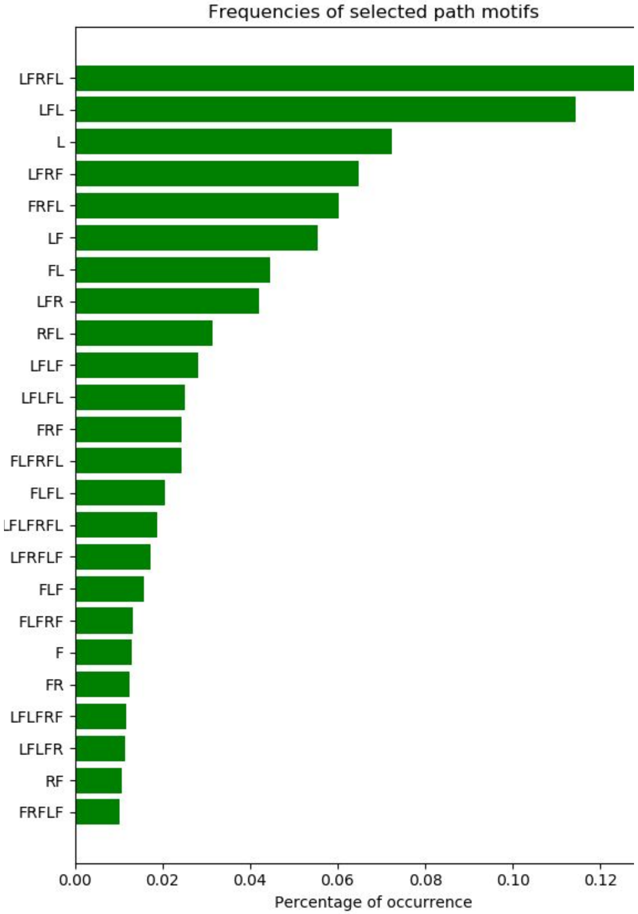
What's the cause?

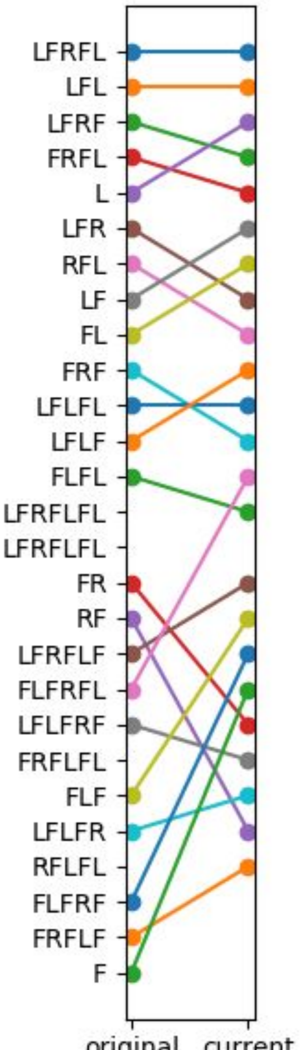
- Randomness from algorithm and implementation details

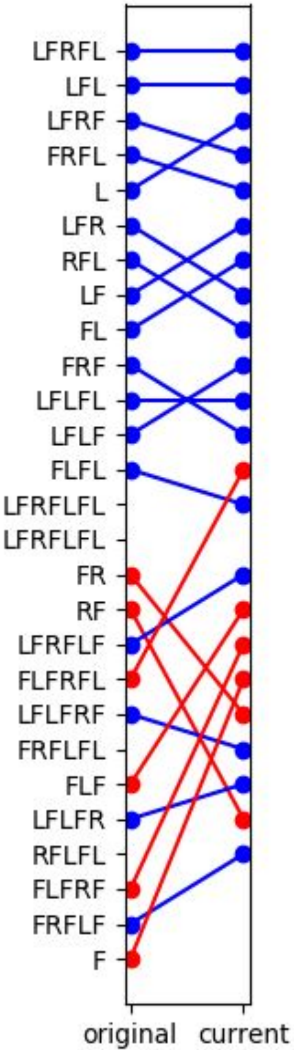
Method 1	Network density	Network cost 1	Network cost 2
rich club	0.12425793	0.32317503	0.23628926
feeder	0.25629478	0.34364441	0.29132679
local	0.61944729	0.33318056	0.47238395

Results (2.3) - More findings

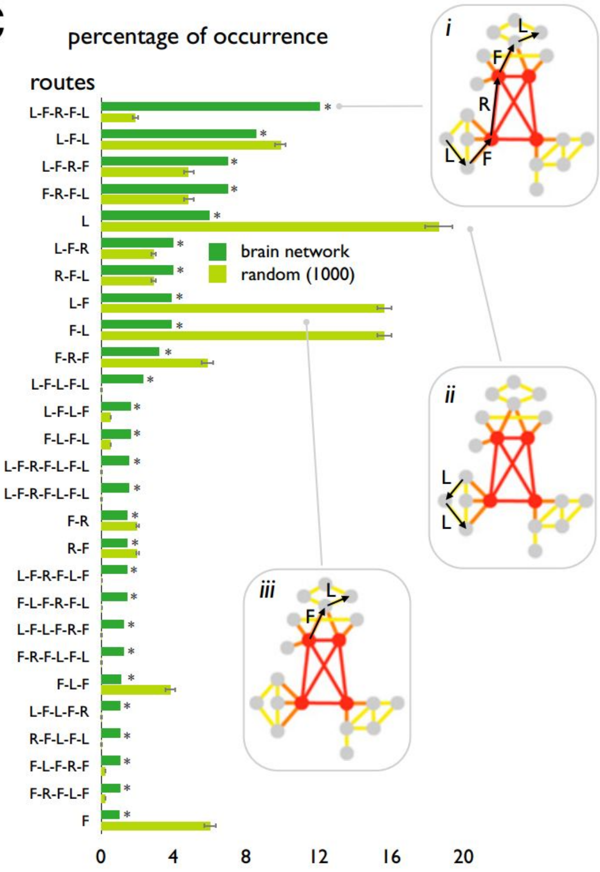


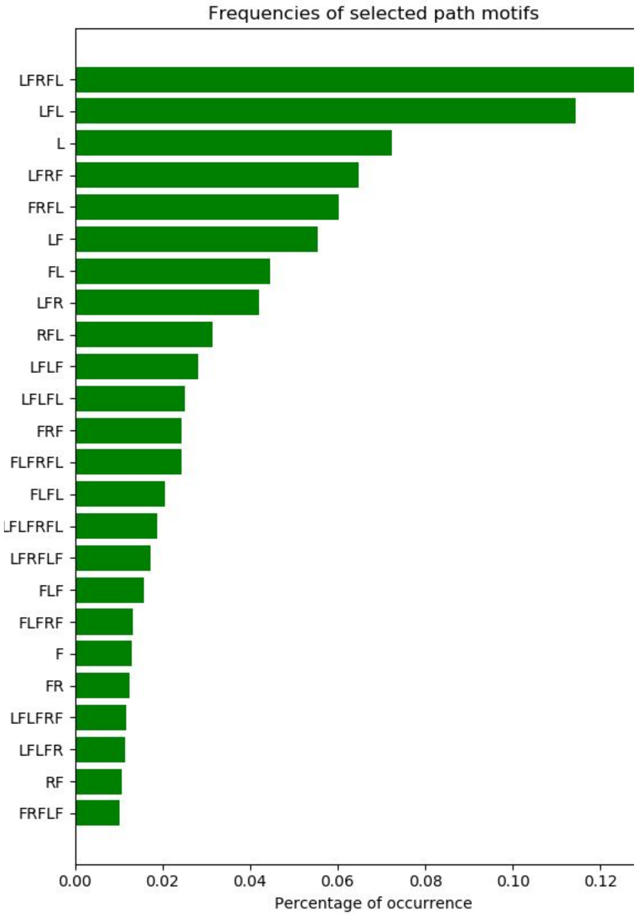






Results (2.3) - More findings





Interesting results emerge from network path motifs, showing communication patterns of LFRFL, LFL, etc.

Reproduced?

- **Yes**, main conclusions match
- Minor shifts in values

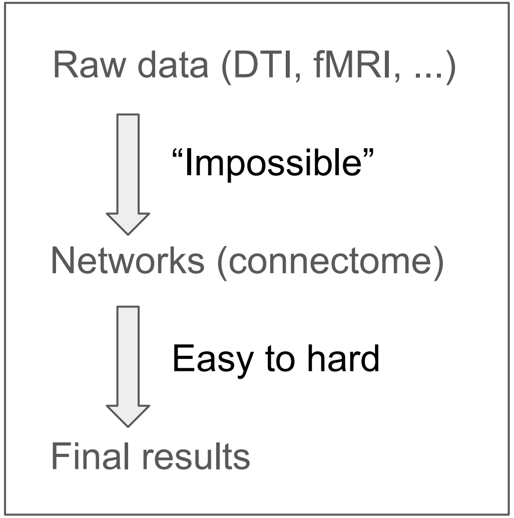
What's the cause?

- Randomness from algorithm and implementation details

Conclusions & Proposals

What can be learned?

- Reproducing an article in network neuroscience is not easy
- Results come back ***OKAY*** **(take-home message)**
    - Major findings stand well
    - Numeral variations exist as the result of…
        - Randomness from null models
        - Algorithm implementation details
        - Randomness from the procedures



What should be done?

- 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