A community-based framework for assessing whole-brain functional connectivity

Duy Duong-Tran[1,2], Jose Cadena[3], Alan D. Kaplan[3], Joaquín Goñi[1,2,4]

1 School of Industrial Engineering, Purdue University;
2 Purdue Institute for Integrative Neuroscience;
3 Lawrence Livermore National Laboratory;
4 Weldon School of Biomedical Engineering, Purdue University

DATA SCIENCE SUMMER INSTITUTE

PURDUE INSTITUTE FOR INTEGRATIVE NEUROSCIENCE — Where science meets engineering

rethink IE

Lawrence Livermore National Laboratory

Problem Statement

- One of the most challenging tasks in computational neuroscience is to *properly* estimate and assess whole-brain functional connectomes (FCs).
- A critical step concerns **thresholding** *spurious* edge(s) in FCs. State-of-the-art thresholding methods are largely **arbitrary** and **non-analytic**.

Significance

- Leveraging recent theoretical developments in Stochastic block model (SBM), we provide a framework to investigate the prominence of Resting State Networks (RSNs);
- This analytic framework assesses the recovery of community labels across different thresholds in two different ways:
  - weak recovery measured by signal-to-noise ratio
  - exact recovery measured by Chernoff-Hellinger divergence
- This work paves the way to an *automated thresholding* of FCs based on prior knowledge of community structures.

Background Information

A) Stochastic Block Model (SBM) [1]

Notations

- $\boldsymbol{G}$: network/graph; $V(\boldsymbol{G})$: vertex set of $\boldsymbol{G}$;
- n: number of nodes;
- k: number of communities;
- p_i : prob. of a node in community $i \in [k]$;
- $\left[\frac{q_{ij}}{n}\right]$: prob. of edge between community i and j;
- X_u: community label of node $u \in V(\boldsymbol{G})$;
- s_n: scalable factor of degree regime in $G_{n\to\infty}$;
- **Community profile matrix:** *denoted as*

$$\mathbf{M} = s_n diag(p)Q = s_n PQ$$

where i^{th} column is the expected number of edges that community i has with other communities.

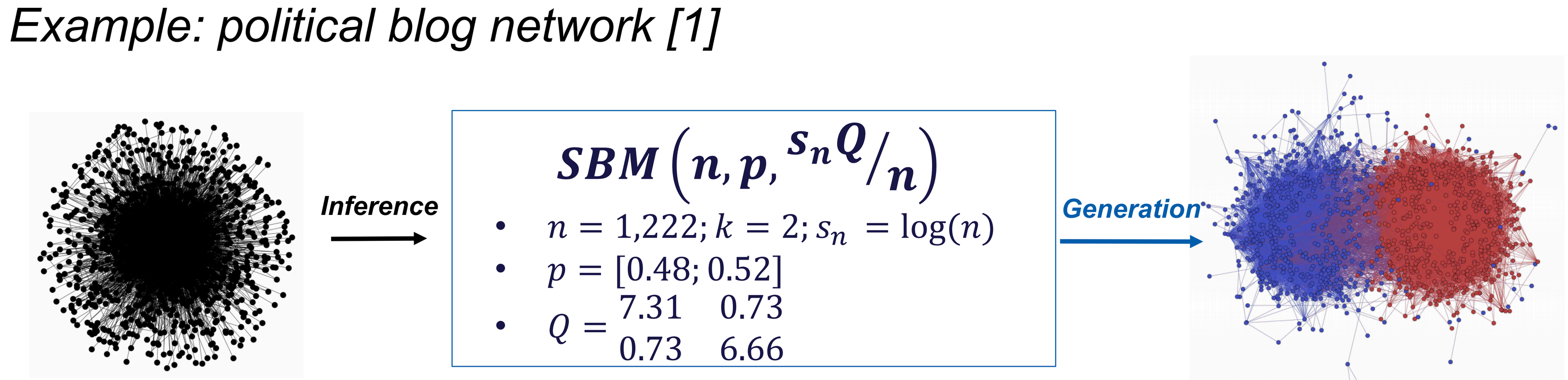
Ensemble Generation

$$\boldsymbol{SBM}\left(\boldsymbol{n}, \boldsymbol{p}, {}^{\boldsymbol{s_n Q}}/_{\boldsymbol{n}}\right) \longrightarrow \begin{matrix} X_u \\ \boldsymbol{G} \end{matrix}$$

Inference

$$\boldsymbol{G} \longrightarrow \boldsymbol{SBM}\left(\boldsymbol{n}, \boldsymbol{p}, {}^{\boldsymbol{s_n Q}}/_{\boldsymbol{n}}\right)$$

Example: political blog network [1]



Real Network — SBM parameters — Clustered Network

B) Network Neuroscience

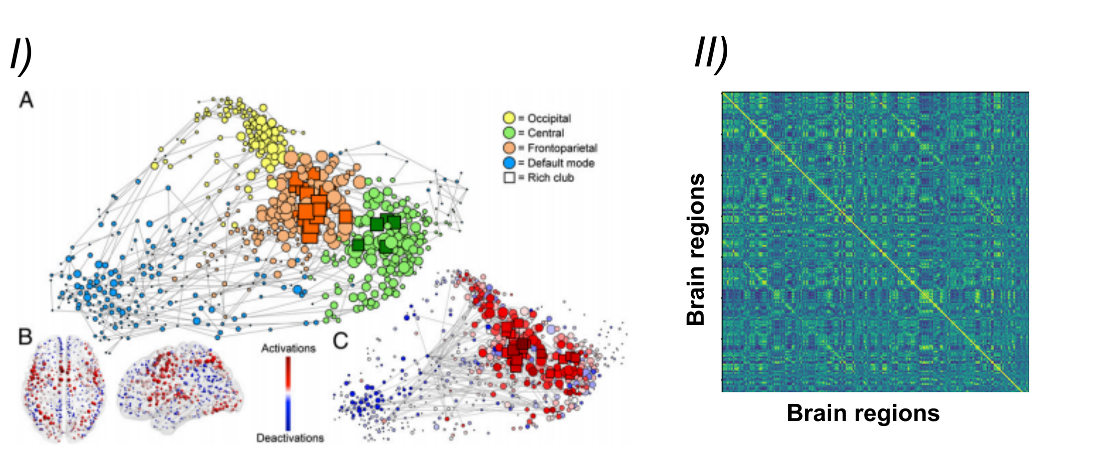


Figure 1: I-A),B),C) are different topological presentation (functional) brain networks[2]. II) Correlation matrix of two regions of interest (ROIs) in the human (functional) connectome.

Stochastic Block Model Framework

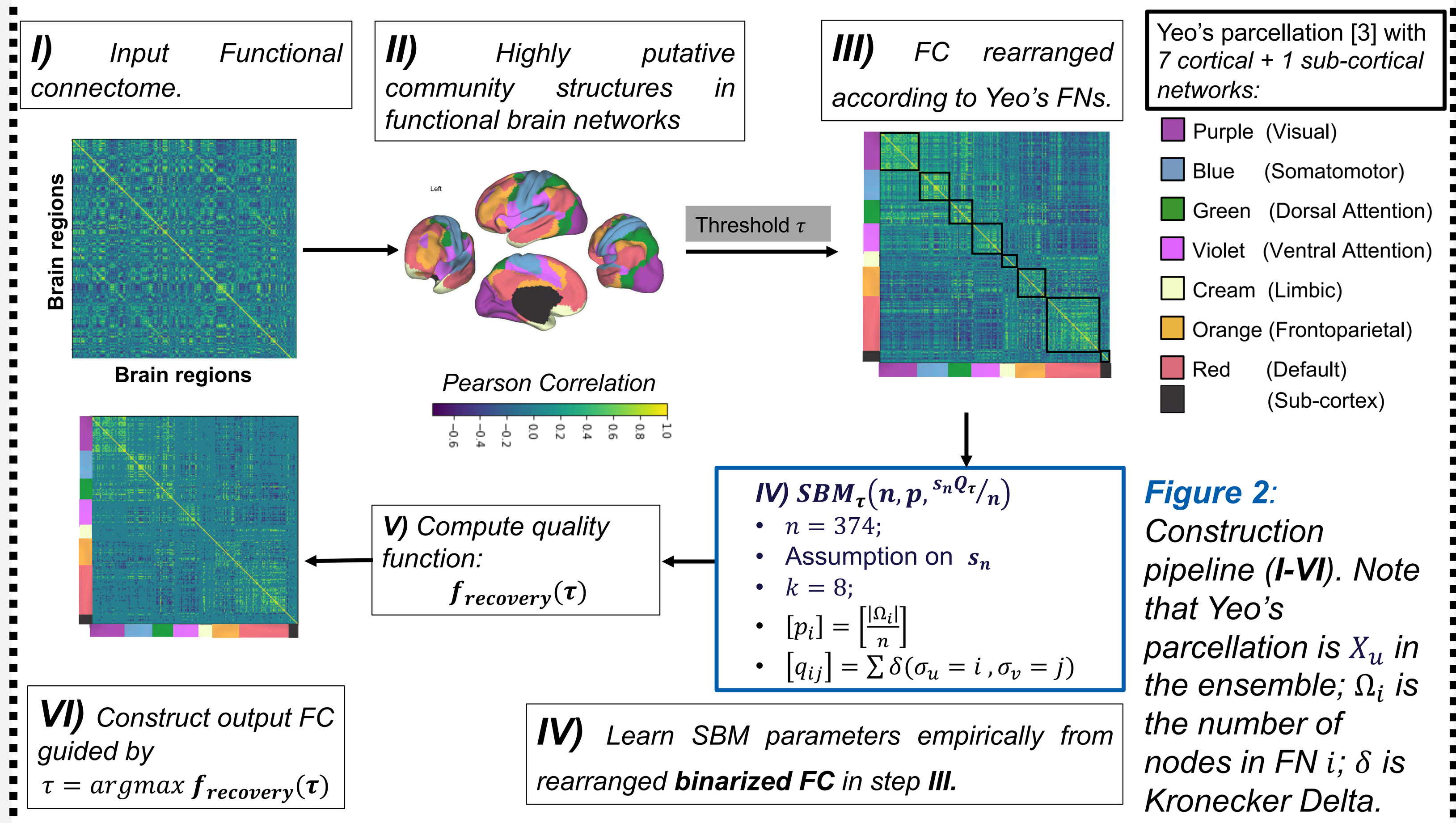


Figure 2: Construction pipeline (I-VI). Note that Yeo's parcellation is X_u in the ensemble; Ω_i is the number of nodes in FN i; δ is Kronecker Delta.

Data Description

100 unrelated subjects in Human Connectome Project with resting state and 7 tasks: Gambling, Emotion, Language, Motor, Relations, Social, and Working memory (WM).

Weak Recovery

Weak recovery: identify node's community label correctly at rates better than chance. Asymptotically, the probability of correctly labelling $1 - O_n(1)$ vertices in $\boldsymbol{G}$ is $\max_{i\in[k]} p_i + \epsilon$.

Theorem (AS 15, 17) – *High signal-to-noise ratio (SNR) implies higher chance to recover community labels.*

Let $(X, G) \sim \boldsymbol{SBM}\left(\boldsymbol{n}, \boldsymbol{p}, \frac{\boldsymbol{Q}}{\boldsymbol{n}}\right)$ for p, Q arbitrary. If $SNR > 1$, then weak recovery is efficiently solvable; where

$$f_{weak}(\tau) = SNR(\tau) = \frac{\lambda_2^2}{\lambda_1}$$

And λ_i is the i^{th} eigen value of the community profile matrix $\boldsymbol{M}$. The network is in **constant degree regime**, i.e. $s_n = 1$, asymptotically.

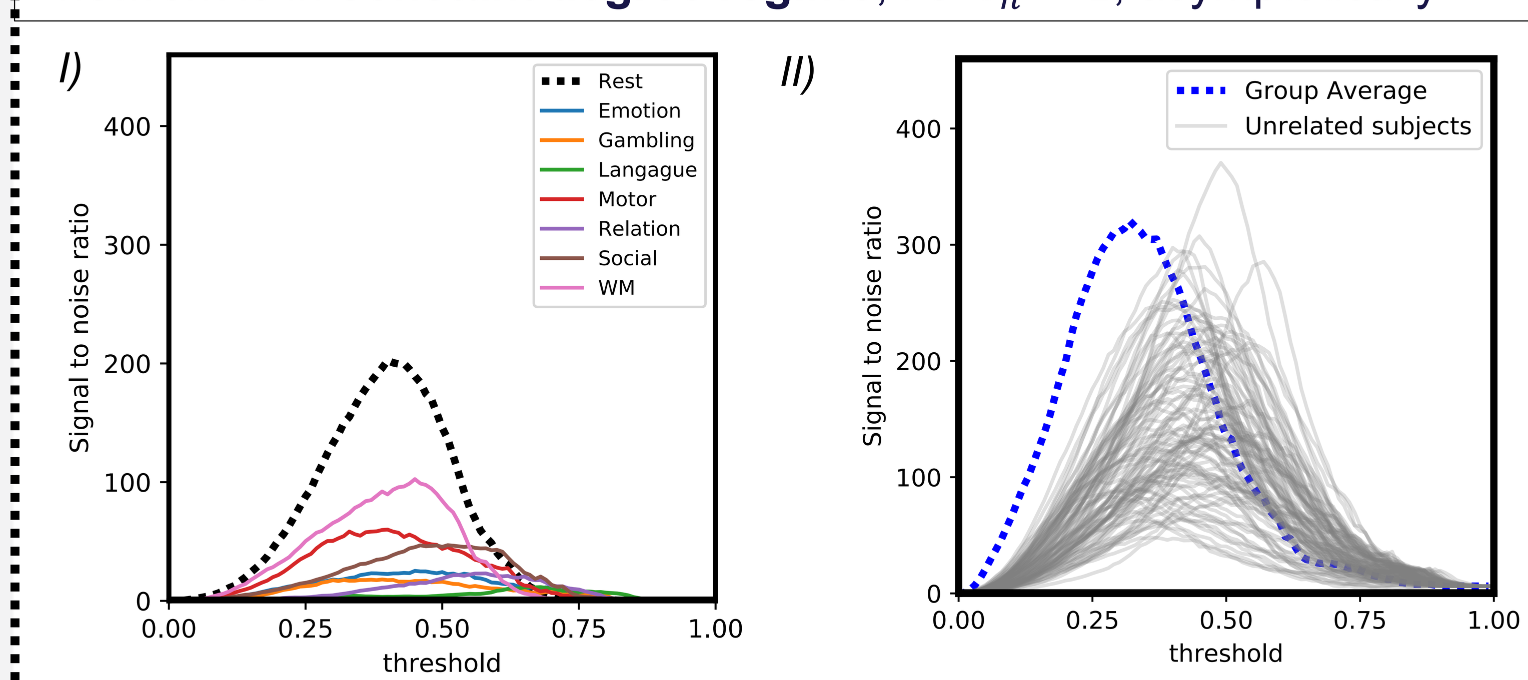


Figure 3: I) SNR for one subject at rest and engaging different tasks. As expected, RSNs are most prominent at rest; they are less prominent in all investigated tasks (for the majority of threshold domain). There are task(s) that are more RSN prominent than other(s). II) SNR for 10 subjects, at rest. Here, there exist subject(s) that has rest SNR that is/are smaller than task SNR of subject on the left plot.

Exact Recovery

Exact recovery: identify node's community label with high probability. Asymptotically, the probability of correctly labelling $1 - O_n(1)$ vertices in $\boldsymbol{G}$ is $1 - o(1)$

Theorem (AS 15) – If all community profiles are distinguishable, we can perform exact recovery.

Exact recovery in $\boldsymbol{SBM}\left(\boldsymbol{n}, \boldsymbol{p}, \log(\boldsymbol{n})\frac{\boldsymbol{Q}}{\boldsymbol{n}}\right)$ is solvable and efficiently so if

$$f_{exact}(\tau) = I_+(p, Q) = \min_{i,j\in[k]} D_+\left((PQ)_i \| (PQ)_j\right) > 1$$

Where

$$D_+\left((PQ)_i \| (PQ)_j\right) = \max_{t\in[0,1]} \sum_x (PQ)_j(x) f_t\left({}^{(PQ)_i(x)}/_{(PQ)_j(x)}\right)$$

The network is in **diverging degree regime**, i.e. $s_n = \log(n)$, assymptotically.

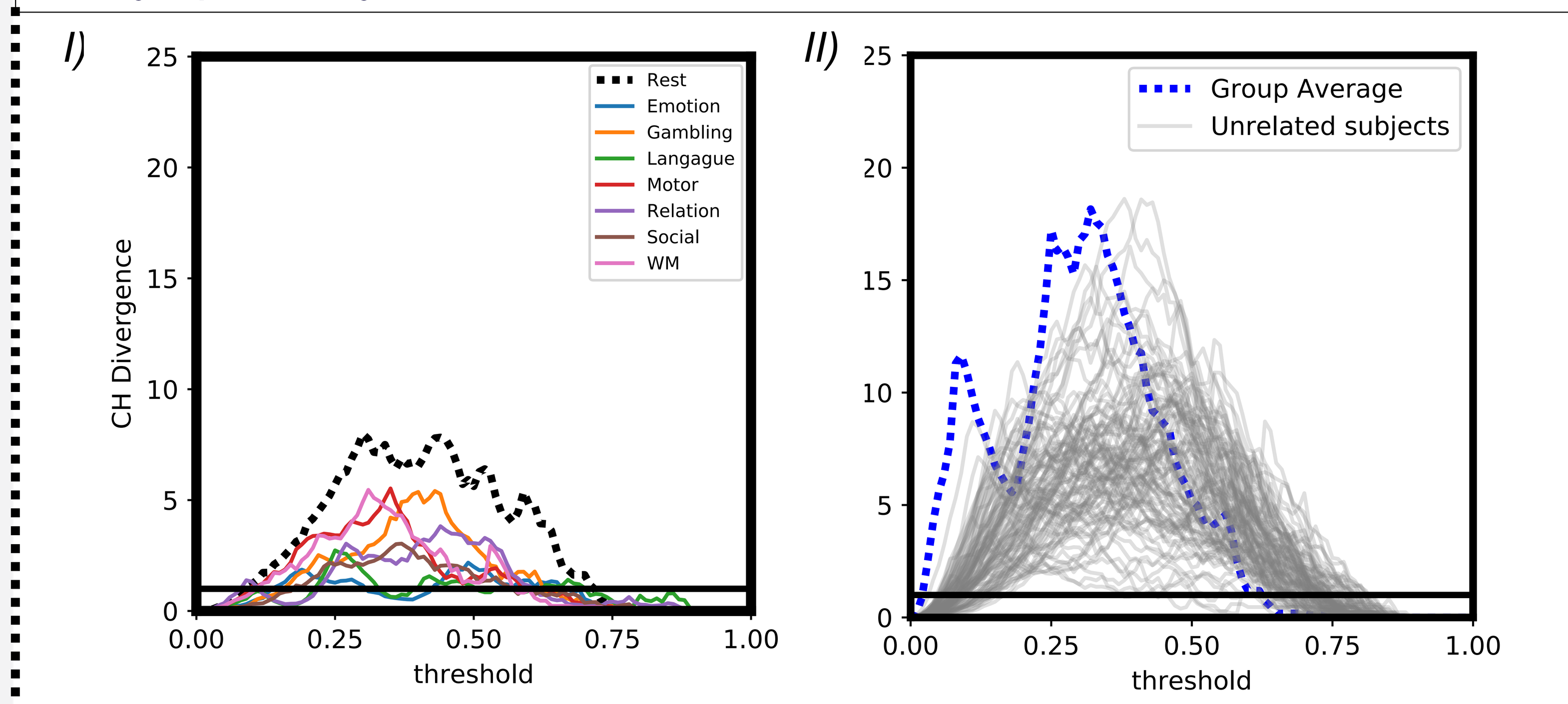


Figure 4: I) Chernoff-Hellinger Divergence (CHD) for one subject at rest and different tasks. CHDs are most prominent at rest; they are less prominent in all investigated tasks (for the majority of threshold domain). There are task(s) that are more RSN prominent than other(s). II) CHDs for 10 subjects, at rest.

Conclusion & Future work

- Allows for a data-driven method for automated thresholding of functional connectomes with prominent functional networks;
- SNR can be interpreted as a quantitative measure of whole-brain FC, since SNR of group average is higher than most subjects;
- Allows the comparisons of FNs' prominence among different tasks (within one subject) or among different subjects;
- ❑ Apply to other upstream processing steps, i.e. at the time series level;
- ❑ Extend the analysis to 447 unrelated subjects;
- ❑ Explore different parcellations other than Yeo [3] and Glasser.

Acknowledgements

We would like to thank Phan Nguyen, Enrico Amico, and Kausar Abbas for providing valuable comments.

References

[1] Abbe, E. (2017). Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research*, *18*(1), 6446-6531.
[2] Crossley, N. A., Mechelli, A., Vértes, P. E., Winton-Brown, T. T., Patel, A. X., Ginestet, C. E., ... & Bullmore, E. T. (2013). Cognitive relevance of the community structure of the human brain functional coactivation network. *Proceedings of the National Academy of Sciences*, *110*(28), 11583-11588.
[3] Thomas Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., ... & Fischl, B. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology*, *106*(3), 1125-1165.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.
IM Release number: 782257