Understanding Rich Club with a Stochastic Block Model

Soroosh Afyouni[1], Dragana Pavlovic[2], Emma Towlson[3], Petra Vertes[4], Theo Arvanitis[1], Edward Bullmore[4,5], and Thomas Nichols[1]

[1]University of Warwick [2]National University of Singapore [3]Northeastern University [4]University of Cambridge [5]GlaxoSmithKline

1. Introduction

Rich Club (RC) analysis is one way to gain insight to the core structure of the human connectome. Impediments to general use of the RC are: **1)** difficulty with summarisation of multi-subject data, **2)** computational inefficiency. In this work we propose a new measure called **Rich Block** which uses the **Generalised Linear Stochastic Block Models (GL-SBM)** to: It fits a statistical model to a network and, then re-define the model based on the GL-SBM parameters, aiming to:

- Identify the rich cores of a network
- Provide a coherent approach to multi-subject inference of Rich Club characteristics multi-subject analysis of Rich Club.
- Forming 'rich clubs' (block) that not only share the same degree but also share the same pattern of connectivity.

2. Methods

Erdos Renyi Mixture Model (ERMM):
The ERMM decomposes a $N \times N$ binary network $G(N, E)$ into Q latent blocks. Prevalence nodes in block $\ell \in \{1, ..., Q\}$ is α_ℓ, such that $\sum_\ell \alpha_\ell = 1$. Edges occur between block q and ℓ as a Bernoulli random variable with parameter $\pi_{q\ell}$. where $\Pi = ((\pi_{q\ell}))_{1 \leq q, \ell \leq Q}$. The degree distribution is then approximately a Poisson mixture with parameters $\lambda = (\lambda_q)_q$. Mixing weights α and λ gives the expected degree of each block: $\lambda_q = (N - 1) \sum_{\ell=1}^{Q} \alpha_\ell \pi_{q\ell}$
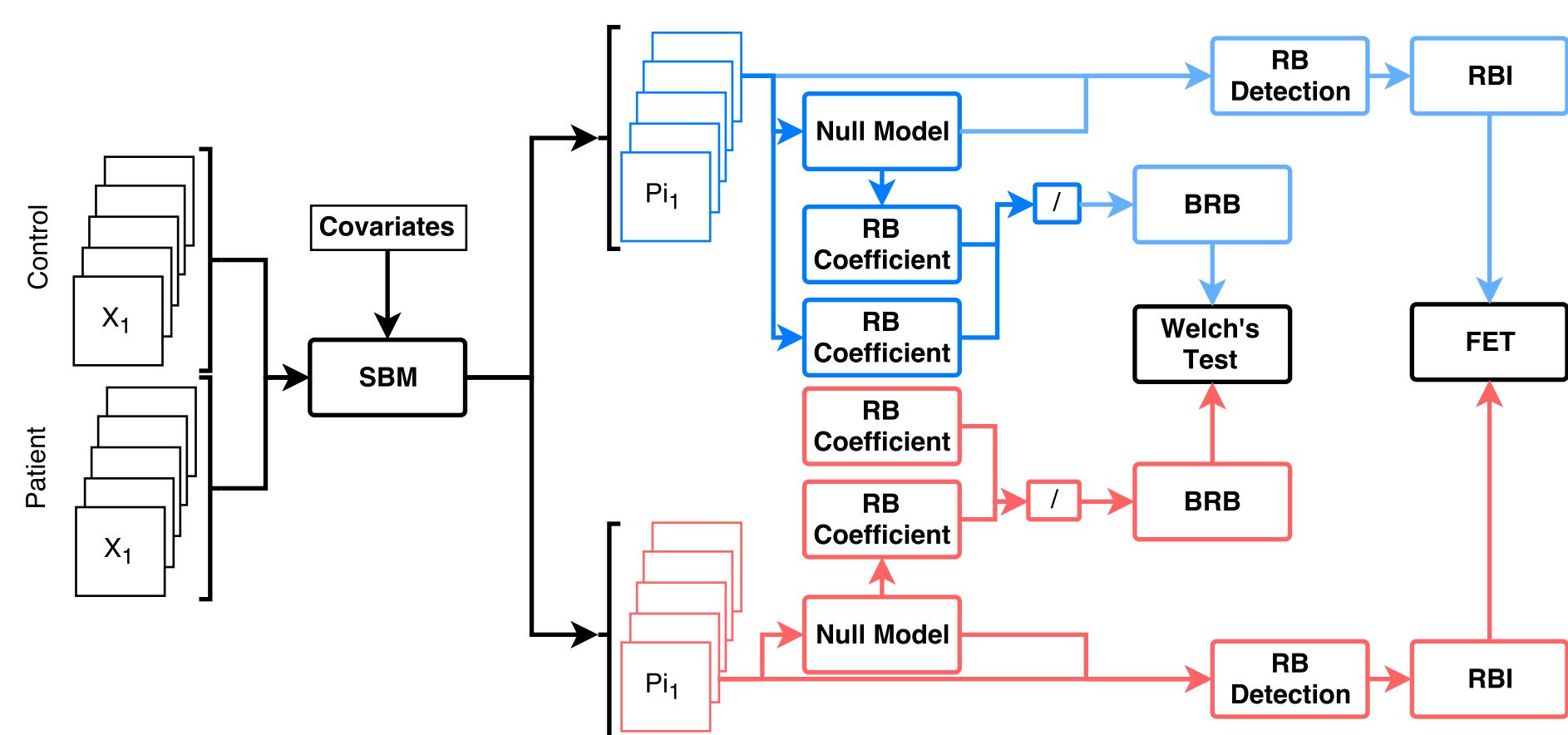
Generalised Linear Stochastic Block Model (GL-SBM):
Fitting the ERMM to multiple subjects will produce different block structures in each subject. In (Pavlovic et al 2015) we introduce a new extension to ERMM that finds a **common** set of Q blocks, allowing subject connectivity to differ as per **covariates** in a logistic regression.

Rich Block (RB):
We propose defining a new version of Rich Club in terms of the ERMM, describing how a set of Erdos-Renyi blocks with expected degree of k or larger interact with each other. Fundamental to RB is how the SBM groups together nodes with similar profile of connectivity. Hence, all nodes in a given block have the same fitted degree, and thus we treat these **'clubs'** as the basic units of the RB. The SBM implies the statistical distribution of RB degree which is a **Binomial mixture** of blocks of certain degree.

Testing of Rich Blocks is conducted by Monte-Carlo simulation of synthetic degrees from the bionomial mixtures of each RB and testing against its corresponding Null-RB with null model specified below. The test would be possible by assigning RB coefficients to corresponding block (BRB) and binary
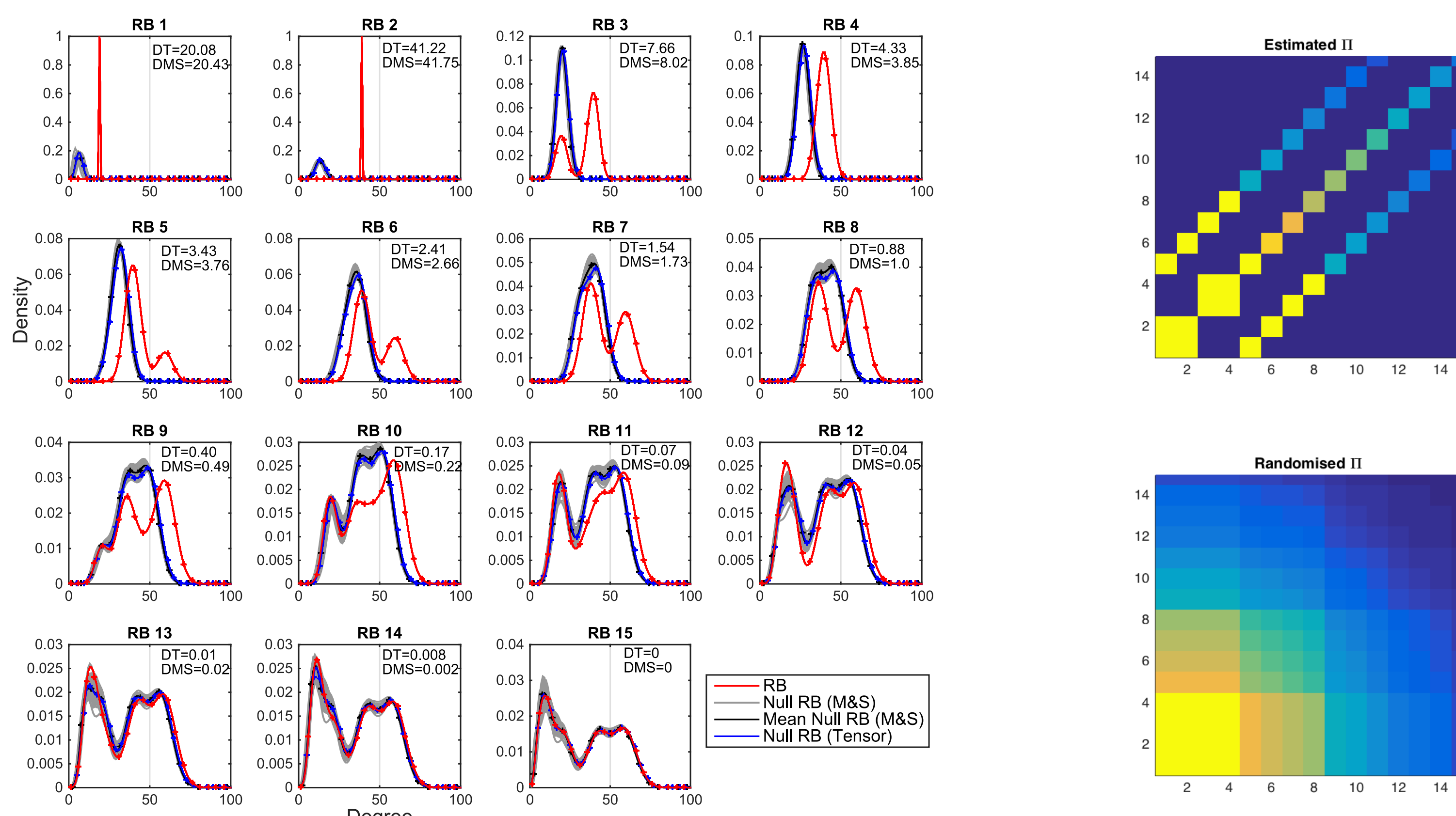

vector where blocks in the RB organisation set to one, otherwise to zero (RBI). The test between $RBI_{control}$ and $RB_{patient}$ is done by Fisher Exact Test (FET).

3. Null Models and Randomised RB

A relevant null model is essential for distinguishing between genuine and spurious RB behaviour. Regardless of empirical network, we merely use the model parameters to estimated the null model, called **Tensor Null Model (TNM)**. The RB null model is formed by calculating the expected connection probability of each block.

Below, we show how a simulated network (top-right) with synthesized RB effect is randomised using TNM (bottom-right). We also validate TNM against the traditional Maslov&Sneppen (M&S) method (left). In each of the Rich Blocks we calculate the **Kullback-Leibler divergence** between the RB's pdf and pdf of TNM and M&S randomised RB. The results suggest the TNM performs almost identical to mean of thousand random networks resulted by M&S
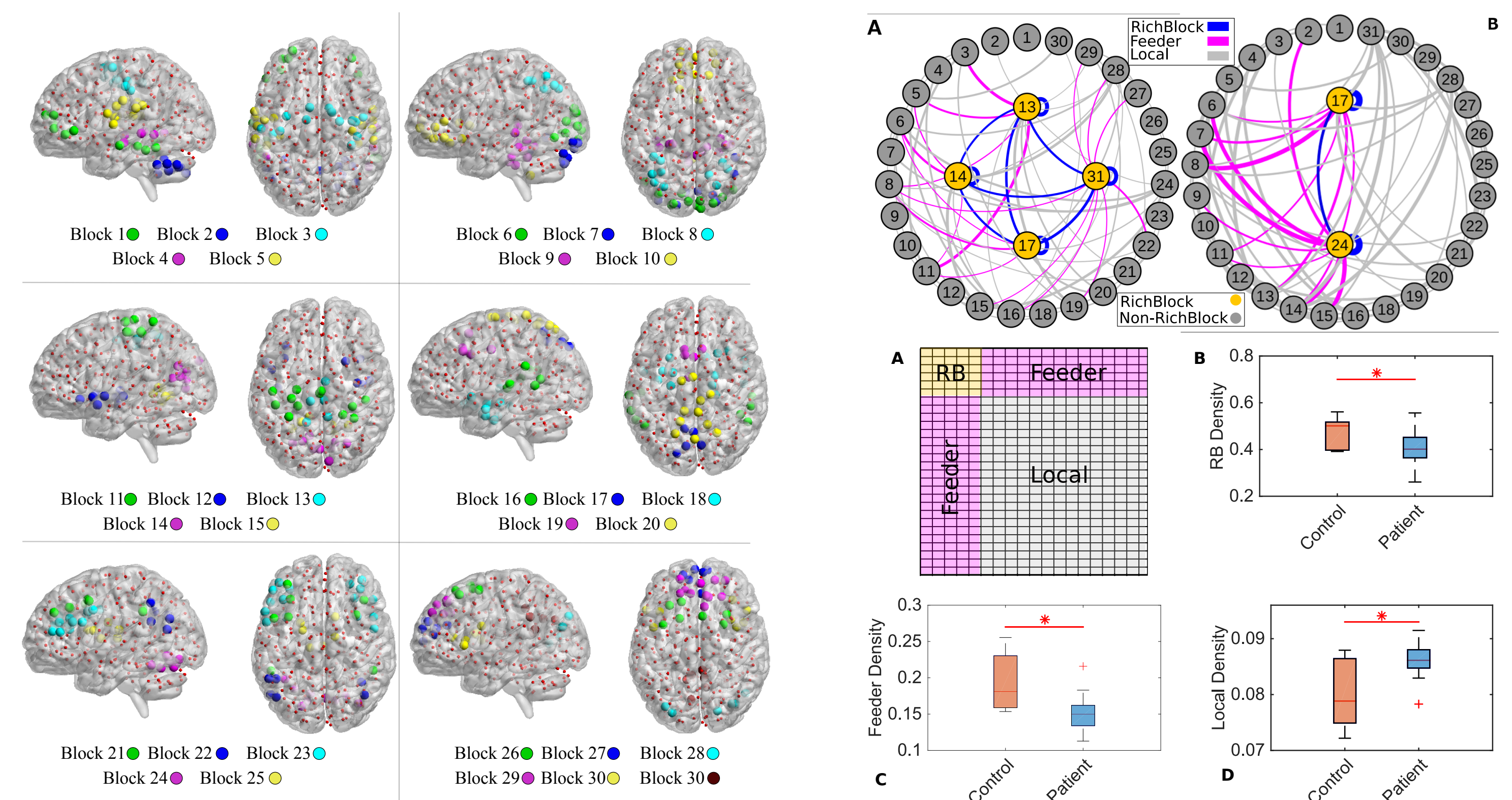

References

Pavlovic et al, Journal of Computational Statistics and Data Analysis, 2015 [under review] — Maslov and Senpenn, Science,2011 — Lynall et al, Journal of Neuroscience, 2010 — Olhede et al, PNAS, 2014 — Glasser et al, Neuroimage , 2015 — Yeo et al, Journal of Neurophysiology, 2011

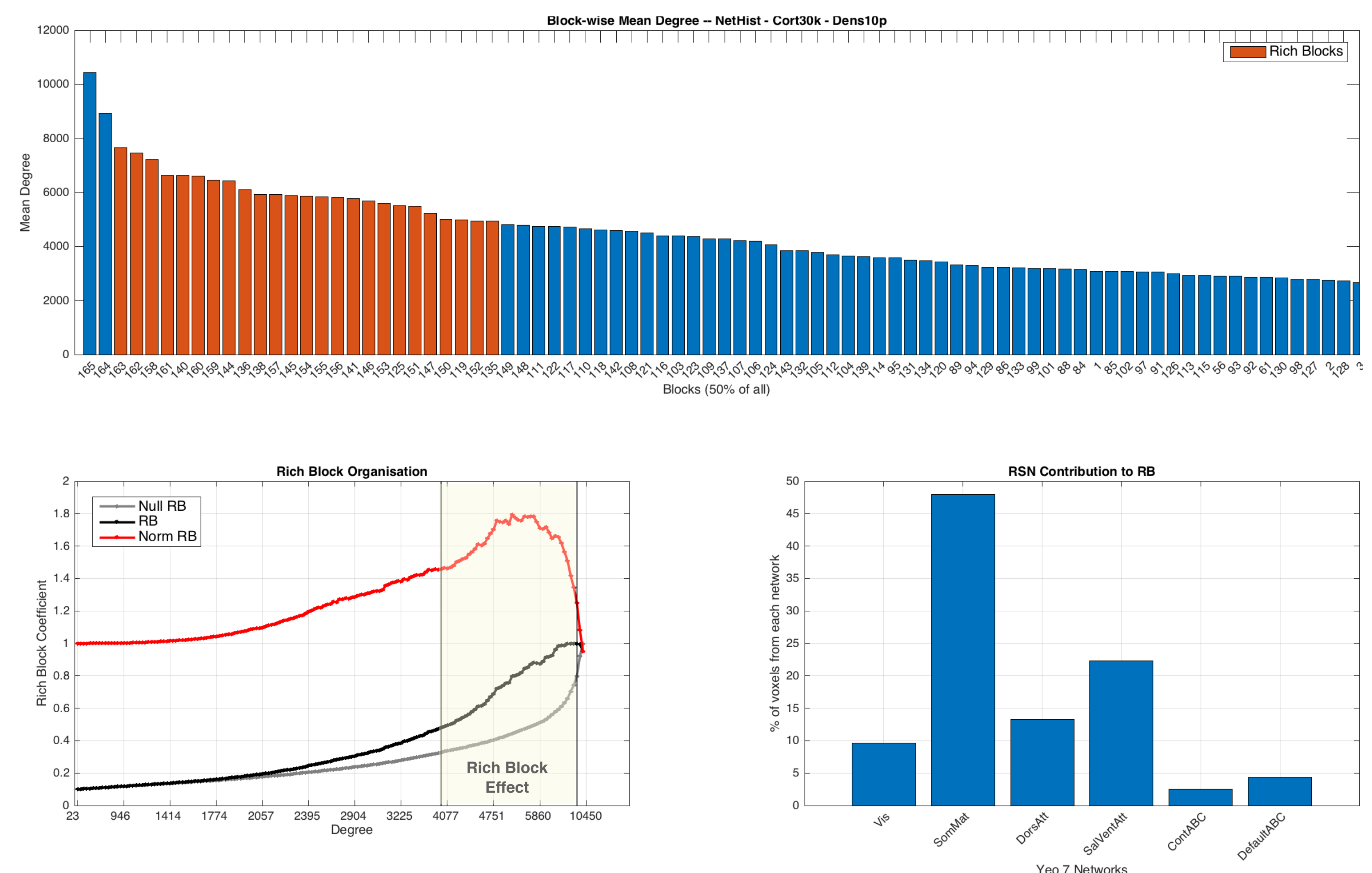## 4. RB application on Connectome of Schizophrenia

Wavelet correlation transformation (scale 2) was applied on 13 healthy control and 12 schizophrenic to form a similarity matrix among **297 regions** of interest obtained from a AFNI atlas. The similarity matrix of each subjects was binarised at 10% density. GL-BMM was used to find a **common block assignment** among all 25 subjects involved in the study. The effect of covariates between subjects such as age, IQ and fMRI head motion were corrected in the model. GL-BMM found **31 blocks** ranging in size from 5 to 16 as well as the probability of forming an edge between each pair of blocks. Rich Block was conducted separately for each group.

The estimated blocks appeared to form interpretable neuro-antomical patterns (above left) despite having no prior knowledge of brain structure. The RB organisation (yellow circles) and their interaction with rest of the network (top-right) suggest a decline in Default Mode Network participation in RB organisation while the cerebellum becomes one of the two blocks identified as RB in patients. Bearing in mind that each subjects has the same density (10%), the investigation of feeder and local connection (bottom-right) suggest that schizophrenic patients have a lower rate of connection in their RB organisation.

5. RB application on large-scale networks

Finding Rich Club effect on large-scale (> 10k) data-sets requires a iterative randomisation of the empirical network which is followed by calculation of the Rich Club coefficients on each of these iterations. On a 8-cored 2.5GHz CPU this process can take up to several weeks. However, by employing an appropriate SBM, the process of finding Rich Block organisation only takes **few hours**. Here we show how Rich Block effect is found within the 0-fold decimated cortical area of dense-connectome ($30k \times 30k$ adjacency matrix). We use the **Network Histogram (NH)**, an SBM method designed for large-scale networks. That, like ERMM, groups nodes with similar profile of connectivity, however, with the constraint that the blocks have equal sizes. We estimate the probability matrix of inter and intra block connection according to the block assignments obtained by NH and then use method described to detect the RB phenomena.

The NH resulted in 164 blocks of 181 nodes and one block of 21 nodes. Blocks with expected degree ranged between $\approx 4k$ to $\approx 10k$ were identified as Rich Block (bottom-left). These block contain the total number of 4343 voxels. The voxels in RB organisation were overlapped with **Yeo's 7-Network** atlas (bottom-right) , which shows that rich blocks were associated with majority of functional network specially Somatosensory and Saliency Ventral Attention network.

Conclusions

In this study we have proposed a new metric called **Rich Block** which is combination of stochastic block model and empirical rich club. This approach simultaneously identifies a simplified or **'compressed'** network structure while identifying the **'richly connected'** components of this structure, and facilitating group comparisons. We also demonstrated how this can be used to detect 'rich' core of the immense dense-connectome networks, where the traditional Rich Club would be computationally infeasible.