Rich Block: Rich Club Analysis with the Erdos-Renyi Mixture Model

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

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

1. Introduction

Rich Club (RC) analysis is one way to gain insight to the structure of the human connectome. One impediment to general use of the RC is difficulty with multi-subject data. In other work we have shown that stochastic block models like the Erdos-Renyi Mixture Model (ERMM) are useful for capturing of biological relevant group structure, in single networks (Pavlovic et al 2014), and in recent work we have extended this to find common network structure in multiple subjects (General Linearised Stochastic Block Model), GL-SBM, (Pavlovic etal, 2015 - See poster OHBM2015 3861 WT-PM). Here we use the GL-SBM to develop a new metric called Rich Block (RB), a stochastic block model implementation of RC, with two main goals:

- Provide an approach to multi-subject analysis of Rich Club.
- Forming clubs (block) that not only share the same degree but also share the same pattern of connectivity.

2. Methods

Rich Club (RC):
Rich Club measures how densly nodes of degree k are connected to other nodes with degree k or larger. Normalised Rich Club coefficient, Φ_k measures the density of club k divided by the average density of a null network's club k. Null samples also permit computation of P-values and selection of significant RCs, and are obtained by random but degree-preserving rewiring (Maslove & Sneppen, 2002).

Erdos Renyi Mixture Model (ERMM):
The ERMM decomposes a $N \times N$ binary network $G(N, E)$ into Q latent blocks. Prevalence of 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 $\Pi = ((\pi_{ql}))_{1 \leq q, \ell \leq Q}$. The degree distribution is then approximately Poisson mixture with parameter λ. Mixing weights α and λ gives the expected degree of each block: $\lambda_q = (n-1) \sum_{l=1}^{Q} \alpha_l \pi_{ql}$

Generalised Linear Stochastic Block Model (GL-SBM):
Fitting the ERMM to multi subjects will produce different block structures in each subject. In [Poster 3861 WT-PM] we introduce the GL-SBM 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 ERMM groups nodes with common a pattern 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 RC. The ERMM implies the statistical distribution of RB degree (again, a Binomial mixture); with the GL-SBM, Π and α can be found for group data, not just single subjects.
normalisation and inference of RB coefficients is done with a null model. In contrast to RC, we can estimate null Π while holding fixed the node-to-block assignments fixed. This allows computationally efficient sampling from the null, as the Binomial mixture can be directly sampled from the fit to one random sample.

3. Simulation Methods

Simulated networks were formed from synthetic probability matrices, Πs (Fig1B). Regardless of network size, 25% of each probability matrix were assigned higher probabilites with their intra-connections also inflated to ensure that Rich Club/Block will be formed. Twelve different types of networks were used to conduct the simulations: Networks of size 60, 120 and 300 were selected. For each category, four Q numbers (6,15,30,60) were investigated. All blocks assumed to have the same size. Three categories were discarded due to low ($\geq$ 4) ratio of node to block
Example for N=300, Q=15
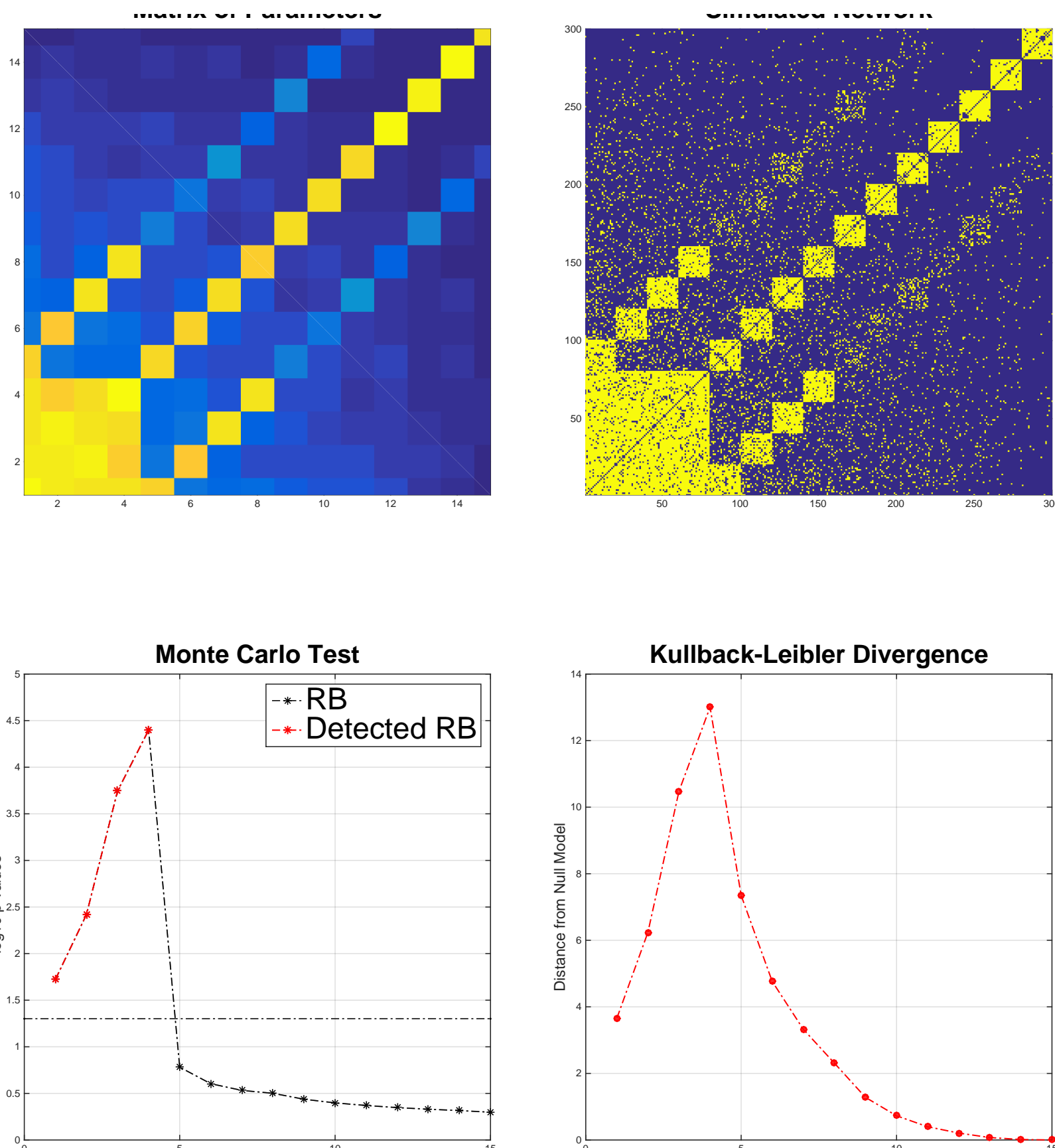
4. Evaluation and Validation

We use two approaches to compare RB with traditional RC:

- Node-wise RB/RC coefficient (**Nodal RB/RC**). The nodal coefficient for node i is the Φ_k of the smallest RC or RB containing i. We compare NodalRB to NodalRC with R^2.
- Node-wise RB/RC labelling (**Identity**). Node i is given a value of 1 if it belongs to the largest RB/RC found significant at level 5%, 0 otherwise. We compare RB Identity to RC Identity with Adjusted Rand Index (ARI).

Results suggest that although distance between Nodal RB and Nodal RC remain respectively the same over different number of nodes and blocks, RB/RC identity values become closer to each other as number of blocks increase.
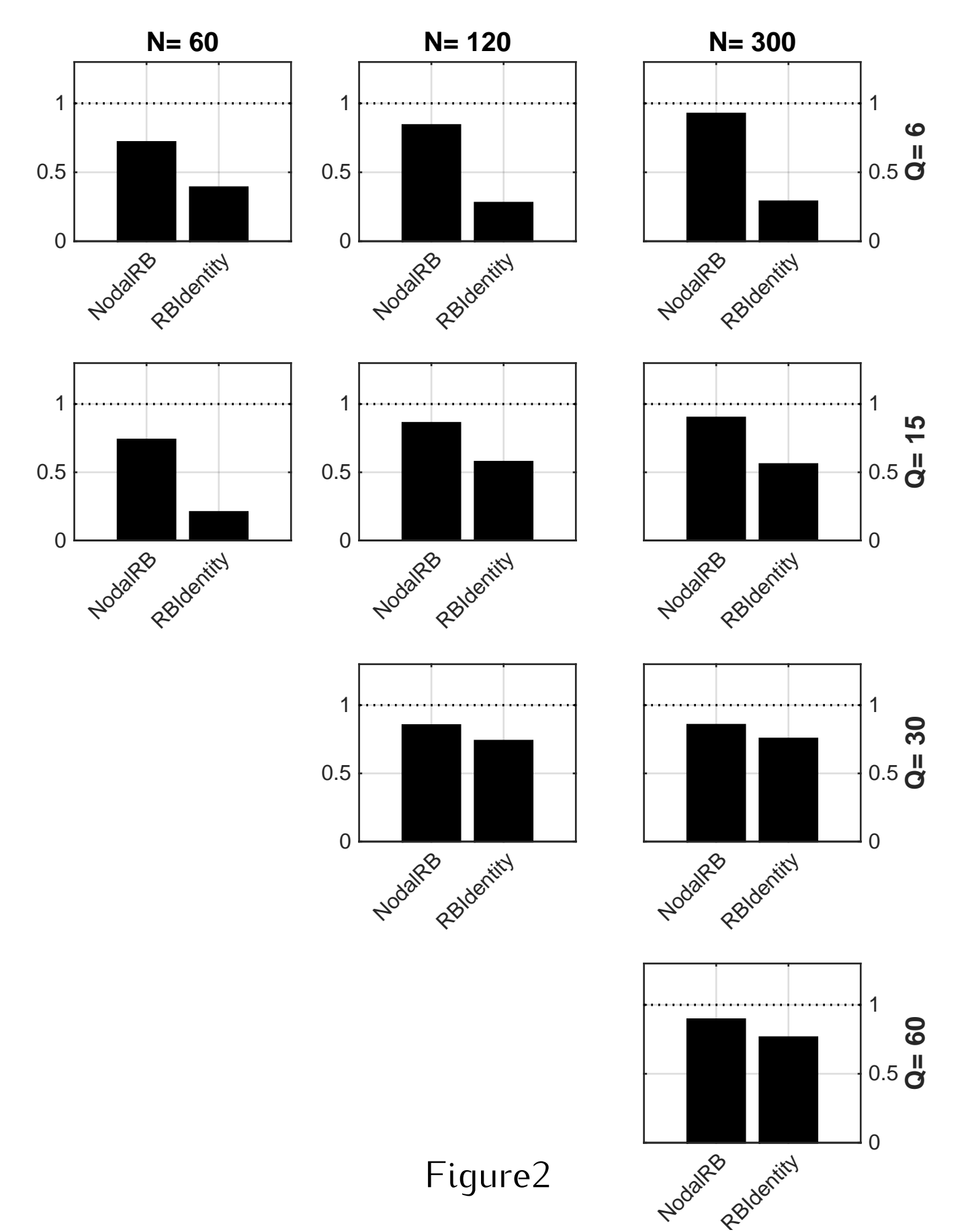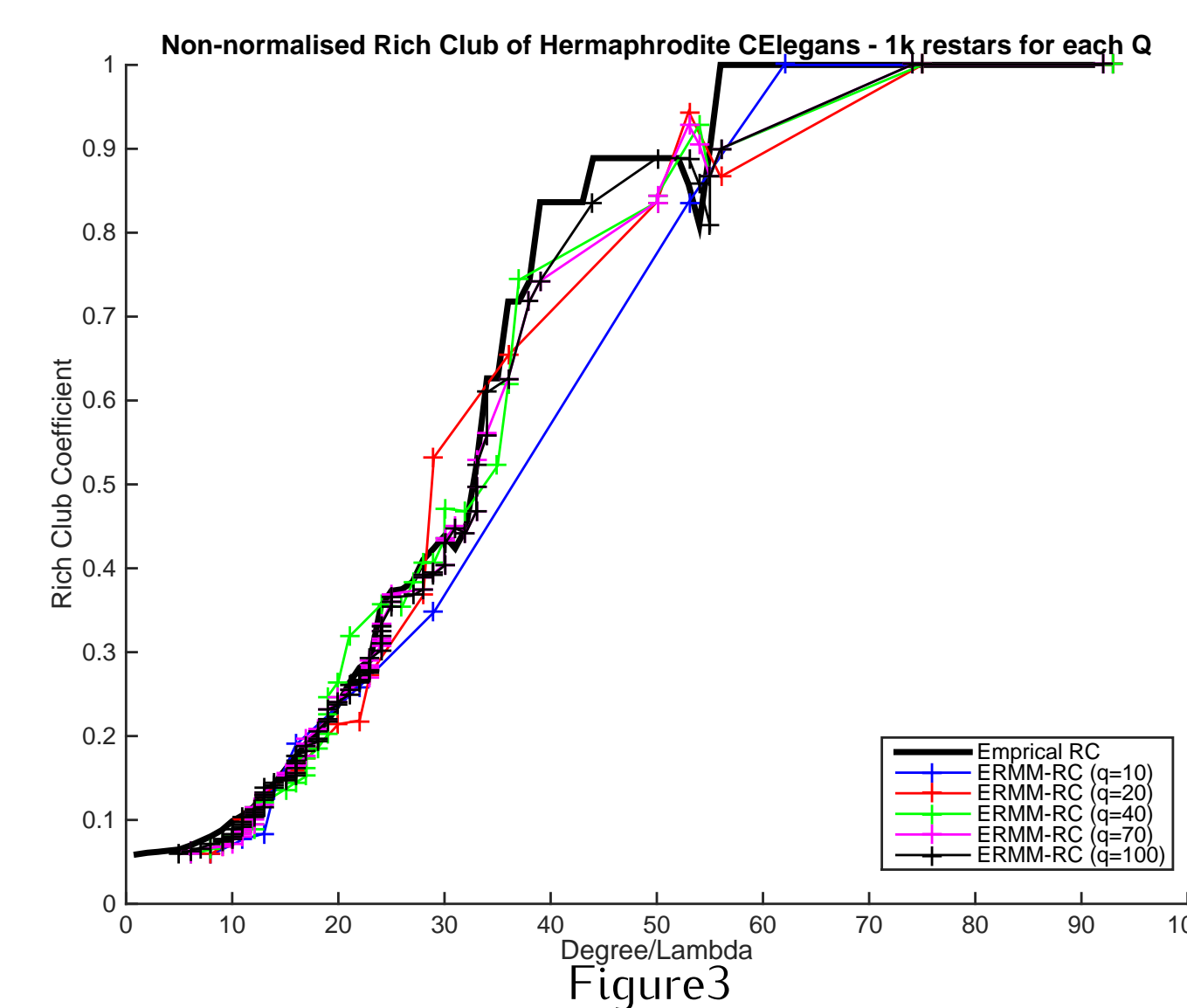

Figure2

Figure3

Fig 3 shows an example with the C.Elegans connectome (N=279 nodes, E=2287 edges, [Towlson et al, 2013]), with RC's Φ with thick black line, and RB's Φ for various Q's. There is close agreement that improves with Q.

5. Application: Resting State fMRI Connectivity in Schizophrenia

Rich Block was used to detect the differenceresting fMRI connectivity between controls (n=13) and schizophrenic subjects (n=12) (Lynall etal 2010). Wavelet correlation transformation (WCT) (scale 2) was used to form a network, after binarisation to give 10% edge density, GL-SBM with patient/control, age and movement covariates found Q=31 latent blocks.
Fig4 Illustrates degree distribution of estimated models for control and patient versus degree distribution of their null models, sorted by descending λ. Despite detectable difference between degree distribution of estimated models and null models in earlier RBs, they significantly overlaps in further RB until they become identical in the last RB (degree distribution of last RB (RB31) is degree distribution of whole network. For sake of visualisation, we just showed 16 RBs below.)
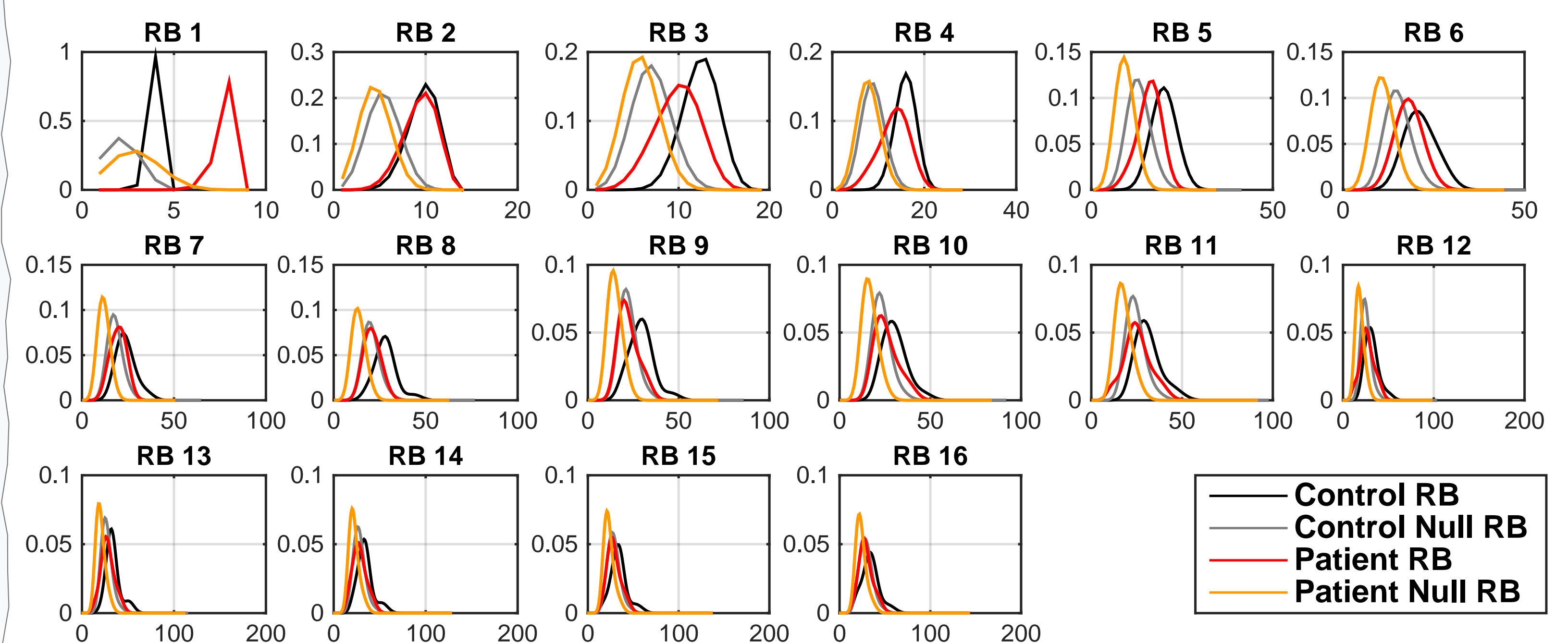

Figure 4

Fig5 shows detected RBs in red on unsorted expected degree bar plots for patients and controls.
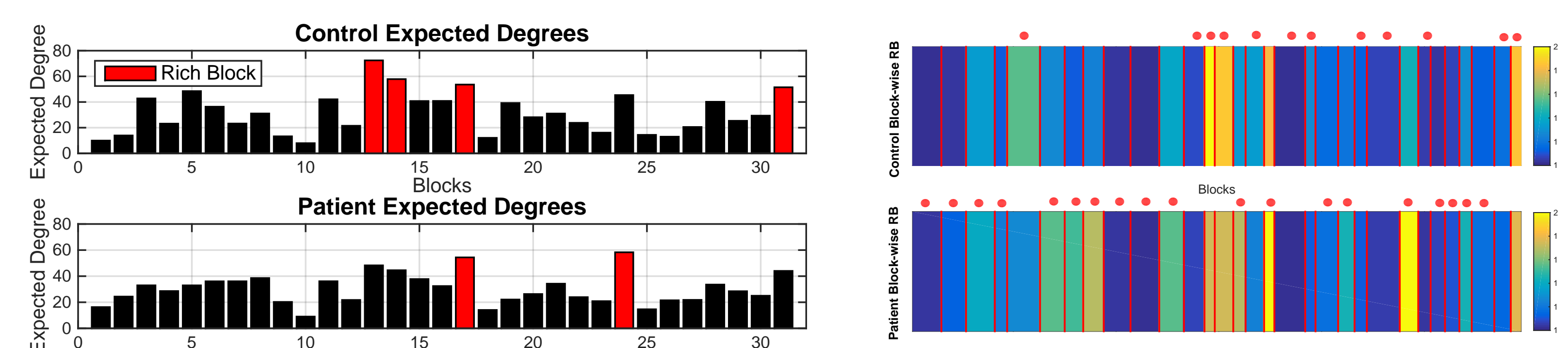

Figure 5

Figure 6

Fig6 shows block-wise RB coefficients, for patients and controls. A Monte Carlo (parametric Boostrap) test was used to compare Φ between populations. The blocks which are significantly larger was assigned by a red circle at the top of it.

References

Heuvel and Sporns. Journal of Neuroscience, 2011 — Lynall et al, Journal of Neuroscience, 2010 — Maslov and Sneppen, Science, 2002 — Pavlovic et al, PloS One, 2014 — Towlson et al, Journal of Neuroscience, 2013

Conclusions

In this study we have proposed a new metric called Rich Block which is combination of stochastic block model and empirical rich block. This approach simultaneously identifies a simplified or 'compressed' network structure while identifying the 'richly connected' components of this structure, and facilitating group comparisons.