Reproducibility Assessment of the First Principal Network Calculation: a Tool for Studying Anatomical Brain Connectivity

E. Biondetti[1], C. A. M. Wheeler-Kingshott[2]

[1]UCL Department of Medical Physics and Biomedical Engineering, London, WC1E 6BT, United Kingdom.
[2]NMR Research Unit, Queen Square MS Centre, UCL Institute of Neurology, London, WC1N 3BG, United Kingdom.

TARGET AUDIENCE: Scientists and physicians interested in methods to study brain connectivity through identification of relevant networks of cortico-cortical connections.

PURPOSES: 1) To evaluate the reproducibility of the main subnetwork in the brain determined by the Principal Networks (PNs) method[1], 2) to cross-validate the result within our data set, 3) to study its dependence on the data set dimension and 4) to compare it with previously reported findings[1].

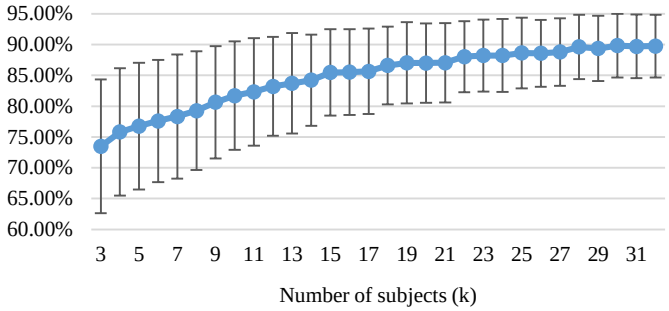
INTRODUCTION: Macroscopic brain connectivity refers to a pattern of structural, statistical or causal association between anatomically segregated brain regions. The pattern is usually represented through a graph, since graph theory allows the quantification of important network properties: segregation and integration. Segregation describes the tendency of a network to form clusters, i.e. groups of nodes in which internal connectivity prevails on the degree of connection with the outside. Integration indicates how easy it is to reach a node from any other node of the network. At a macro-scale the human brain is a "small-world network", i.e. a highly integrated network with a certain level of segregation, so not all the regions of the brain take part to every cognitive process and a meaningful characterization of neural connectivity is possible only identifying the relevant subnetworks in the brain. Recently, a method called "Principal Networks"[1] has been proposed. The authors described whole brain connectivity through between-subject correlation of cortical thickness measured in 64 areas. The resulting 64x64 matrix was decomposed with a principal component analysis (PCA)-based approach, giving a set of influential subnetworks associated to distinct eigenvalues of the matrix, in which each interconnection was weighted by the between-subject correlation of thickness of the corresponding vertices. Notably, the subnetwork associated to the main eigenvalue was the "first Principal Network" (first PN). In this work, the first PN is calculated from a new cohort of healthy subjects, to investigate of result reproducibility and its robustness in relationship with the number of subjects used.

METHODS: Subjects: 32 healthy subjects selected in a previous study[2]. Images acquisition: images were acquired on a 3-T Philips Achieva MRI scanner (Philips Medical Systems, Best, Netherlands) with a 32-channel head coil and 3-D sagittal T1-w FFE scan protocol (1-mm isotropic voxel size, 6.9 ms TR, 3.1-ms TE). All participants gave written informed consent. Cortical thickness measurement: white matter (WM) segmentations computed by FreeSurfer were visually inspected and corrected for errors in voxels inclusion/exclusion and intensity levels computation, and cortical reconstruction process was repeated when needed. Each subject's brain cortex was then measured in 64 areas segmented by FreeSurfer (labels[3]: 1002-3, 1005-34, 2002-3, 2005-34, renumbered[1] with indices 1-64). First PN: the network was calculated using the previously reported technique[1]. The resulting matrix was called **A**; vertices and edges of the corresponding graph were thresholded at 0.1 and 0.2, respectively. Leave-one-out cross-validation: the first PN was recalculated 32 times by excluding the t-th subject from the data set of the t-th iteration ($1 \le t \le 32$). Cross-validation outputs were indicated as $\mathbf{A}|_t = (a_{ij})|_t$ ($1 \le i,j \le 64$, $1 \le t \le 32$). Two matrices were determined: $\mathbf{P} = (p_{ij})$ and $\mathbf{D} = (d_{ij})$ ($1 \le i,j \le 64$). Given two cortical regions i and j, p_{ij} described the frequency of appearance of their interconnection, while d_{ij} evaluated the difference in calculating the interconnection weight from the full data set or by using the leave-one-out approach, based on the absolute difference between the elements of **A** and the corresponding mean elements resulting from the leave-one-out iterations. Information from **P** and **D** was combined to assess reliability of **A**. Variability related to data set dimension: for each number k of subjects ($1 \le k \le 32$), for each of 1000 replicates ($1 \le n \le 1000$) a new data set of cortical thickness measurements was obtained by randomly sampling k subjects from the original 32 (with replacement), then the first PN $\mathbf{A}_{k,n}$ was recalculated. The overlap between regions in **A** and regions in $\mathbf{A}_{k,n}$ was determined as $\boldsymbol{x}_k(n)$ = dim(regions in **A**'s graph ∩ regions in $\mathbf{A}_{k,n}$'s graph) * 100 / dim(regions in **A**'s graph). For each value of k information was summarized through mean and standard deviation of $\boldsymbol{x}_k(n)$.

Table 1 Leave-one-out cross-validation results.

	Regions **58-44**	Regions **22-20**	Other regions (not in **A**)
Values of **P**	$p_{ij} \ge 0.9$	$0.5 < p_{ij} < 0.9$	$p_{ij} < 0.2$
Values of **D**	$d_{ij} < 0.025$	$0.05 < d_{ij} < 0.25$	$d_{ij} \le 0.05$

Figure 2 Agreement with **A** related to the number of subjects.



RESULTS: Calculation of the first PN was 1) reproducible, 2) accurate within our data set, 3) strongly dependent on the number of subjects used for its determination. Measured thickness values were within the physiological range (min: 1.63 ± 0.13 mm, max: 3.72 ± 0.30 mm). **A** comprised 48 fully-connected regions: 58, 55, 5, 60, 38, 23, 10, 61, 26, 6, 29, 43, 45, 59, 27, 7, 48, 34, 37, 50, 39, 57, 46, 16, 41, 42, 9, 13, 28, 56, 24, 18, 35, 49, 2, 12, 1, 47, 52, 11, 32, 25, 21, 44, 22, 54, 53, 20 (ordered according to their decreasing influence over **A**). **A** overlapped at the 73% with the previously reported first PN[1], and, notably, the overlap increased at the 78% if only highly-interconnected regions (weights ≥ 0.5) were considered. Table 1 shows that first PN calculation was reproducible within our data set, and that the PNs technique could reliably and accurately determine the existence of a connection between two regions (first row, columns 1 and 3) along with the associated weight (second row, columns 1 and 3) . The only exception was given by a group of 4 cortical areas (second column), for which values of **P** were lower and values of **D** were higher compared to the other regions. However, those areas had very low influence over our first PN, so we can reasonably think that their inclusion/exclusion affected the final network only marginally. Finally, from Figure 1 we observed that the overlap between the first PN determined from resampling and **A** was in direct proportion with the number of subjects involved. The standard deviation of the overlap, instead, showed the opposite trend. In particular, using 28-32 subjects, overlap of regions in the first PN was reproducible at the 90 ± 5 %. Interestingly, in the reference paper[1] PNs were determined from data of 28 healthy subjects.

DISCUSSION AND CONCLUSION: We suggest the PNs technique to calculate the brain main influential subnetwork along with a leave-one-out approach to cross-validate the result. We also underline that accuracy of the outcome is strongly related to the number of subjects involved. In fact, since this is a group technique, we always obtain an average result, which gets more robust and less sensitive to possible outliers as the number of subjects increases. Differences between our first PN and the reference[1] could have been caused by different imaging protocols and MRI scanners or by FreeSurfer cortical reconstruction process itself. FreeSurfer, in fact, has 0.5-mm accuracy, while thickness of the human cerebral cortex varies between 1 and 4.5-mm. However, good overlap of the networks obtained from two different data sets encourage future use and development of this technique.

REFERENCES: 1) J. D. Clayden et al. Principal Networks. PLoS ONE 2013. 2) V. Lippolis et al. Statistical modeling to assess the impact of cortical parameters on cognition in Multiple Sclerosis. Proc. Intl. Soc. Mag. Reson. Med. 2014;22:0959.
3) https://surfer.nmr.mgh.harvard.edu/fswiki/FsTutorial/AnatomicalROI/FreeSurferColorLUT