Clinically-Feasible Brain Morphometric Similarity Network Construction Approaches with Restricted Magnetic Resonance Imaging Acquisitions and their Relationship with Cognition

Daniel J. King[1] & Amanda G. Wood[1,2]

[1] School of Life and Health Sciences & Aston Neuroscience Institute, Aston University, Birmingham, B4 7ET, UK

[2] School of Psychology, Faculty of Health, Melbourne Burwood Campus, Deakin University, Geelong, Victoria, Australia

Author note

Correspondence concerning this article should be addressed to Amanda G. Wood, School of Life and Health Sciences & Aston Brain Centre, Aston University, Birmingham, UK. E-mail: a.wood4@aston.ac.uk

Abstract

Morphometric Similarity Networks (MSNs) estimate structural 'connectivity' as a biologically meaningful set of statistical similarities between cyto-architectural features derived in-vivo from multiple MRI sequences. These networks have shown to be clinically relevant, predicting 40% variance in IQ. However, the sequences required (T1w and T2w 3D anatomical, DWI) to produce these networks typically have long acquisition times, which are less feasible in some populations. Thus, estimating MSNs using features from only a T1w MRI is attractive to both clinical and developmental neuroscience. We aimed to determine whether reduced-feature approaches approximate the original MSN model as a potential tool to investigate brain structure. Using Human Connectome Project data, we extended previous investigations of reduced-feature MSNs by comparing not only T1w-derived networks but additional MSNs generated with fewer MR sequences to their full acquisition counterparts. We produce MSNs which are highly similar at the edge-level, to those generated with multi-modal imaging. We also find that, regardless of the number of features, these networks have limited predictive validity of generalised cognitive ability scores in contrast to previous research. Overall, settings in which multi-modal imaging is not available or clinically/developmentally appropriate, T1w-restricted MSN construction provides a valid estimate of the MSN.

Keywords: Morphometric similarity networks, Structural, brain development,

1. Introduction

Cortical grey-mater structural covariance networks (SCNs) model the degree to which the morphology of brain regions (measured by a single morphometric feature, cortical thickness or volume for instance) statistically co-varies across all possible pairs of regions of interest (ROIs; (Alexander-Bloch, Giedd, & Bullmore, 2013; Alexander-Bloch, Raznahan, Bullmore, & Giedd, 2013; Evans, 2013; Mechelli, Friston, Frackowiak, & Price, 2005). Whilst these types of networks represent region to region similarity of GM region metrics rather than causal interactions or tracked anatomical connections (Zheng et al., 2019), they are built on the premise that regions which are cytoarchitecturally similar are more likely to be anatomically connected (Goulas, Uylings, & Hilgetag, 2017; Wei, Scholtens, Turk, & van den Heuvel, 2019). These whole-brain network approaches to morphometric data, within a graph theoretic framework (Bullmore & Sporns, 2009), allow us to investigate additional information beyond that which is offered by univariate, local approaches (Bullmore & Sporns, 2009; Pagani, Bifone, & Gozzi, 2016).

The potential role of disruption to the SCN to understanding functional outcomes has been explored within a graph theoretic framework in relation to a range of conditions. These include broad psychiatric diagnoses such as bulimia, depression and schizophrenia (Chen et al., 2017; Mak, Colloby, Thomas, & O'Brien, 2016; Palaniyappan, Park, Balain, Dangi, & Liddle, 2015; Tijms et al., 2015; Westwater, Seidlitz, Diederen, Fischer, & Thompson, 2017), neurodegenerative disorders, such as Alzheimer's disease (AD) and multiple sclerosis (Kim et al., 2016; Pereira et al., 2015; Pereira et al., 2016; Raamana, Weiner, Wang, Beg, & Alzheimer's Disease Neuroimaging, 2015; Tewarie et al., 2014), epilepsies (Garcia-Ramos et al., 2017; Sone et al., 2016; Yasuda et al., 2015) and autism spectrum disorders (Balardin et al., 2015). In all of these studies, the methodology requires multiple participants to sample enough cortical measurements to generate a correlation between all possible regional pairs. Thus, this framework approach generates group-level brain networks, expressing population-level covariance in neuroanatomy (Alexander-Bloch, Raznahan, et al., 2013). This limits the ability of these approaches to quantify network- and system- level deficits within individual patients, which would benefit stratified diagnosis and prognosis (Zheng, Yao, Xie, Fan, & Hu, 2018).

Existing methodological approaches have attempted to investigate these structural relationships between regions at the individual-patient level (i.e. (Kim et al., 2016; Kong et al.,

2015; Kong et al., 2014; Tijms, Series, Willshaw, & Lawrie, 2012; Yu et al., 2018)). The majority of these methodologies have two major limitations; they either divide ROIs into sub regions that do not respect the underlying structure and convolutions of the cortex (Tijms et al., 2012), or the edge weights are defined as the simple subtraction of the feature in region A minus region B, rather than covariance. Both of these methodological deviations represent marked changes to the structural covariance paradigm under which many of the previous SCN validation studies have operated, potentially limiting the validity of these studies.

An alternative approach to investigate the covariance structure between multiple morphometric features can provide individual-level networks of covariance. Morphometric Similarity Networks (MSNs; Seidlitz et al. (2018)) estimate structural 'connectivity' as a biologically meaningful set of similarities between cyto-architectural properties at both the macro- and micro- structural level (Morgan et al., 2018). This is achieved through combination of features derived from a large set of imaging sequences, which may not always be possible in clinical settings. Data include morphometry measurements (such as cortical thickness, volume, curvature etc from T1w structural MRI), tissue diffusion properties (such as fractional anisotropy (FA) and mean diffusivity (MD) from diffusion-weighted images) and myelination indices (i.e. magnetization transfer from a multi-parameter mapping sequence or T1w/T2w ratio).

MSNs have been shown to be clinically useful, predicting ~40% variance in IQ, as well as being biologically meaningful, with edges of the MSN highly aligned with gene co-expression between regions in human data and with axonal tract tracing data in the rhesus macaque (Seidlitz et al., 2018). These findings likely reflect the fact that cortical regions that are less cortically differentiated from one another (that is, more anatomically similar) are more likely to also be anatomically connected (Goulas et al., 2017; Wei et al., 2019). Given the alignment between MSNs and other biological networks, these networks represent a new connectivity phenotype which may provide additional biologically-relevant information beyond existing network approaches.

MSNs have already been utilised in a small number of studies in clinical populations. For example, Morgan et al. (2018) used the multi-feature (grey matter volume, surface area, cortical thickness, gaussian curvature, mean curvature, FA, and mean diffusivity) network approach using both T1w and DWI MRI and found a robust and replicable pattern of differences in cortical grey-matter networks for patients with psychosis compared to controls.

Galdi et al. (2018) used a similar multi-feature model with macrostructural (volume and T1/T2 ratio) and multiple miscrostructural features (diffusion tensor-derived metrics and Neurite Orientation Dispersion and Density Imaging (NODDI) parameters). They trained a model to predict the post-menstrual age of infants born at term or pre-term. This model was able to detect a dysmaturation of the brain in the preterm infants, consistent with previous findings in similar cohorts. Seidlitz et al. (2019) also used MSNs to empirically test a 'transcriptional vulnerability model' of neurodevelopmental disorders of known genetic origin, with anatomical disruptions being spatially associated with regional gene expression within the region of the causal copy number variant. Overall, these findings seem to suggest that MSNs appear to offer a useful and clinically-relevant, individualised imaging phenotype.

Despite these existing clinical applications, it is important to note that multiple, high quality MRI sequences are required to recreate such methodologies. These may not be feasible for all research requirements and/or settings. For instance, in large existing clinical ('legacy') cohorts, the availability of this 'advanced' imaging may be limited or only a minimal number being consistent across multiple sites for instance. Also, due to the longer acquisition time of these MRI scans (especially DWI), the risk is that these MRI are more vulnerable to being of lower quality due to potential of movement artefacts over time for instance, especially in some paediatric or clinical applications where movement is more prevalent (Rosen et al., 2018).

Subsequently, estimating cyto-architectural similarity based on metrics from a single T1w 3D anatomical MRI, which is quickly and commonly acquired in clinical settings, is attractive to the fields of clinical and developmental neuroscience (Batalle, Edwards, & O'Muircheartaigh, 2018). Both Seidlitz et al. (2018) and Li et al. (2017) estimated connectivity in this way and found the edge weights of these networks to be similar to the multi-modal MSNs (r = .68, Seidlitz et al. (2018)), with 'good' test-retest reliability in terms of network topology (ICC = .60, Li et al. (2017)). However these networks had reduced precision in their estimation with greater standard deviation of edge-level weights seen across participants (Seidlitz et al., 2018). Of these previous studies, limited assessment has been conducted of the performance of these methods across characteristics of reliability, consistency with group-networks, biological validity and predictive ability. However very little attention has been given to directly comparing the performance of models with a reduced number of structural features with which the network is estimated. No previous study has conducted an assessment of the reliability and performance of models across a number of models, each using reduced number of

cytoarchitectural features indicative of a more restricted MRI acquisition sequence. These networks using only T1w MRI have already been seen in clinical applications. Zheng et al. (2019) generated networks using seven morphological features from T1w MRI. These networks were used to predict classification of ASD and controls. A machine learning approach using individual morphological features produced near-chance prediction accuracy, however, utilising only connection-weights from multi-feature networks there was a significant improvement in the model's prediction. Zheng et al. (2018) conducted a similar classification task and found that multi-feature MSNs classify patients with AD and mild cognitive impairment against controls, with a very high accuracy (~96%).

However, without an evidence-based comparison of MSNs constructed from only T1w MRI features and those constructed from a wider selection of MRI acquisitions, it is unclear as to whether the addition of added MRI sequences would necessarily lead to more reliable estimates of the network. If this were the case, then one would also posit that the increased reliability of MSN estimation would better position MSNs as a biomarker of brain structure, with less measurement error, and thus provide better prediction than simpler, T1w only models such as those in Zheng et al. (2019) and Zheng et al. (2018).

Recent research has shown that multi-feature MSNs are biologically meaningful and have potential clinical applicability, but MSNs generated with T1w features may be more amenable to certain patient groups/samples. The current study aimed to determine whether reduced-feature approaches approximate the 'original' MSN model as a potential tool to investigate brain structure. We extended previous investigations of reduced-feature MSNs by comparing not only T1w-derived networks, but additional MSNs generated with fewer MR sequences to their full-acquisition counterparts. No previous work has specifically investigated three MSN models, each using fewer metrics from a reduced number of specific MRI scan acquisitions, assessing a number of replication properties. These models were hierarchically organised, with reduced acquisition complexity from model a) to c) seen below;

a) MSN (T1w + T1w/T2w ratio + DWI; ten-features ($MSN_{10\text{-feat.}}$)),

b) MSN (T1w + T1w/T2w ratio; eight-features ($MSN_{8\text{-feat.}}$)),

c) MSN (T1w; seven-features ($MSN_{7\text{-feat.}}$))

Model a), hereto referred to as $MSN_{10\text{-feat.}}$, is the best approximation of the Siedlitz (2018) approach, with magnetization transfer replaced with T1w/T2w ratio mapping (Glasser & Van

Essen, 2011) in the current study. Thus, for each participant, three connectivity matrices (one per model) were estimated, across multiple thresholds. We predicted that, for each measure of reliability/replicability, performance would be ordered in a hierarchical fashion, with $MSN_{10\text{-feat.}}$ outperforming $MSN_{8\text{-feat.}}$ which subsequently outperforms $MSN_{7\text{-feat.}}$. However, we also predicted that between model comparisons would suggest that the models themselves were highly similar. We also predicted that we would conceptually replicate previously found associations between cognition and MSN organisation (Seidlitz et al., 2018) and that we could generalise this finding to a novel domain of cognition, specifically executive functioning.

2. Methods

2.1 Participants - HCP data

The current study uses open access, 3T MRI data provided by the Human Connectome Project (Van Essen et al 2013, Neuroimage), shared via ConnectomeDB (https://db.humanconnectome.org) under the HCP1200 and HCP Test-retest release. Favourable ethical approval for the secondary analysis of this data was granted by the Aston University ethics panel.

2.1.1 HCP 1200 Release

The HCP 1200 release contains data from n = 1206 subjects (550 Males, 656 Females). Subjects are grouped into age bins from '22-25' to '36+' (median age = 26-30). Whilst n = 1206 subjects provided behavioural data, only 1113 subjects had MRI data available. These were the subjects for which data was accessed and downloaded from ConnectomeDB for the current study.

2.1.2 HCP Test-Retest Release

For 46 subjects from the HCP-1200 release, a second 'retest' dataset is available to assess test-retest reliability of analyses. These second MRI visits occurred within time bins from '1-2 months' to '11 months' post initial scanning session. The median retest-interval bin was '5 months'. Of these subject 45 had available MRI data, and these were the subjects used for subsequent analyses.

2.2 Methods

2.2.1 Data Quality Control

Subjects were selected for inclusion if, in the 1200-subject HCP release, they had T1w, T2w and diffusion data uploaded. This led to exclusion of $n = 76$ cases.

Also, utilising QC data shared by the HCP project, any data labelled as with QC issue code B (which flags cases as having focal segmentation and surface errors when the corresponding Freesurfer outputs were checked) was further excluded from the current study ($n = 33$). The final dataset consisted of $n = 1004$ subjects. In the test-retest cohort, only one subject was excluded as flagged with QC issue B by the HCP project.

2.2.1 MRI Processing

The current study utilises data shared in its pre-processed format, including the output of the HCP Freesurfer pipeline (Fischl et al., 2002; Glasser et al., 2013; Jenkinson, Bannister, Brady, & Smith, 2002; Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012), processed DWI (gradient non-linearity, eddy-current and EPI distortion corrected (Andersson, Skare, & Ashburner, 2003; Andersson & Sotiropoulos, 2015, 2016), and calculated T1/T2w ratio myelin maps (Glasser & Van Essen, 2011). For further details of HCP processing pipelines see Glasser et al. (2013).

Once cases were selected, measures indexing the underlying cyto-architecture were derived from multiple imaging modalities (see Table 1). Seidlitz et al. (2018) leverage near-identical MRI-derived metrics for the construction of the MSN network. However, we are using the T1/T2 ratio as a proxy for myelin content, rather than the magnetization transfer scan used by Seidlitz et al. (2018). The rationale for this modification was both pragmatic and clinically-driven; i) the T1/T2w ratio maps are already implemented by the HCP project and thus this data is available for use with the rest of the high-quality HCP acquisition data and ii) in clinical populations, for which the methods may provide greatest benefit, multi-parameter mapping MRI sequences may not be acquired as part of a clinical protocol, whereas T1w and T2w sequences are.

Table 1. Morphometric measures and the modality of MRI from which they were derived

Modality	Metrics
T1w	Cortical thickness (CT), surface area (SA), mean (extrinsic) curvature (MC), Gaussian (intrinsic) curvature (GC), folding index (FI), curvature index (CI) and grey matter volume (GMV)
T2w	Myelination (T1/T2w ratio)
DWI	Fractional Anisotropy (FA), Mean Diffusivity (MD)

Preprocessed DWI (b = 1000) in T1w space were fitted to a tensor model using FMRIB's 'dtifit' function, and the subsequent FA and MD maps were mapped to the individual subject's Freesurfer generated surface model in MNI space, using the connectome workbench (Marcus et al., 2011) function 'volume-to-surface-mapping'. These, and the Tw1/T2w ratio myelin maps, were parcellated based on the Desikan-Killany atlas (Desikan et al., 2006), by generating a dense-cifti (using the 'cifti-create-dense-from-template' function) and parcellating the output (using 'cifti-parcellate'). Freesurfer metrics were also extracted for each parcellated region using the 'aparcstats2table' function.

2.2.2 MSN Construction

To generate MSNs we apply the methods of Seidlitz et al. (2018) to the HCP data. The Desikan-Killany atlas was mapped to the individual subjects with a surface-based registration, using the Freesurfer pipeline. The Desikan-Killany atlas ROIs were used as the nodes for all network construction.

Morphometric features (parcellated to the Desikan-Killany atlas) for each participant can be expressed as a set of n vectors of length 10, with each vector as a different anatomical region (n = 68), and each element of the vector a different morphometric measure. However, these features are not all measured at the same magnitude of scale. For instance, volume (mm^3) is measured at the order of 10^3, whereas folding index is measured to the order of 10^1. Thus, to normalize within this length 10 vector, each of these morphometric features is normalized across the 68 regions, using Z-scores (demeaned and SD scaled). This brings the measures across the feature vector into a comparable range.

Using the normalized features, a correlation matrix is generated for each participant, where each element of the matrix is the correlation between the feature vectors for every possible pairwise combinations of regions. Because each feature is zero-centred, the resultant distribution of correlation coefficients is normally distributed about zero. This correlation matrix represents the MSN-estimated connectivity for each participant. This procedure was repeated across the three MSN models ($MSN_{10\text{-feat.}}$, $MSN_{8\text{-feat.}}$, and $MSN_{7\text{-feat.}}$), each using fewer metrics from a reduced number of scan acquisitions.

2.3 Demographic and Behavioural Data

Demographic variables were selected from the unrestricted data table accessed via 'ConnectomeDB'. These included age bin, sex recorded at birth and recorded quality control issues. Behavioural data were also extracted to assess the relationship between the MSNs and both general cognitive ability (measured with both fluid and crystallized intelligence measures) and executive functioning. These neuropsychological assessments were conducted contemporaneously in relation to the MRI scans. Further details of the tasks and measures acquired in the HCP dataset can be found in (Barch et al., 2013).

2.3.1 General Cognitive Ability

General cognitive functioning is measured with the Cognitive Function Composite (CogComp) score (Heaton et al., 2014), derived from the average of the normalized, scaled scores of Fluid and Crystallized cognition measures, then subsequently age-adjusted, and scaled. The Fluid Cognition Composite score is derived by averaging the normalized scores of each of the fluid ability measures in the NIH-toolbox (Flanker, Dimensional Change Card Sort, Picture Sequence Memory, List Sorting and Pattern Comparison), whilst the Crystallized Cognition Composite score is derived by averaging the normalized scores of each of the crystallized measures in the NIH-toolbox (Picture Vocabulary and Reading Tests). Higher Cognitive Function Composite scores indicate higher levels of cognitive functioning.

2.3.2 Executive Functioning

Behavioural executive function (EF) measures were selected based on an evidence-based, 3-factor model of executive function (Karr et al., 2018); measures selected from the HCP cognitive battery to model EF were the same as previous studies of EF utilising the HCP data (Lerman-Sinkoff et al., 2017; Nomi et al., 2017). These tests assessed multiple cognitive aspects of executive functioning including cognitive flexibility/shifting (Dimensional Change

Card Sort test,(Zelazo, 2006; Zelazo et al., 2014)), inhibition (Flanker Inhibitory Control and Attention task, (Zelazo et al., 2014)), working memory (List Sorting task, (Tulsky et al., 2013)). Age-adjusted scores were used for all behavioural data.

Due to the fact we have only one neuropsychological measure per sub-domain of EF and there is therefore potential risk of measurement error, a principal component analysis (using the 'prcomp' function in the R 'stats' base package (R Core Team, 2016)) was used to find a common EF component across all three EF measures. This produced a single principal component with an eigenvalue above 1, upon which all measures positively loaded onto, and thus this component was used as a 'summary' score of EF (see supplementary materials for further details). Higher summary EF scores reflect greater EF functioning.

2.4 Statistical comparison

When comparing weighted networks produced by each model, we use multiple metrics to assess the (dis)similarity of the subsequent covariance matrices.

To reduce number of comparisons and, based on our premise that the $MSN_{10\text{-feat.}}$ is the most precise estimation of the MSN network (as shown by Seidlitz et al. (2018)), all inter-model comparisons were done in a hierarchical fashion in comparison to this 'gold-standard' network. That is to say that model $MSN_{10\text{-feat.}}$ was compared to the $MSN_{8\text{-feat.}}$ and then the $MSN_{10\text{-feat.}}$ was subsequently compared to the $MSN_{7\text{-feat.}}$.

In order to test differences in the topological organisation of the networks produced by each model, we calculate average nodal strength for each graph. Nodal strength is the 'magnitude' of structural covariance for each node, this is the sum of the connectivity weights of all edges connected to node *i* (Fornito, Zalesky, & Bullmore, 2016). We did not normalize this measure based on number of edges as we averaged the nodal measures over the graph, where the number of edges was consistent across models due to density thresholding. This metric was calculated per subject, per density for each MSN model. For each comparison, we calculate the difference in distributions of graph strength using a paired t-test test. Due to the large number of comparisons (across densities, and contrasts) we do not report p-values, but instead report the effect sizes for comparisons.

We also calculate the Pearson correlation coefficient between all edge weights for both models (as per Seidlitz et al. (2018)), and also specifically between all non-zero edge weights (those elements where a zero is present in the correlation matrix for each model are excluded).

However, because of the symmetric, undirected nature of the correlation matrix, this correlation coefficient may inflate/bias the supposed 'similarity' between the sets of edge weights. Thus we also employed the Mantel test, which calculates the Pearson correlation on either half of the off-diagonal elements of the correlation matrix (Mantel, 1967).

To compare the binary networks produced by each model at each density (where edges retained after thresholding are set to 1 and those excluded are set to zero), we assessed the number of edges in the reduced model which replicated as a proportion of the fuller model, as per the following formula:

$$\frac{\sum(x_i \neq 0 \ \& \ y_i \neq 0)}{\sum(x_i \neq 0)}$$

where x_i and y_i represent the correlation matrices estimated from two of the MSN models for a given subject i.

Secondly, we calculate these similarity measures between the subject-level network and the group average network, across all densities and models. This allows the assessment of the inter-subject reliability of the networks being constructed by each model. Thirdly, we similarly test the intra-subject reliability of the produced networks, based on test-retest data from a subset of the overall dataset. Due to the categorical and inaccurate nature of the 'binned' measurement of time between initial and retest scan, this was not controlled for in this analysis.

In order to assess the functional relevance of these networks, we assess their ability to predict CogComp and EF scores using a supervised-learning approach, namely partial least squares (PLS) regression (similarly to Seidlitz et al. (2018)) using the 'plsRglm' package in R (Bertrand & Maumy-Bertrand, 2018). This multivariate approach finds the optimal low dimensional relationship between a high dimensional set of predictors (in this case the MSN networks) and a univariate predictor variable (either CogComp or EF). This approach is commonly use when the number of predictors exceeds the number of observations (Krishnan, Williams, McIntosh, & Abdi, 2011).

A PLS regression was used to find the maximal low-dimensional covariance between components derived from the MSN and cognitive outcomes. The PLS regression was used to decompose the predictor variables into latent variables (components) which simultaneously model the predictors and predict the response variable (Krishnan et al., 2011). The predictor matrix consisted of either the degree or strength of each node of the MSN, for each participant.

Using a linear model, the potential confounding effect of age, gender and age*gender interaction was regressed out of values for nodal degree/strength (but not our cognitive outcome variable as these were already age-adjusted within the HCP dataset). For each model (at each threshold), a PLS regression model was fitted between principal components derived from the resultant predictor matrix (68 x 991) and the outcome variable. This was repeated across 100 instances of 9-fold cross-validation.

Cross-validated R^2 (R^2_{CV}) otherwise known as the Q^2 statistic (Consonni, Ballabio, & Todeschini, 2010; Stone, 1974), was used to select the number of components to retain in the predictor matrix. Q^2 was defined as:

$$Q^2 = R^2_{CV} = 1 - \frac{PRESS}{TSS} = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

where PRESS is the predictive residual error sum of squares and TSS is the total sum of squares.

The number of components to retain in the predictive model was selected as the number of components which resulted in the greatest Q^2 value. This was repeated over the cross-validations and resulted in a count measure of the number of times a model with a given number of components were selected. Hence the final model was the given number of components which was most commonly selected as having the greatest Q^2 statistic. Given the model with the retained number of components, we report the variance explained by the model and the bias corrected and accelerated bootstrapped (Bastien, Vinzi, & Tenenhaus, 2005) weightings of each predictor. This allows us to assess which brain regions are contributing most to the prediction.

Due to the normal distribution of the cognitive measures (CogComp and EF) data, there may be an issue of class-imbalance for more 'extreme' cases (Torgo, Branco, Ribeiro, & Pfahringer, 2015). As there are fewer subjects who fall within the tails of the continuous distribution on our cognition measures, the cross-validation approach may lead to training samples where there are too few 'extreme' cases (those with particularly high/low cognitive abilities) to 'learn' from. This may result in a model where there is accurate prediction around the mean but not at the tail ends of the distribution. To ensure the training samples contain subjects from stratified sampling approach, we repeated the analyses discretizing the performance on cognitive

measures into four discrete bins across the distribution and training a model based on equally-sized, random samples from each bin.

3. Results

3.1 Inter-model comparisons

3.1.1 Magnitude of morphometric similarity: graph-level strength

In terms of the topology of the networks, global graph strength for each model, across densities, can be seen in Figure 1. This plot shows the similar trajectories across densities for all models tested, however the observed average graph strength was different between models, with lower strength being see in the MSN models with greater features. The effect size of differences (estimated with a paired t-test) between $MSN_{10\text{-feat.}}$ vs $MSN_{8\text{-feat.}}$ and $MSN_{10\text{-feat.}}$ vs $MSN_{7\text{-feat.}}$ can be also be seen in Figure 1. Effect sizes (r) were extremely large, especially between $MSN_{10\text{-feat.}}$ vs $MSN_{7\text{-feat.}}$.

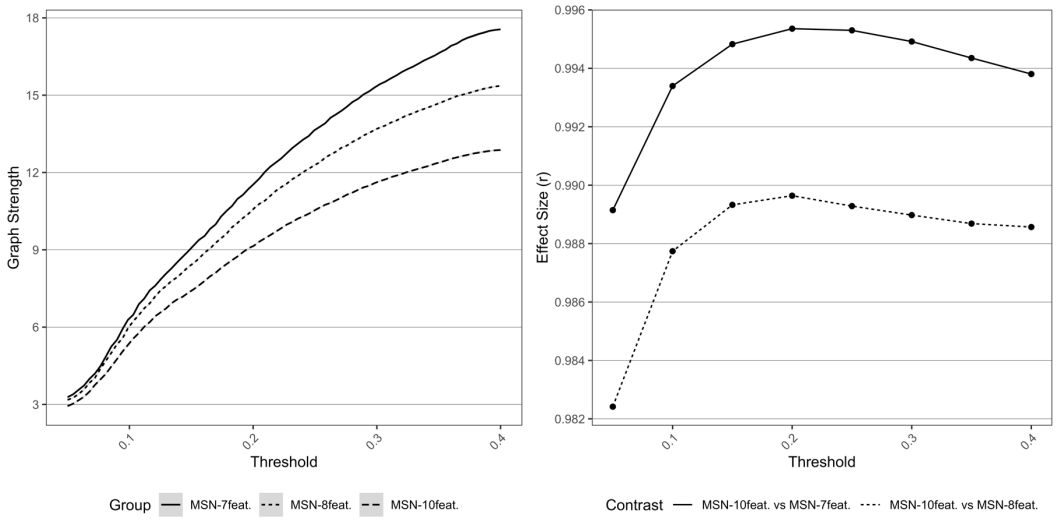


Figure 1 Left: Graph metrics describing average network strength for each MSN model, across all densities. Right: Effect sizes of differences between a) $MSN_{10\text{-feat.}}$ vs $MSN_{8\text{-feat.}}$ and b) $MSN_{10\text{-feat.}}$ vs $MSN_{7\text{-feat.}}$. for differing graph metrics, across densities.

3.1.2 Edge Weights

Figure 2 shows the inter-model comparisons between $MSN_{10\text{-feat.}}$ and $MSN_{8\text{-feat.}}$, and between $MSN_{10\text{-feat.}}$ and $MSN_{7\text{-feat.}}$. There is a gradual increase in correlation of edge weights across densities with the peak mean correlation being found between $MSN_{10\text{-feat.}}$ and $MSN_{8\text{-feat.}}$ at a 40% threshold ($r(M\pm SD)$ = .849 (± .025)), with slightly weaker correlations found between $MSN_{10\text{-feat.}}$ and $MSN_{7\text{-feat.}}$ ($r(M\pm SD)$ = .736 (± .031)). When considering only the non-zero edge weights (only edge weights remaining after thresholding), a slightly weaker peak correlation was found for both contrasts at 5% threshold ($MSN_{10\text{-feat.}}$ vs $MSN_{8\text{-feat.}}$ $r(M\pm SD)$ = .738 (± .053); $MSN_{10\text{-feat.}}$ vs $MSN_{7\text{-feat.}}$ $r(M\pm SD)$ = .670 (± .066)). However, as the threshold increased, the dispersion of individual level non-zero edge correlation decreases, especially in the $MSN_{10\text{-feat.}}$ vs $MSN_{7\text{-feat.}}$ contrast.

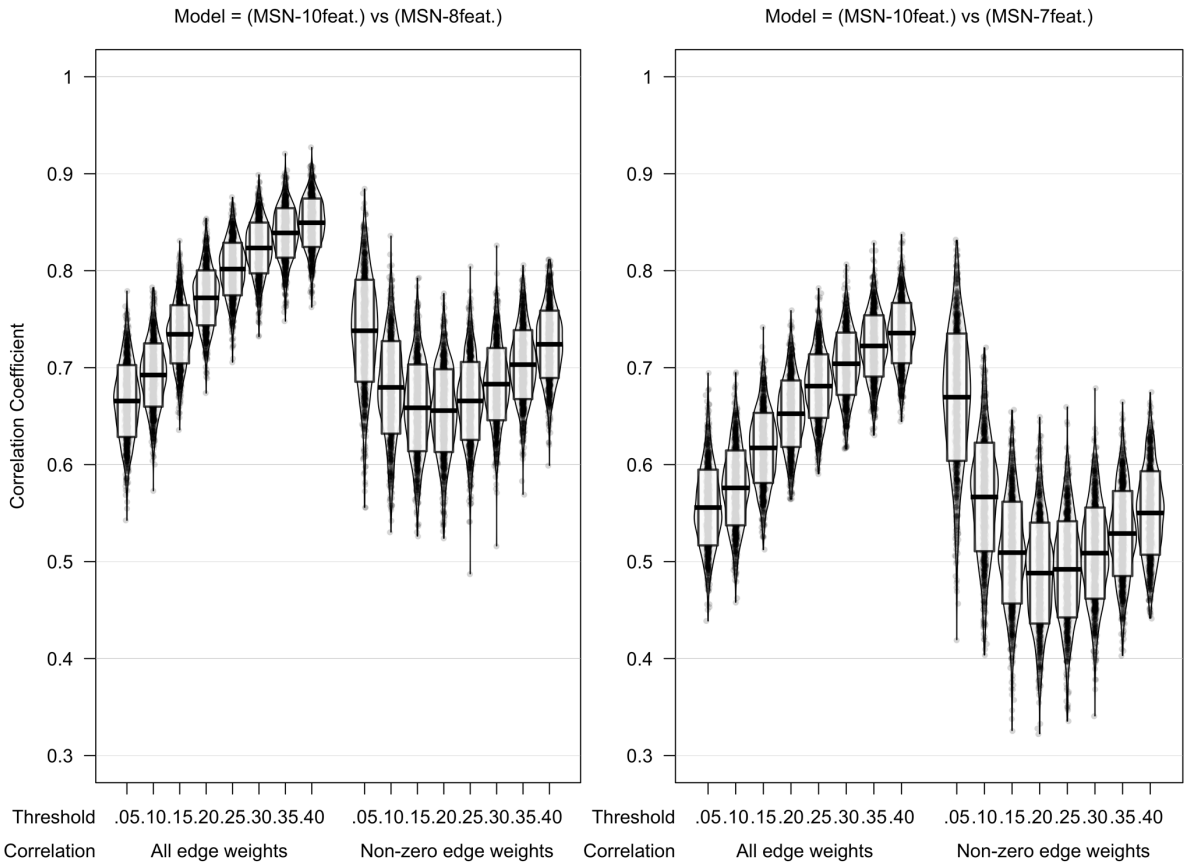


Figure 2 Violin plot of correlation of edgeweights between a) $MSN_{10\text{-feat.}}$ vs $MSN_{8\text{-feat.}}$ and b) $MSN_{10\text{-feat.}}$ vs $MSN_{7\text{-feat.}}$. Midline of the box-plot component of the violin represents the mean of all correlation coefficients, with the box itself representing the SD of this mean. Individual data points are also plotted.

When considering correlation coefficients calculated using the Mantel test, similarly strong correlations were found between edge weights across all models however, as predicted, the

$MSN_{10\text{-feat.}}$ vs $MSN_{8\text{-feat.}}$ were most similar (At 40% threshold: $MSN_{10\text{-feat.}}$ vs $MSN_{8\text{-feat.}}$ Mantel $r(M\pm SD)$ = .835 (± .028); $MSN_{10\text{-feat.}}$ vs $MSN_{7\text{-feat.}}$ Mantel $r(M\pm SD)$ = .715, (± .034)). For the binarized networks, the proportion of edges replicated also peaked at 40% threshold ($MSN_{10\text{-feat.}}$ vs $MSN_{8\text{-feat.}}$ proportion of replicated edges = 85%, (± 2%); $MSN_{10\text{-feat.}}$ vs $MSN_{7\text{-feat.}}$ proportion of replicated edges = 77%, (± 2%;. Figure 3)).

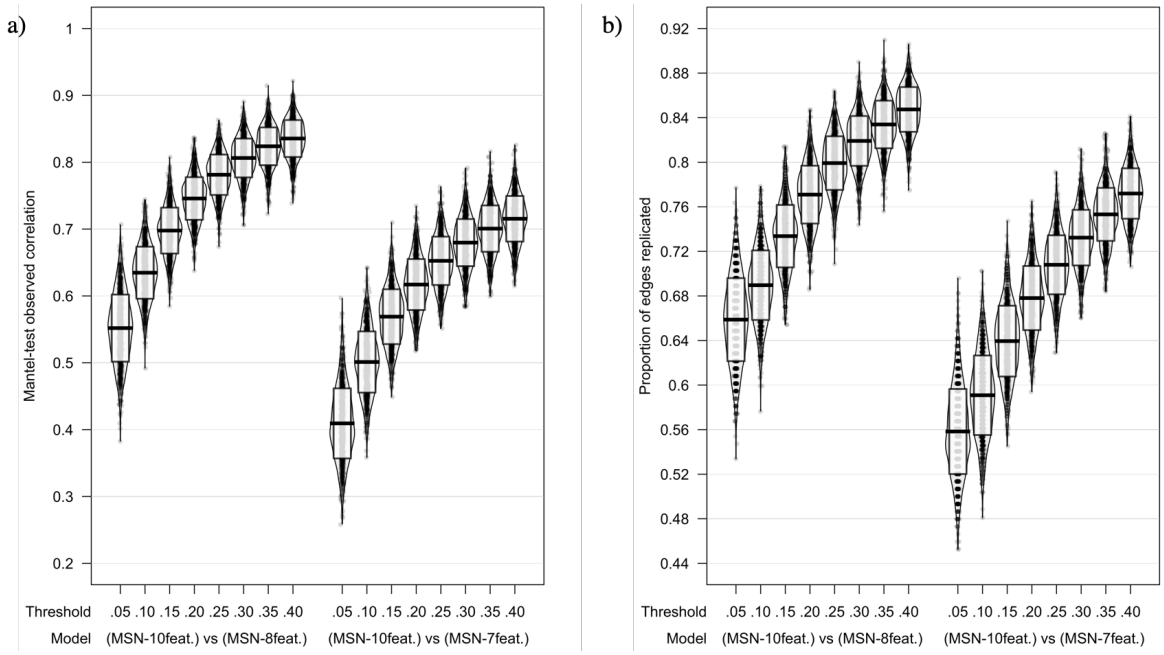


Figure 3 Model comparisons across thresholds using a) Mantel-test correlation coefficient and b) proportion of edges replicated as measures of model similarities. Midline of the box-plot component of the violin represents the mean whilst the box itself representing the SD

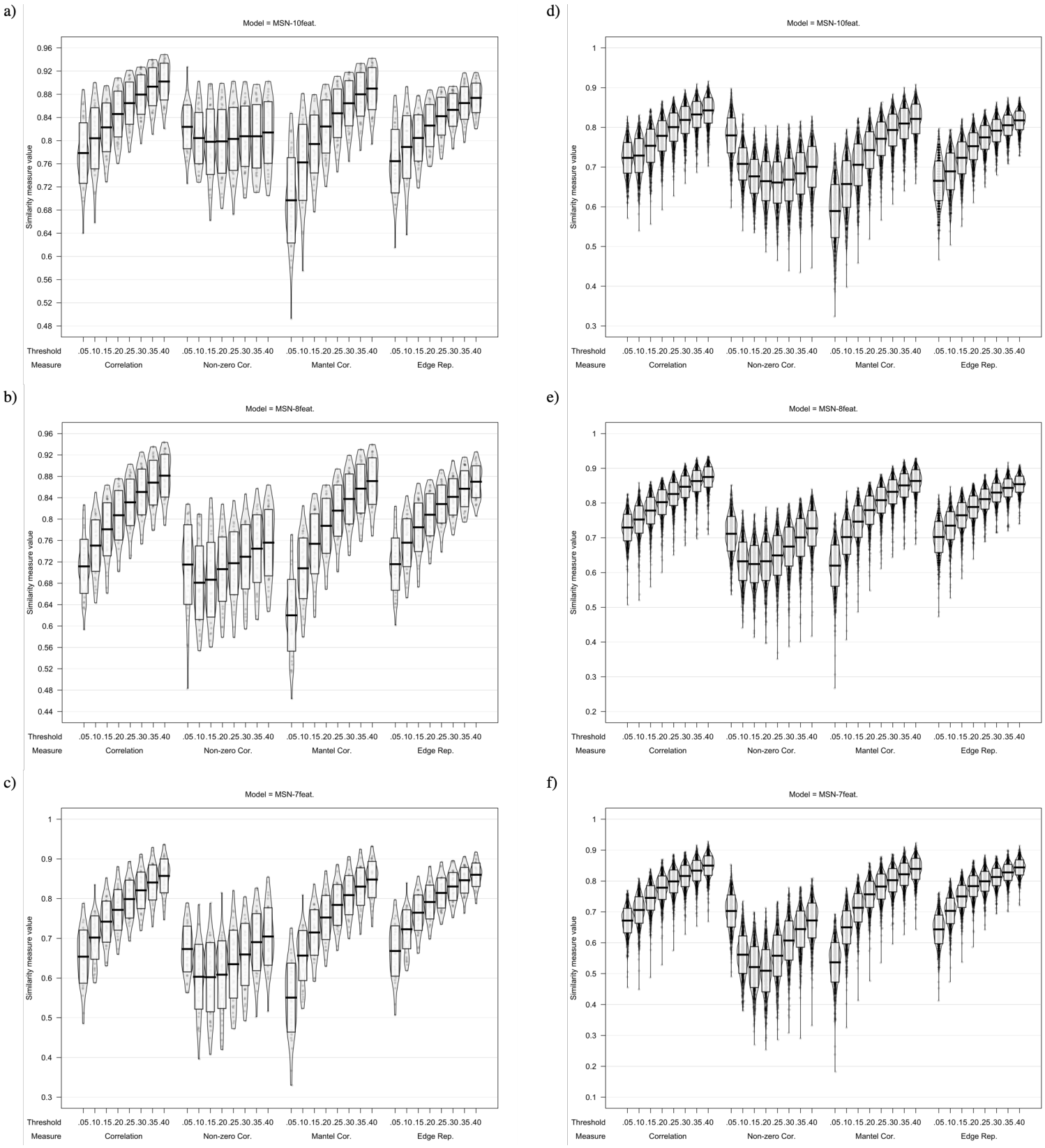
3.2 Intra-model comparisons

3.2.1 Test-retest reliability of MSN models

We compared the MSN models at the initial scan with those calculated from test-retest scans acquired between 1 and 11 months after the initial MRI. All models showed high test-retest reliability of the MSN (correlation of all edge weights at 40% threshold: $MSN_{10\text{-feat.}}$ $r(M\pm SD)$ = .902 (± .032); $MSN_{8\text{-feat.}}$ $r(M\pm SD)$ = .881 (± .040), $MSN_{7\text{-feat.}}$ $r(M\pm SD)$ = .857 (± .043)). This high test-retest reliability of networks held even when networks were binarized (At 40% threshold: $MSN_{10\text{-feat.}}$ proportion of replicated edges = 87 % (± 3%); $MSN_{8\text{-feat.}}$ proportion of replicated edges = 87% (± 3%), $MSN_{7\text{-feat.}}$ proportion of replicated edges = 86% (± 3%)). See Figure 3 for plots.

3.2.2 Similarity with average MSN

For each model, at each threshold, a group-level network was produced as the mean of the correlation matrices for all subjects. Across all models ($MSN_{10\text{-feat.}}$, $MSN_{8\text{-feat.}}$, and $MSN_{7\text{-feat.}}$), regardless of similarity metric used, the individual-level MSNs were highly similar to the group-mean network (see Figure 4). Interestingly, the $MSN_{8\text{-feat.}}$ model showed greatest correlation between edge weights (At 40% threshold: $MSN_{10\text{-feat.}}$ $r(M \pm SD)$ = .843 (± .032); $MSN_{8\text{-feat.}}$ $r(M \pm SD)$ = .875 (± .029), $MSN_{7\text{-feat.}}$ $r(M \pm SD)$ = .850, (± .031)). Similar to the inter-model analyses, correlation peaked at the highest threshold tested (40%) for all models.



435 Figure 4 Plots showing MSN similarity (across thresholds, with multiple similarity measures) between
436 a,b,c) individual MSNs generated with test-retest MRI scans and d,e,f) individual-level MSNs and the
437 group-average MSN network.

3.3 Relationship with cognitive scores

Only participants who had available a full dataset comprising of the three EF subtests and the CogComp measure were included in the following analyses (n = 991). For both cognitive variables, using 100 instances of 9-fold cross validation, the greatest Q^2 was found most frequently when zero-components were retained and thus no models were built.

This suggests that no PLS-derived components of nodal degree, strength or normalised strength of the MSN provided greater explanation than the intercept alone. After the stratified sampling of the training cohort, there was no improvement in the result outlined above; cross-validation still recommended retention of zero components for all MSN models.

4. Discussion

Within the morphometric similarity network model, we assume that those regions which are high in morphometric similarity have high concordance of cyto- and myelo- architectural features at a resolution unobservable in-vivo with current MRI capabilities (Morgan et al., 2018). These cortico-cortico regions which are less cortically differentiated from one another are more likely to be anatomically connected (Goulas et al., 2017; Wei et al., 2019). However, the methods presented here are not causal, the represent the region to region similarity in terms of the GM morphology of the cortex (Zheng et al., 2019). Whilst Seidlitz et al. (2018) and Li et al. (2017) performed some assessment of T1w MSNs, the current study is the first to formally investigate the potential for generation of multiple MSNs based on a reduced number of cyto- and myelo- architectural features dependant on the complexity of the MRI acquisition sequence. We found that the weighted networks generated from these models are highly similar, across a number of correlation measures investigating edge weightings. Overall our results suggest that these meso-scale relationships can be captured (to a considerable degree) within a more limited number of cyto- and myelo- architectural features from a lesser number of MR-sequences.

Seidlitz et al. (2018) investigated the similarity of a T1w MSN (using only 5 morphometric features compared to our 7) with the full $MSN_{10\text{-feat.}}$ model and found a high level of similarity, although the $MSN_{10\text{-feat.}}$ model had a greater level of precision with a lower standard deviation of edge weights. Seidlitz et al. (2018) also did not systematically investigate the consequences of removing MRI acquisitions from the features with which to estimate the MSN model.

In the current study we expanded previous comparisons of T1w MSNs to the 'origiinal' MSN model to include multiple MSN models. We found that the between-model similarity was nearly always hierarchical between models, with greater similarity seen between $MSN_{10\text{-feat.}}$ and $MSN_{8\text{-feat.}}$ compared to that between $MSN_{10\text{-feat.}}$ and $MSN_{7\text{-feat.}}$. Weaker similarity was found for sparser networks at a much lower density (i.e. .05). Even when binarized (that is to say the edge weightings were ignored) the replication rates were high, suggesting that the models are sensitive to specific edges within the network.

However, our results show that, in terms of average network strength, the three models differed significantly in their topology. Whilst previous studies had investigated the correlation between nodal similarity for full and reduced models of MSN estimation (Seidlitz et al., 2018), this is the first study to investigate differences in this topology. On average, the magnitude of morphometric covariance across the nodes of the graph are higher when fewer features are used to generate the network. The topology of networks generated from different MSN models is fundamentally different and, dependant on metric used, this difference can be of a large effect size. Hence, as more cytoarchitechtural features are added to the MSN, specifically estimated myelin content (T1w/T2w ratio) and macro-structural diffusion properties (FA & MD), regions appear less similar and more differentiated, hence the lower average graph strength. This may because these features index cytoarchitectural properties which show greater variation, and are more discriminatory between regions, across the cortex. This difference in network topology is important to consider, as it means that network topology between these models is not comparable across studies.

Each model seemed to achieve high-levels of congruence with the group average network, suggesting that we are able to use these methods to index individual differences from a relatively consistent meso-scale-phenotype of the structure of the brain. Li et al. (2017) found high levels of test-retest reliability of the T1w MSN, we replicated this and found that each of the reduced-feature MSNs seemingly had similar reproducibility in terms of test-retest MRI.

It is important to consider that none of the models tested in the current manuscript showed perfect or even near-perfect concordance across these measures of performance. These between-model differences may be due to the fact that these models are generated with less features, rather than being specific to the modality of feature being dropped. Beyond the scope of the current paper but could look at this in future by generating MSN with 10, 8 and 7 randomly selected features, irrespective of modality of MRI sequence used to derive said

feature. If this is the case, then the 'gap' between the $MSN_{10\text{-feat.}}$ and $MSN_{7\text{-feat.}}$ models could potentially be rectified using software such as 'mindboggle' to generate/sample a larger number of morphometric features from the T1w image.

Overall, our findings suggest that, even with a reduced number of cytoarchitectural features, the MSN seems to capture a group-level phenotype of the structure of the brain which shows a reasonable level test-retest reliability. However, whilst these models may capture enough shared variance to be meaningful in a number of fields, it must be considered that the loss of information due to a reduced number of MR-acquisitions may result in a 'noisier' measure of the connectivity phenotype being indexed by the MSN approach. This will inherently limit generalisability across findings utilising these methods.

However, the main benefit of the reduced MR-acquisition approaches (specifically the $MSN_{7\text{-feat.}}$ model) is the applicability to those populations where multiple MR sequence acquisition is more challenging or difficult. For instance, in clinical populations where research MRI are acquired alongside routine examination and therefore time is limited, or in developmental populations where acquisition time needs to be kept short in order to ensure child participants can remain still for the length of the scan to ensure the images are free of motion artefact. Estimating cyto-architectural similarity based on metrics from a single T1w 3D anatomical MRI, which is commonly and quickly acquired clinically, is therefore particularly attractive to the field of clinical and developmental neuroscience (Batalle et al., 2018). It also validates these models for use in legacy datasets for instance, where the full array of MRI acquisition sequences required to estimate the 'original' MSN were not acquired and are therefore not available. Overall, the current study validates the use of these reduced-feature networks in recent studies estimating cyto-architectural similarity utilising theMSN (Galdi et al., 2018; Li et al., 2017; Morgan et al., 2018; Seidlitz et al., 2019; Zheng et al., 2019; Zheng et al., 2018).

One could argue that one-acquisition connectivity is already available in the form of DWI tractography, or even fMRI resting state connectivity. However, these are still much longer sequences compared to a 3D T1w MPRAGE for instance and therefore face inherent difficulties in the face of clinical realities of restricted time and potentially greater motion. Also, both fMRI and DTI inevitably suffer from a lower signal-to-noise ratio and a greater sensitivity to motion artefacts compared to anatomical MRI (Wang, Jin et al 2016 #614). It could also be argued that, in terms of legacy/existing datasets, it is more likely that a high quality, 3D T1w MRI has been acquired than the specific DWI/fMRI protocol required.

Overall, this therefore positions MSNs as a useful in-vivo connectivity phenotype for studying both clinical and developmental populations, with the T1w-only model potentially being of greatest potential benefit.

These approaches have potential utility in these fields of research, with one use being assessing relationships between brain structure and neuropsychological functioning. The current zeitgeist in the field of cognitive neuroscience is that the topological organization of the brain networks (across multiple MR modalities), as quantified within a graph theoretic framework, captures physiologically relevant information (Bullmore & Sporns, 2009; Fornito, Zalesky, & Breakspear, 2013; Hahn, Lanzenberger, & Kasper, 2019). However, a recent study failed to replicate one of the most prominent findings for the field relating rsfMRI connectivity to fluid and crystallized intelligence in the HCP dataset (Kruschwitz, Waller, Daedelow, Walter, & Veer, 2018). The current study investigated this by assessing the relationships between cognition and organisation of the MSN models.

We assessed the predictive validity of the MSN models in the current study by comparing the predictive validity of the 3 MSN models in relation to general intelligence, with previous research suggesting the organization of the MSN network (modelled similarly to the $MSN_{10\text{-}feat.}$) was able to predict ~40% variance in WASI IQ (verbal and non-verbal, (Seidlitz et al., 2018)). We were unable to replicate the predictive validity of the MSN with regard to general cognitive functioning or generalize previous relationships to a novel domain of cognitive functioning (in this case executive functioning). Our results showed that, when using 9- fold cross-validation, no model (at any density) recommended retention of any PLS components.

One important strength of the current study is the fact that we used a quantitative methodology of cross-validation to validate retained number of components whereas previous studies have retained either a single or two components which explains the greatest amount of variance (Seidlitz et al., 2019; Seidlitz et al., 2018). This may mean that previous findings are less generalizable to new datasets, hence why we were unable to replicate findings of Seidlitz et al. (2018), and instead found that nodal topological characteristics (i.e. strength) did not predict cognitive abilities in the current sample.

However, there are several other potential hypotheses as to why we were unable to replicate the previous findings. Most importantly, there were developmental differences between our sample and that of Seidlitz et al. (2018). The current study investigated a healthy young adult population between the 3rd and 4th decades of life whereas Seidlitz et al. (2018) studied a late

adolescent (15-25yrs) sample. The brain undergoes substantial structural change over development with this adolescent period being a time of peak maturation (Gogtay et al., 2004; Sowell et al., 2004) It is across these years in which some of the neurocognitive skills investigated in the current study, executive functioning for instance, are fully established. For instance, the NIH-toolbox total cognition composite highlights this quite clearly with a greater magnitude of age effects seen in childhood compared to adulthood (Akshoomoff et al., 2013; Heaton et al., 2014). This is likely because, throughout childhood, the regions subsuming these functions are reaching structural maturity. Therefore, it is reasonable to believe that, it is within the child/adolescent period where the most variance in these neurocognitive skills can be explained by structural networks (as seen by the ~40% variance in IQ explained by the MSN in Seidlitz et al. (2018)).

In the age-range that the current study has sampled, the brain should have reached structural maturity (with only mild age related effects in this age-group) and so there is likely less between-individual variance in the MSN. This was seen in the fact that there was greater congruence between individual MSNs and the group-average MSN in the current study compared to previous adolescent MSNs (correlation of all edge weights: mean r = .60, (Seidlitz et al., 2018)). Therefore, the limited variance in the MSN within this age group may mean that there is not enough variance to relate to cognitive functioning, hence our current findings.

We therefore propose that the MSN may in fact be a useful phenotype for assessing neuropsychological functioning, but only in populations where there is sufficient variation in the structure of the brain. This may be populations in the infant/child/adolescent period where structural networks are likely to see greatest variability due to developmentally-mediated change (such as Galdi et al. (2018) & Seidlitz et al. (2018)) or clinical populations where atypical brain structure is seen in the pathophysiology of the disorder (such as Seidlitz et al. (2019), Morgan et al. (2018)& Zheng et al. (2019)). It may be the case that these networks hold utility in populations such as these, rather than healthy, matured populations (where measures of brain structure are likely to heavily regress to the mean), where these methodologies may be of much lesser utility in explaining cognitive functioning.

However, it is also important to consider that the variation in our results could be due to other variations in analysis. Firstly, differences may be driven as an artefact of using differing measures of general intelligence, with Seidlitz et al. (2018) utilising the Weschler Abbreviated Scale of Intelligence (WASI; (Wechler, 1999)), whilst we used the NIH Toolbox Cognition

composite scores (Heaton et al., 2014). However, it is important to remember that the composite score shows high convergent validity with other Weschler assessments of general intelligence (with the Weschler Adult Intelligence Scale (WAIS-IV, (Wechler, 2008)) r = .89 (Heaton et al., 2014), and with the Weschler Intelligence Scale for Children (WISC-IV; (Wechsler, 2003)) r = .88 (Akshoomoff et al., 2013).

Also, we calculated the MSN at a much lower spatial scale (68 ROIs) compared to this previous work (308 ROIs). This lower spatial resolution may result in more regionally specific effects being difficult to detect, however it may also have allowed us to detect more subtle effects due to increased power. Yet it is important to note that the 308 ROIs are derived by subdividing the 68 ROI atlas used in the current study into equally sized 'patches' and thus still respects the anatomy of the brain in the same way. Therefore, it is highly unlikely that this would explain our non-replication of previous findings.

One potential issue with these metrics is that these similarity measures only investigate graph properties which only partially describe the whole network (Schieber et al., 2017). By using correlational measures of 'replicability' we only consider edge-weightings, rather than the structure of the network, hence why we also included comparisons of network strength to begin to investigate this in terms of network topology. We could have investigated additional metrics which characterize network topology (i.e. global efficiency) however, due to the fact that the SC networks do not adhere to typical assumptions of networks (edges representing definitive real connections) we utilised strength as a simpler metric which makes less assumptions about the underlying neurophysiology of the network. Thus, we have taken the assumption that SC represents a graph of higher-order inter-relationships between morphometry and not necessarily 'connectivity'.

Conclusion

We have demonstrated that, when we generate the MSN based on a reduced/limited number of MR features, we produce correlation matrices which are highly similar to those generated with multi-modal imaging. However, the networks generated are differentially, topologically organised based on the number of features. We also find that, regardless of number of features, these networks have limited predictive validity of generalised cognitive ability scores, although this may be specific to the current age range under study. Overall, our study recommends that, in situations where multi-modal imaging is not available or clinically/developmentally inappropriate, T1w-restricted MSN construction may give a useful estimate of the MSN,

629 however between model comparisons should be aware of potentially methodologically-driven
630 changes to network topology.

Acknowledgements

Data were provided [in part] by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

Funding: Professor Amanda Wood is supported by a European Research Council Consolidator Fellowship [682734]. Daniel J. King is supported by a studentship awarded by the School of Life and Health Sciences, Aston University, Birmingham UK.

640 References

641

642 Akshoomoff, N., Beaumont, J. L., Bauer, P. J., Dikmen, S. S., Gershon, R. C., Mungas, D., . . . Heaton, R.
643 K. (2013). VIII. NIH Toolbox Cognition Battery (CB): composite scores of crystallized, fluid, and
644 overall cognition. *Monogr Soc Res Child Dev, 78*(4), 119-132. doi: 10.1111/mono.12038
645 Alexander-Bloch, A., Giedd, J. N., & Bullmore, E. (2013). Imaging structural co-variance between
646 human brain regions. *Nat Rev Neurosci, 14*(5), 322-336. doi: 10.1038/nrn3465
647 Alexander-Bloch, A., Raznahan, A., Bullmore, E., & Giedd, J. (2013). The convergence of maturational
648 change and structural covariance in human cortical networks. *J Neurosci, 33*(7), 2889-2899.
649 doi: 10.1523/JNEUROSCI.3554-12.2013
650 Andersson, J. L. R., Skare, S., & Ashburner, J. (2003). How to correct susceptibility distortions in spin-
651 echo echo-planar images: application to diffusion tensor imaging. *Neuroimage, 20*(2), 870-
652 888. doi: 10.1016/S1053-8119(03)00336-7
653 Andersson, J. L. R., & Sotiropoulos, S. N. (2015). Non-parametric representation and prediction of
654 single- and multi-shell diffusion-weighted MRI data using Gaussian processes. *Neuroimage,*
655 *122*, 166-176. doi: 10.1016/j.neuroimage.2015.07.067
656 Andersson, J. L. R., & Sotiropoulos, S. N. (2016). An integrated approach to correction for off-
657 resonance effects and subject movement in diffusion MR imaging. *Neuroimage, 125*, 1063-
658 1078. doi: 10.1016/j.neuroimage.2015.10.019
659 Balardin, J. B., Comfort, W. E., Daly, E., Murphy, C., Andrews, D., Murphy, D. G., . . . Sato, J. R. (2015).
660 Decreased centrality of cortical volume covariance networks in autism spectrum disorders. *J*
661 *Psychiatr Res, 69*, 142-149. doi: 10.1016/j.jpsychires.2015.08.003
662 Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., . . . Consortium,
663 W.-M. H. (2013). Function in the human connectome: Task-fMRI and individual differences in
664 behavior. *Neuroimage, 80*, 169-189. doi: 10.1016/j.neuroimage.2013.05.033
665 Bastien, P., Vinzi, V. E., & Tenenhaus, M. (2005). PLS generalised linear regression. *Computational*
666 *Statistics & Data Analysis, 48*(1), 17-46. doi: 10.1016/j.csda.2004.02.005
667 Batalle, D., Edwards, A. D., & O'Muircheartaigh, J. (2018). Annual Research Review: Not just a small
668 adult brain: understanding later neurodevelopment through imaging the neonatal brain. *J*
669 *Child Psychol Psychiatry, 59*(4), 350-371. doi: 10.1111/jcpp.12838
670 Bertrand, F., & Maumy-Bertrand, M. (2018). plsRglm: Partial least squares linear and generalized linear
671 regression for processing incomplete datasets by cross-validation and bootstrap techniques
672 with R. *Arxiv*. doi: arXiv:1810.01005.

673 Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and
674 functional systems. *Nat Rev Neurosci, 10*(3), 186-198. doi: 10.1038/nrn2575
675 Chen, T., Kendrick, K. M., Wang, J., Wu, M., Li, K., Huang, X., . . . Gong, Q. (2017). Anomalous single-
676 subject based morphological cortical networks in drug-naive, first-episode major depressive
677 disorder. *Hum Brain Mapp*. doi: 10.1002/hbm.23534
678 Consonni, V., Ballabio, D., & Todeschini, R. (2010). Evaluation of model predictive ability by external
679 validation techniques. *Journal of Chemometrics, 24*(3-4), 194-201. doi: 10.1002/cem.1290
680 Desikan, R. S., Segonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., . . . Killiany, R. J. (2006).
681 An automated labeling system for subdividing the human cerebral cortex on MRI scans into
682 gyral based regions of interest. *Neuroimage, 31*(3), 968-980. doi:
683 10.1016/j.neuroimage.2006.01.021
684 Evans, A. C. (2013). Networks of anatomical covariance. *Neuroimage, 80*, 489-504. doi:
685 10.1016/j.neuroimage.2013.05.054
686 Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., . . . Dale, A. M. (2002). Whole
687 brain segmentation: automated labeling of neuroanatomical structures in the human brain.
688 *Neuron, 33*(3), 341-355.
689 Fornito, A., Zalesky, A., & Breakspear, M. (2013). Graph analysis of the human connectome: promise,
690 progress, and pitfalls. *Neuroimage, 80*, 426-444. doi: 10.1016/j.neuroimage.2013.04.087
691 Fornito, A., Zalesky, A., & Bullmore, E. (2016). *Fundamentals of Brain Network Analysis*: Academic
692 Press.
693 Galdi, P., Blesa, M., Sullivan, G., Lamb, G. J., Stoye, D. Q., Quigley, A. J., . . . Boardman, J. P. (2018).
694 Neonatal Morphometric Similarity Networks Predict Atypical Brain Development Associated
695 with Preterm Birth. *10842*, 47-57. doi: 10.1007/978-3-030-00755-3_6
696 Garcia-Ramos, C., Bobholz, S., Dabbs, K., Hermann, B., Joutsa, J., Rinne, J. O., . . . Group, T. S. (2017).
697 Brain structure and organization five decades after childhood onset epilepsy. *Hum Brain*
698 *Mapp*. doi: 10.1002/hbm.23593
699 Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., . . .
700 Consortium, W.-M. H. (2013). The minimal preprocessing pipelines for the Human
701 Connectome Project. *Neuroimage, 80*, 105-124. doi: 10.1016/j.neuroimage.2013.04.127
702 Glasser, M. F., & Van Essen, D. C. (2011). Mapping Human Cortical Areas In Vivo Based on Myelin
703 Content as Revealed by T1- and T2-Weighted MRI. *Journal of Neuroscience, 31*(32), 11597-
704 11616. doi: 10.1523/Jneurosci.2180-11.2011

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., . . . Thompson, P. M. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proc Natl Acad Sci U S A, 101*(21), 8174-8179. doi: DOI 10.1073/pnas.0402680101

Goulas, A., Uylings, H. B., & Hilgetag, C. C. (2017). Principles of ipsilateral and contralateral cortico-cortical connectivity in the mouse. *Brain Struct Funct, 222*(3), 1281-1295. doi: 10.1007/s00429-016-1277-y

Hahn, A., Lanzenberger, R., & Kasper, S. (2019). Making Sense of Connectivity. *International Journal of Neuropsychopharmacology, 22*(3), 194-207. doi: 10.1093/ijnp/pyy100

Heaton, R. K., Akshoomoff, N., Tulsky, D., Mungas, D., Weintraub, S., Dikmen, S., . . . Gershon, R. (2014). Reliability and Validity of Composite Scores from the NIH Toolbox Cognition Battery in Adults. *Journal of the International Neuropsychological Society, 20*(6), 588-598. doi: 10.1017/S1355617714000241

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage, 17*(2), 825-841. doi: 10.1006/nimg.2002.1132

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). Fsl. *Neuroimage, 62*(2), 782-790. doi: 10.1016/j.neuroimage.2011.09.015

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The Unity and Diversity of Executive Functions: A Systematic Review and Re-Analysis of Latent Variable Studies. *Psychol Bull, 144*(11), 1147-1185. doi: 10.1037/bul0000160

Kim, H. J., Shin, J. H., Han, C. E., Kim, H. J., Na, D. L., Seo, S. W., . . . Alzheimer's Disease Neuroimaging, I. (2016). Using Individualized Brain Network for Analyzing Structural Covariance of the Cerebral Cortex in Alzheimer's Patients. *Front Neurosci, 10*(394), 394. doi: 10.3389/fnins.2016.00394

Kong, X. Z., Liu, Z., Huang, L., Wang, X., Yang, Z., Zhou, G., . . . Liu, J. (2015). Mapping Individual Brain Networks Using Statistical Similarity in Regional Morphology from MRI. *PLoS One, 10*(11), e0141840. doi: 10.1371/journal.pone.0141840

Kong, X. Z., Wang, X., Huang, L. J., Pu, Y., Yang, Z. T., Dang, X. B., . . . Liu, J. (2014). Measuring individual morphological relationship of cortical regions. *Journal of Neuroscience Methods, 237*, 103-107. doi: 10.1016/j.jneumeth.2014.09.003

Krishnan, A., Williams, L. J., McIntosh, A. R., & Abdi, H. (2011). Partial Least Squares (PLS) methods for neuroimaging: A tutorial and review. *Neuroimage, 56*(2), 455-475. doi: 10.1016/j.neuroimage.2010.07.034

738 Kruschwitz, J. D., Waller, L., Daedelow, L. S., Walter, H., & Veer, I. M. (2018). General, crystallized and
739 fluid intelligence are not associated with functional global network efficiency: A replication
740 study with the human connectome project 1200 data set. *Neuroimage, 171*, 323-331. doi:
741 10.1016/j.neuroimage.2018.01.018
742 Lerman-Sinkoff, D. B., Sui, J., Rachakonda, S., Kandala, S., Calhoun, V. D., & Barch, D. M. (2017).
743 Multimodal neural correlates of cognitive control in the Human Connectome Project.
744 *Neuroimage, 163*, 41-54. doi: 10.1016/j.neuroimage.2017.08.081
745 Li, W., Yang, C., Shi, F., Wu, S., Wang, Q., Nie, Y., & Zhang, X. (2017). Construction of Individual
746 Morphological Brain Networks with Multiple Morphometric Features. *Front Neuroanat, 11*,
747 34. doi: 10.3389/fnana.2017.00034
748 Mak, E., Colloby, S. J., Thomas, A., & O'Brien, J. T. (2016). The segregated connectome of late-life
749 depression: a combined cortical thickness and structural covariance analysis. *Neurobiol Aging,*
750 *48*, 212-221. doi: 10.1016/j.neurobiolaging.2016.08.013
751 Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer*
752 *research, 27*(2 Part 1), 209-220.
753 Marcus, D. S., Harwell, J., Olsen, T., Hodge, M., Glasser, M. F., Prior, F., . . . Van Essen, D. C. (2011).
754 Informatics and data mining tools and strategies for the human connectome project. *Front*
755 *Neuroinform, 5*, 4. doi: 10.3389/fninf.2011.00004
756 Mechelli, A., Friston, K. J., Frackowiak, R. S., & Price, C. J. (2005). Structural covariance in the human
757 cortex. *J Neurosci, 25*(36), 8303-8310. doi: 10.1523/JNEUROSCI.0357-05.2005
758 Morgan, S. E., Seidlitz, J., Whitaker, K., Romero-Garcia, R., Clifton, N. E., Scarpazza, C., . . . Bullmore, E.
759 T. (2018). Cortical patterning of abnormal morphometric similarity in psychosis is associated
760 with brain expression of schizophrenia related genes. *bioRxiv*, 501494. doi: 10.1101/501494
761 Nomi, J. S., Vij, S. G., Dajani, D. R., Steimke, R., Damaraju, E., Rachakonda, S., . . . Uddin, L. Q. (2017).
762 Chronnectomic patterns and neural flexibility underlie executive function. *Neuroimage, 147*,
763 861-871. doi: 10.1016/j.neuroimage.2016.10.026
764 Pagani, M., Bifone, A., & Gozzi, A. (2016). Structural covariance networks in the mouse brain.
765 *Neuroimage, 129*, 55-63. doi: 10.1016/j.neuroimage.2016.01.025
766 Palaniyappan, L., Park, B., Balain, V., Dangi, R., & Liddle, P. (2015). Abnormalities in structural
767 covariance of cortical gyrification in schizophrenia. *Brain Struct Funct, 220*(4), 2059-2071. doi:
768 10.1007/s00429-014-0772-2
769 Pereira, J. B., Aarsland, D., Ginestet, C. E., Lebedev, A. V., Wahlund, L. O., Simmons, A., . . . Westman,
770 E. (2015). Aberrant cerebral network topology and mild cognitive impairment in early
771 Parkinson's disease. *Hum Brain Mapp, 36*(8), 2980-2995. doi: 10.1002/hbm.22822

Pereira, J. B., Mijalkov, M., Kakaei, E., Mecocci, P., Vellas, B., Tsolaki, M., . . . AddNeuroMed consortium, f. t. A. s. D. N. I. (2016). Disrupted Network Topology in Patients with Stable and Progressive Mild Cognitive Impairment and Alzheimer's Disease. *Cereb Cortex, 26*(8), 3476-3493. doi: 10.1093/cercor/bhw128

R Core Team. (2016). R: A language and environment for statistical computing (Version 3.3.2). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Raamana, P. R., Weiner, M. W., Wang, L., Beg, M. F., & Alzheimer's Disease Neuroimaging, I. (2015). Thickness network features for prognostic applications in dementia. *Neurobiol Aging, 36 Suppl 1*(S1), S91-S102. doi: 10.1016/j.neurobiolaging.2014.05.040

Rosen, A. F. G., Roalf, D. R., Ruparel, K., Blake, J., Seelaus, K., Villa, L. P., . . . Satterthwaite, T. D. (2018). Quantitative assessment of structural image quality. *Neuroimage, 169*, 407-418. doi: 10.1016/j.neuroimage.2017.12.059

Schieber, T. A., Carpi, L., Diaz-Guilera, A., Pardalos, P. M., Masoller, C., & Ravetti, M. G. (2017). Quantification of network structural dissimilarities. *Nat Commun, 8*, 13928. doi: 10.1038/ncomms13928

Seidlitz, J., Nadig, A., Liu, S., Bethlehem, R. A. I., Vértes, P. E., Morgan, S. E., . . . Raznahan, A. (2019). Transcriptomic and Cellular Decoding of Regional Brain Vulnerability to Neurodevelopmental Disorders. *bioRxiv*. doi: 10.1101/573279

Seidlitz, J., Vasa, F., Shinn, M., Romero-Garcia, R., Whitaker, K. J., Vertes, P. E., . . . Bullmore, E. T. (2018). Morphometric Similarity Networks Detect Microscale Cortical Organization and Predict Inter-Individual Cognitive Variation. *Neuron, 97*(1), 231-247 e237. doi: 10.1016/j.neuron.2017.11.039

Sone, D., Matsuda, H., Ota, M., Maikusa, N., Kimura, Y., Sumida, K., . . . Sato, N. (2016). Graph Theoretical Analysis of Structural Neuroimaging in Temporal Lobe Epilepsy with and without Psychosis. *PLoS One, 11*(7), e0158728. doi: 10.1371/journal.pone.0158728

Sowell, E. R., Thompson, P. M., Leonard, C. M., Welcome, S. E., Kan, E., & Toga, A. W. (2004). Longitudinal mapping of cortical thickness and brain growth in normal children. *Journal of Neuroscience, 24*(38), 8223-8231. doi: 10.1523/Jneurosci.1798-04.2004

Stone, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society Series B-Statistical Methodology, 36*(2), 111-147.

Tewarie, P., Steenwijk, M. D., Tijms, B. M., Daams, M., Balk, L. J., Stam, C. J., . . . Hillebrand, A. (2014). Disruption of Structural and Functional Networks in Long-Standing Multiple Sclerosis. *Hum Brain Mapp, 35*(12), 5946-5961. doi: 10.1002/hbm.22596

Tijms, B. M., Series, P., Willshaw, D. J., & Lawrie, S. M. (2012). Similarity-based extraction of individual networks from gray matter MRI scans. *Cereb Cortex, 22*(7), 1530-1541. doi: 10.1093/cercor/bhr221

Tijms, B. M., Sprooten, E., Job, D., Johnstone, E. C., Owens, D. G., Willshaw, D., . . . Lawrie, S. M. (2015). Grey matter networks in people at increased familial risk for schizophrenia. *Schizophr Res, 168*(1-2), 1-8. doi: 10.1016/j.schres.2015.08.025

Torgo, L., Branco, P., Ribeiro, R. P., & Pfahringer, B. (2015). Resampling strategies for regression. *Expert Systems, 32*(3), 465-476. doi: 10.1111/exsy.12081

Tulsky, D. S., Carlozzi, N. E., Chevalier, N., Espy, K. A., Beaumont, J. L., & Mungas, D. (2013). V. NIH Toolbox Cognition Battery (CB): measuring working memory. *Monogr Soc Res Child Dev, 78*(4), 70-87. doi: 10.1111/mono.12035

Wechler, D. (1999). *Manual for the Wechsler abbreviated intelligence scale (WASI)*. San Antonio, TX: The Psychological Corporation.

Wechler, D. (2008). *Wechsler adult intelligence scale–Fourth Edition (WAIS–IV)*. San Antonio, TX: NCS Pearson.

Wechsler, D. (Ed.). (2003). *Wechsler Intelligence Scale for Children: Fourth Edition*. San Antonio, TX, : The Psychological Corporation

Wei, Y. B., Scholtens, L. H., Turk, E., & van den Heuvel, M. P. (2019). Multiscale examination of cytoarchitectonic similarity and human brain connectivity. *Network Neuroscience, 3*(1), 124-137. doi: 10.1162/netn_a_00057

Westwater, M. L., Seidlitz, J., Diederen, K. M., Fischer, S., & Thompson, J. C. (2017). Alterations in cortical thickness and structural connectivity are associated with symptom severity in bulimia nervosa. *bioRxiv*. doi: 10.1101/127910

Yasuda, C. L., Chen, Z., Beltramini, G. C., Coan, A. C., Morita, M. E., Kubota, B., . . . Gross, D. W. (2015). Aberrant topological patterns of brain structural network in temporal lobe epilepsy. *Epilepsia, 56*(12), 1992-2002. doi: 10.1111/epi.13225

Yu, K. X., Wang, X. T., Li, Q. L., Zhang, X. H., Li, X. W., Li, S. Y., & Initi, A. s. D. N. (2018). Individual Morphological Brain Network Construction Based on Multivariate Euclidean Distances Between Brain Regions. *Front Hum Neurosci, 12*. doi: Artn 204 10.3389/Fnhum.2018.00204

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): a method of assessing executive function in children. *Nat Protoc, 1*(1), 297-301. doi: 10.1038/nprot.2006.46

Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., Conway, K. P., . . . Weintraub, S. (2014). NIH Toolbox Cognition Battery (CB): validation of executive function measures in adults. *J Int Neuropsychol Soc, 20*(6), 620-629. doi: 10.1017/S1355617714000472

Zheng, W., Eilamstock, T., Wu, T., Spagna, A., Chen, C., Hu, B., & Fan, J. (2019). Multi-feature based network revealing the structural abnormalities in autism spectrum disorder. *IEEE Transactions on Affective Computing*, 1-1. doi: 10.1109/taffc.2018.2890597

Zheng, W., Yao, Z., Xie, Y., Fan, J., & Hu, B. (2018). Identification of Alzheimer's Disease and Mild Cognitive Impairment Using Networks Constructed Based on Multiple Morphological Brain Features. *Biol Psychiatry Cogn Neurosci Neuroimaging, 3*(10), 887-897. doi: 10.1016/j.bpsc.2018.06.004

Supplementary Materials

Dimension reduction of Executive functioning task performance

A principal component analysis (PCA, using the 'prcomp' function in the base R 'stats' package (R Core Team, 2016)) was used to find a common EF component across all three EF measures.

Data reduction using the PCA was done for two main reasons; a) to reduce dimensionality, and the number of multiple predictor models being built and b) to ensure that we were predicting (a latent variable of) executive functioning ability, rather than ability linked to task-specific performance.

The PCA suggested a three-component solution, however only the first component had an eigen-value > 1 (eigenvalue=1.607) and so only this component was retained. This component explained ~54% variance across our measures. All three measures; list-sort, card-sort and flanker, positively loaded onto this component (rotated sums of squares loading = .362, .673 and .646 respectively).