

RUNNING HEAD: MSN-CONSTRUCTION WITH RESTRICTED MR ACQUISITIONS



Clinically-Feasible Brain Morphometric Similarity Network Construction Approaches with
Restricted Magnetic Resonance Imaging Acquisitions

Daniel J. King¹ & 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 organisation of the cortex as a biologically meaningful set of similarities between anatomical features at the macro- and micro- structural level, derived from multiple structural MRI (sMRI) sequences. These networks are clinically relevant, predicting 40% variance in IQ. However, the sequences required (T1w, T2w, DWI) to produce these networks are longer acquisitions, less feasible in some populations. Thus, estimating MSNs using features from T1w sMRI is attractive to clinical and developmental neuroscience. We studied whether reduced-feature approaches approximate the original MSN model as a potential tool to investigate brain structure. In a large, homogenous dataset of healthy young-adults (from the Human Connectome Project), 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, however the nodal-topology of the networks differed. These networks had limited predictive validity of generalised cognitive ability. Overall, when multi-modal imaging is not available or appropriate, T1w-restricted MSN construction is feasible, provides an appropriate estimate of the MSN, and could be a useful approach to examine outcomes in future studies.

Keywords: Morphometric similarity networks, Structural MRI, Morphology, Connectome, Cognition

43 **Author Summary**

44 We can estimate the higher-order organisation of cortical grey matter as a connectome using
45 structural MRI. However, this methodology, termed morphometric similarity, requires multiple
46 ‘advanced’ neuroimaging protocols that are unsuitable, unavailable or intolerable to certain
47 populations, including children and some clinical groups. In a large, homogenous dataset of
48 healthy young-adults, we estimated these connectomes using three different feature sets, each
49 extracted from fewer MRI sequences. Even when produced using only T1-weighted structural
50 MRI scans, these connectomes were broadly similar to those produced with more complex or
51 numerous MRI sequences. We did not replicate previous findings linking variation in the MSN
52 to individual-differences in cognitive abilities. We highlight potential reasons for this,
53 including the developmental stage of the young adult imaging cohort in which our hypotheses
54 were tested, and conclude that this study provides putative evidence that, in those populations
55 where advanced imaging is not plausible, MSNs generated from T1-weighted structural MRIs
56 are a promising alternative.

1. Introduction

Cortical grey-matter structural covariance networks (SCNs) model the degree to which the morphology of cortical regions (e.g. cortical thickness or volume) statistically co-varies across all possible pairs of regions (Alexander-Bloch, Giedd, & Bullmore, 2013; Evans, 2013). These whole-brain, network approaches to morphometry, within a graph-theoretic framework, allow us to investigate additional information beyond univariate approaches to neuroanatomy (Pagani, Bifone, & Gozzi, 2016). Disruption to the SCN has been explored in a range of neurological and neuropsychiatric conditions. However, this methodology necessarily generates group-level networks indexing population-level covariance in neuroanatomy (Alexander-Bloch, Raznahan, Bullmore, & Giedd, 2013). This limits the ability to quantify system-level abnormalities within individual patients, which could benefit stratified diagnosis and prognosis (Zheng, Yao, Xie, Fan, & Hu, 2018).

An alternative approach to investigate the regional covariance structure between multiple morphometric features at an individual-level, is Morphometric Similarity Networks (MSNs; Seidlitz et al. (2018)). This approach estimates meso-scale organisation of the cortex as a biologically meaningful set of similarities between anatomical features at both the macro- and micro- structural level (Morgan et al., 2018), including meso-scale morphometry, tissue diffusion properties and myelination indices. 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 regional gene co-expression in human data and axonal tract-tracing in the rhesus macaque (Seidlitz et al., 2018). These findings reflect the fact that cortical regions that are more anatomically similar are likely to be anatomically connected (Goulas, Uylings, & Hilgetag, 2017; Wei, Scholtens, Turk, & van den Heuvel, 2019). The MSN represents a new neuroimaging phenotype, which may provide additional, biologically-relevant information beyond existing network approaches.

MSNs have already been utilised in a small number of clinical studies (characteristics of these studies, including neuroanatomical features extracted, are listed below in Table 1). For example, previous studies identified a robust, replicable pattern of differences in morphometric similarity between patients with psychosis compared to controls (Morgan et al., 2018) and detected dysmaturation of the brain in preterm infants (Galdi et al., 2018). Overall, these findings suggest that MSNs represent a useful and clinically-relevant phenotype.

Multi-modal, high-quality MRI sequences are required for these approaches. These may not be feasible or tolerable for all research settings and/or populations. Existing clinical ('legacy') cohorts often lack 'advanced' imaging. Longer acquisition times for advanced sequences may reduce image quality, especially in paediatric or clinical applications where non-compliance and movement is more likely as acquisition time increases (Rosen et al., 2018).

Estimating MSNs using a single T1w 3D anatomical MRI, is appealing for clinical and developmental neuroscience (Batalle, Edwards, & O'Muircheartaigh, 2018). Seidlitz et al. (2018) and Li et al. (2017) estimated morphometric similarity in this way and found that network edge weights were similar to 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)). Using only T1w MRI it is possible to identify patterns of morphometric similarity that classify autism spectrum disorder (Zheng et al., 2019), Alzheimer's disease and mild cognitive impairment (Zheng et al., 2018) from controls, albeit with reduced estimation precision, and greater standard deviation of edge-level weights across participants (Seidlitz et al., 2018). This method also identified a spatial pattern of anatomical disruptions associated with regional gene-expression, findings consistent with a 'transcriptional vulnerability model' of neurodevelopmental disorders (Seidlitz et al., 2019). Characteristics of these studies, including neuroanatomical features extracted, are listed below in Table 1. However, existing studies

- 106 provide limited assessment of reliability, consistency with group-networks, biological validity
- 107 and predictive ability.

Table 1. Characteristics of previous studies utilizing Morphometric Similarity Networks to investigate cortical organization

Study	Population	MRI Acquisitions	MSN feature set
Seidlitz et al. (2018)	Typically-Developing Adolescents	Multi-parametric mapping sequence & DWI	FA, MD, MT, GM, SA, CT, GC, MC, CI & FI (Reduced feature MSN; CT, SA, GM, GC, MC)
Morgan et al. (2018)	Schizophrenia or Non-affective psychotic disorders	T1w MRI & DWI	GM, SA, CT, GC, MC, FA, MD
Galdi et al. (2018)	Term & Preterm Neonates	T1w MRI, T2w MRI & DWI	GM, T1/T2 ratio, FA, MD, RD, AD, VISO, IVF, ODI
Seidlitz et al. (2019)	Neurodevelopmental disorders of known genetic origin	T1w MRI	CT, SA, GM, MC, GC
Li et al. (2017)	Health Adults	T1w MRI	Vertices, GM, SA, CT, SD-CT, MC, GC, CI, FI
Zheng et al. (2018)	Mild Cognitive Impairment & Alzheimer's Disease	T1w MRI	CT, SA, GM, LGI, Sulcul Depth, Gyrus Height
Zheng et al. (2019)	Autism Spectrum Disorders	T1w MRI	CT, SA, GM, LGI, Sulcul Depth, Gyrus Height, MC

Note. FA=Fractional Anisotropy, MD=Mean Diffusivity, MT=Magnetization Transfer, GM=Gray Matter Volume, SA=Surface Area, CT=Cortical Thickness, GC=Gaussian Curvature, MC=Mean Curvature, CI=Curvature Index, FI=Folding Index, RD=Radial Diffusivity, AD=Axial Diffusivity, VISO=Isotropic Volume Fraction, IVF=Intracellular Volume Fraction, ODI=Orientation Dispersion Index, Vertices=Number of vertices, SD-CT=Standard Deviation of Cortical Thickness, LGI=Local Gyrification Index

The current study aimed to determine whether reduced-feature approaches approximate the ‘original’ MSN model. We extended previous investigations of reduced-feature MSNs by comparing not only T1w-derived networks, but also additional MSN models, each using fewer metrics from a reduced number of MRI acquisitions.

We predicted that, for each measure of reliability/replicability, performance would be ordered hierarchically with MSNs generated with the greatest number of features outperforming those generated with fewer. We also predicted that between-model comparisons would suggest that the models themselves were highly similar. In line with previous work (Seidlitz et al., 2018), we predicted that there would be an association between cognition and MSN organisation and this would generalise to a novel cognitive-domain, 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 (.7mm isotropic), T2w (.7mm isotropic) 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 neuroanatomical structure were derived from multiple imaging modalities (see Table 2). 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 2. 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 (1.25mm isotropic, $b = 1000$) 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.

We constructed these networks across three different MSN models. 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 ($\text{MSN}_{10\text{-feat.}}$)),

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

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

Model a), hereto referred to as MSN_{10-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 MSNs (one per model) were estimated.

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 $MSN_{10-feat.}$ 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 $MSN_{10-feat.}$ was compared to the $MSN_{8-feat.}$ and then the $MSN_{10-feat.}$ was subsequently compared to the $MSN_{7-feat.}$.

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

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

320 3. Results

321 3.1 Inter-model comparisons

322 3.1.1 Magnitude of morphometric similarity: graph-level strength

323 In terms of the topology of the networks, global graph strength for each model, across densities,
324 can be seen in Figure 1. This plot shows the similar trajectories across densities for all models
325 tested, however the observed average graph strength was different between models, with lower
326 strength being seen in the MSN models with greater features. The effect size of differences
327 (estimated with a paired t-test) between $MSN_{10\text{-feat.}}$ vs $MSN_{8\text{-feat.}}$ and $MSN_{10\text{-feat.}}$ vs $MSN_{7\text{-feat.}}$
328 can be also be seen in Figure 1. Effect sizes (r) were extremely large, especially between
329 $MSN_{10\text{-feat.}}$ vs $MSN_{7\text{-feat.}}$.

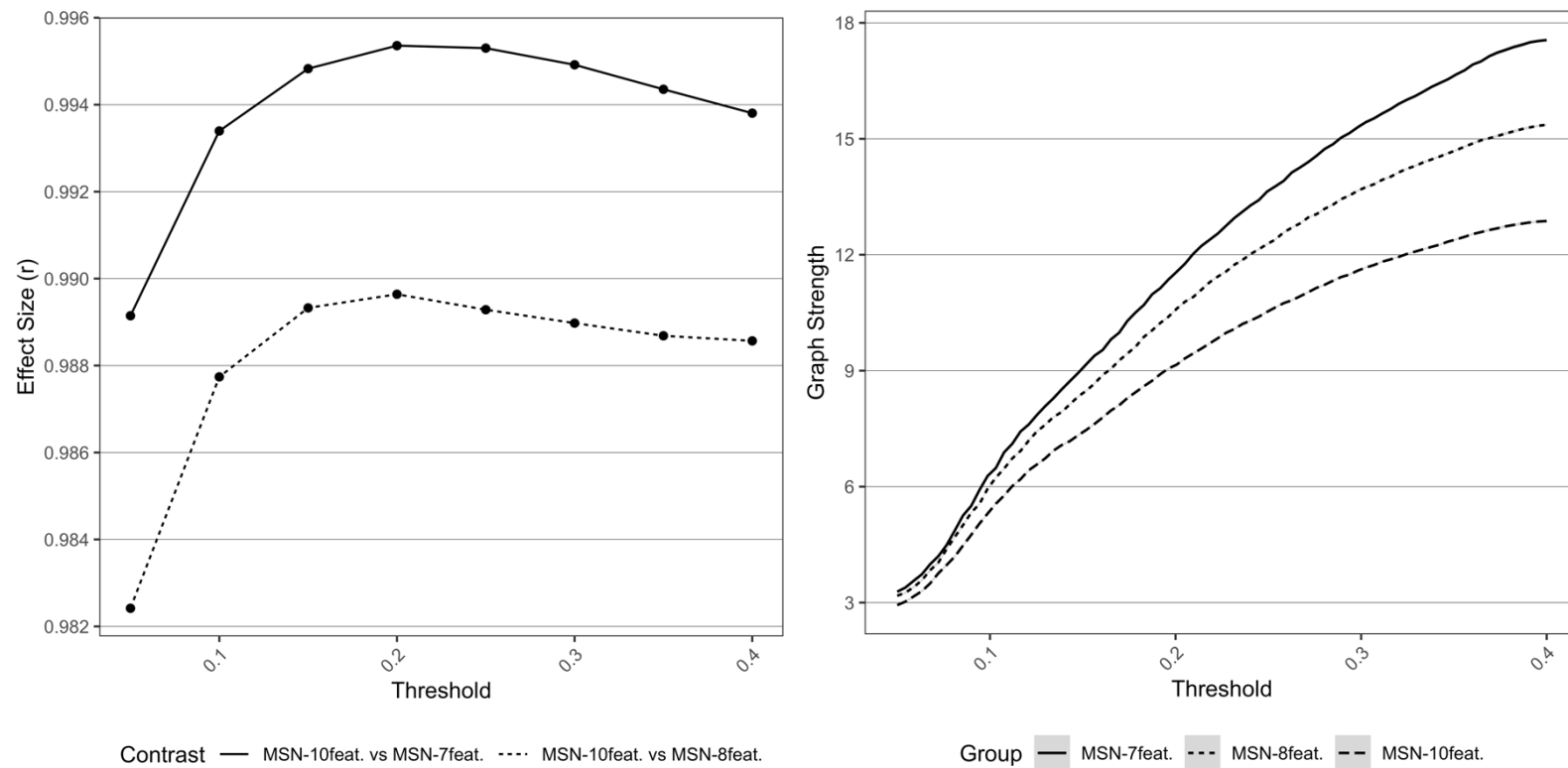


Figure 1 Left: Graph metrics describing average network strength for each MSN model, across all densities. Right: Effect sizes of differences between a) $MSN_{10feat.}$ vs $MSN_{8feat.}$ and b) $MSN_{10feat.}$ vs $MSN_{7feat.}$ for differing graph metrics, across densities.

3.1.2 Edge Weights

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

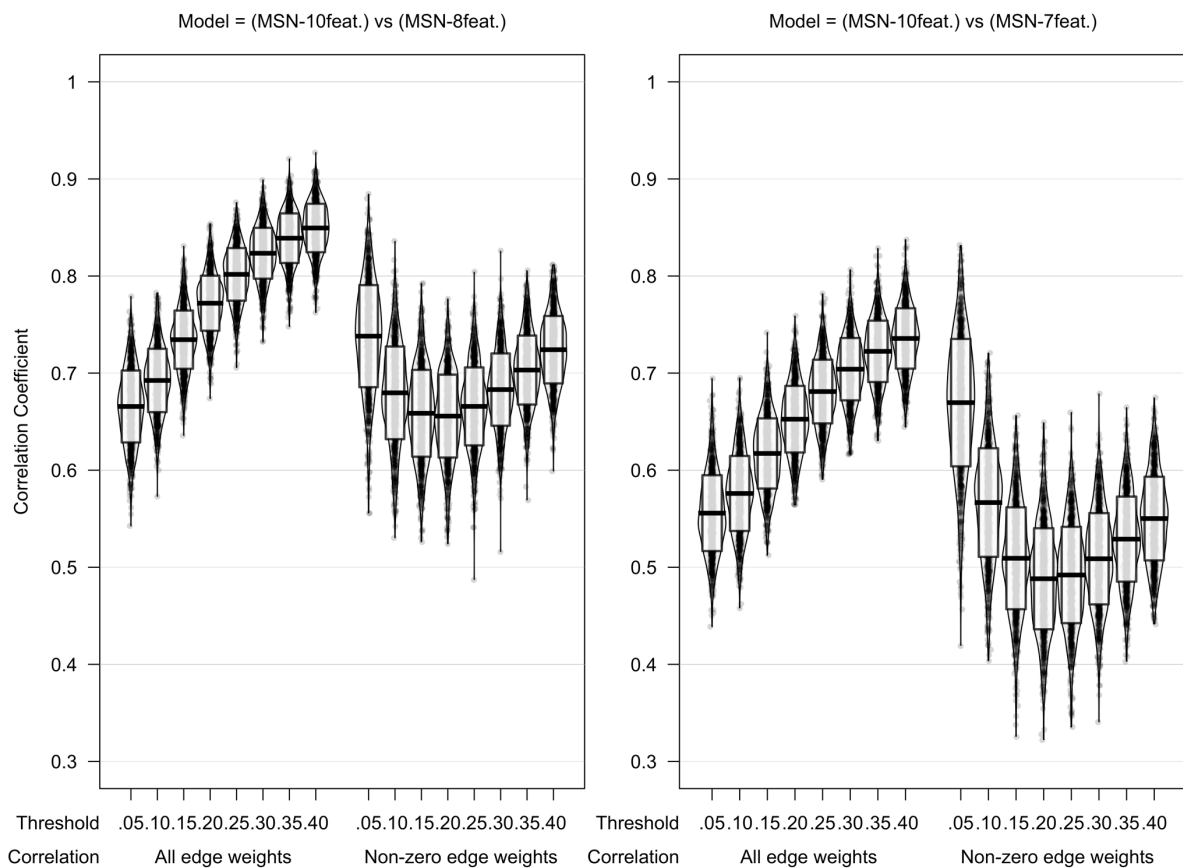


Figure 2 Violin plot of correlation of edgeweights between a) $MSN_{10\text{-feat.}}$ vs $MSN_{8\text{-feat.}}$ and b) $MSN_{10\text{-feat.}}$ vs $MSN_{7\text{-feat.}}$. Midline of the box-plot component of the violin represents the mean of all correlation coefficients, with the box itself representing 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 $r(M \pm SD) = .835 (\pm .028)$; $MSN_{10\text{-feat.}}$ vs $MSN_{7\text{-feat.}}$ Mantel $r(M \pm SD) = .715, (\pm .034)$). 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%, ($\pm 2\%$); $MSN_{10\text{-feat.}}$ vs $MSN_{7\text{-feat.}}$ proportion of replicated edges = 77%, ($\pm 2\%$; Figure 3)).

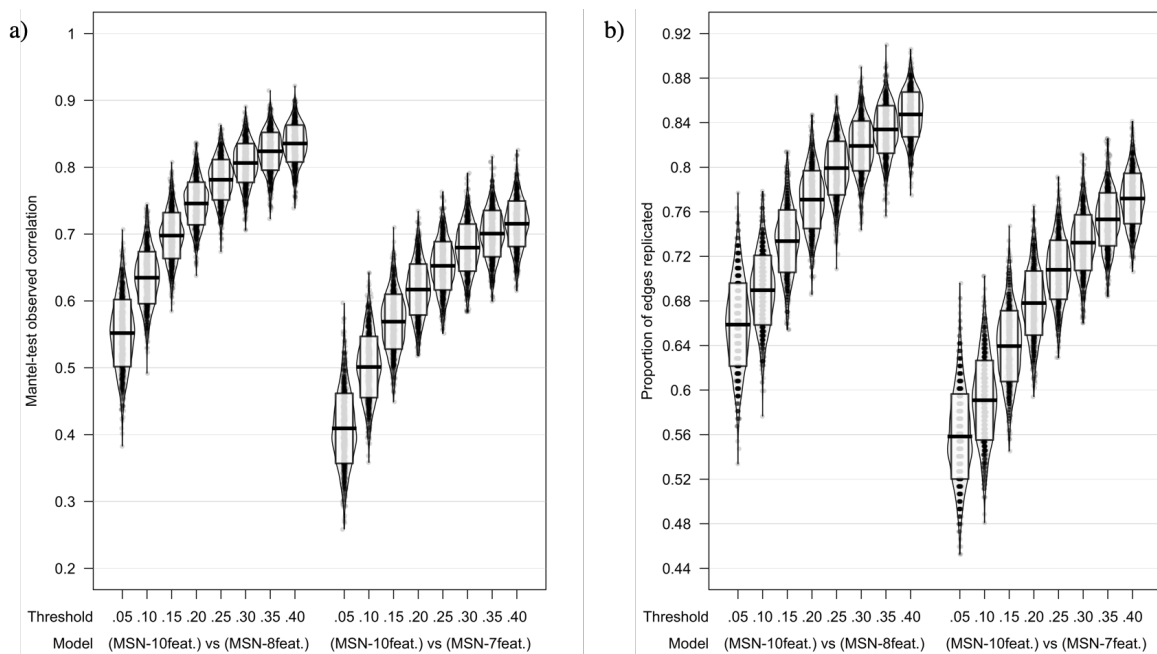


Figure 3 Model comparisons across thresholds using a) Mantel-test correlation coefficient and b) proportion of edges replicated as measures of model similarities. Midline of the box-plot component of the violin represents the mean whilst the box itself representing the SD.

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: $MSN_{10\text{-feat.}} r(M \pm SD) = .902 (\pm .032)$; $MSN_{8\text{-feat.}} r(M \pm SD) = .881 (\pm .040)$, $MSN_{7\text{-feat.}} r(M \pm SD) = .857 (\pm .043)$). 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 % ($\pm 3\%$); $MSN_{8\text{-feat.}}$ proportion of replicated edges = 87% ($\pm 3\%$), $MSN_{7\text{-feat.}}$ proportion of replicated edges = 86% ($\pm 3\%$)). 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. Example correlation matrices are found in Figure 4. 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 5). Interestingly, the $MSN_{8\text{-feat.}}$ model showed greatest correlation between edge weights (At 40% threshold: $MSN_{10\text{-feat.}} r(M \pm SD) = .843 (\pm .032)$; $MSN_{8\text{-feat.}} r(M \pm SD) = .875 (\pm .029)$, $MSN_{7\text{-feat.}} r(M \pm SD) = .850, (\pm .031)$). Similar to the inter-model analyses, correlation peaked at the highest threshold tested (40%) for all models.

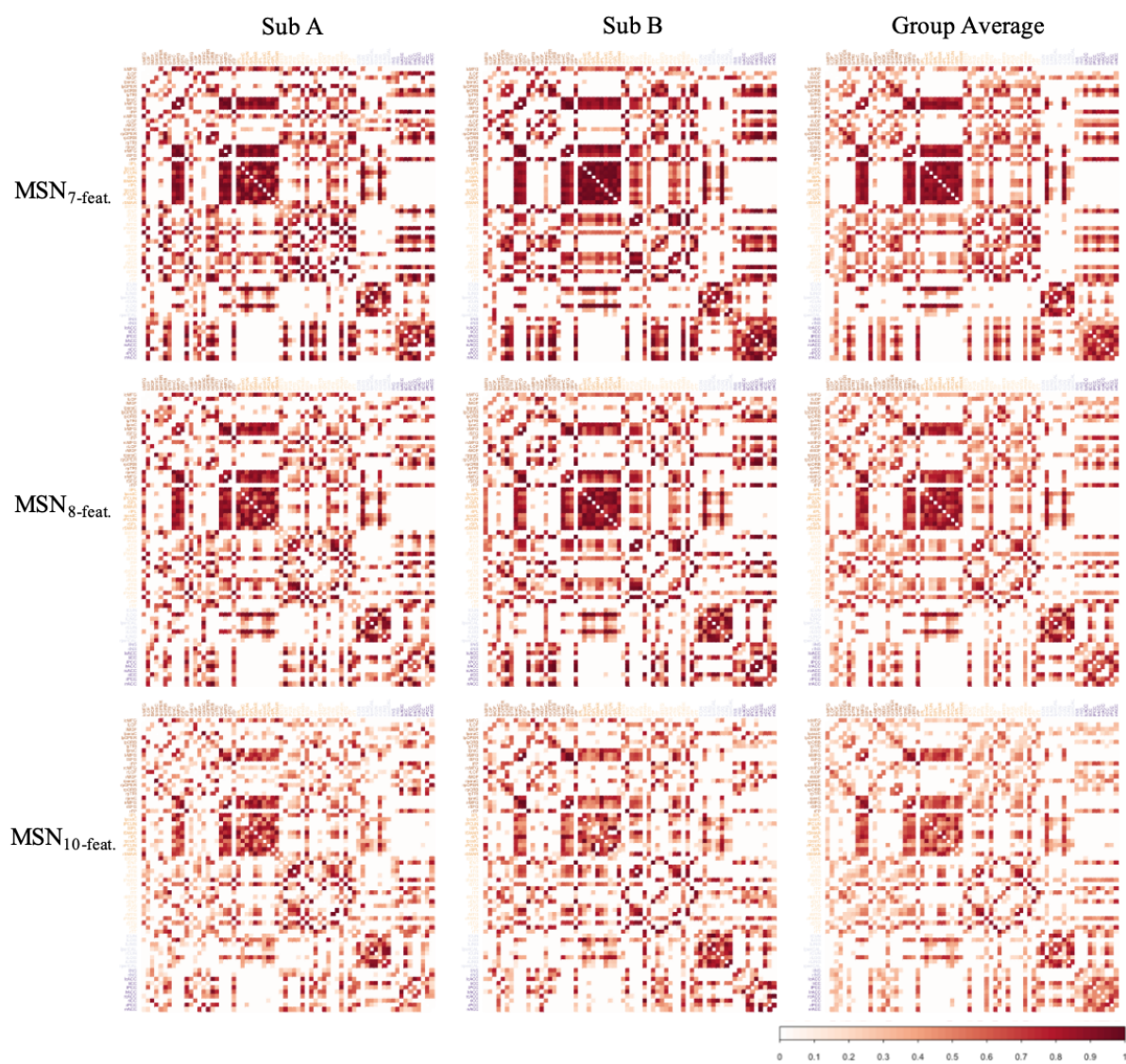
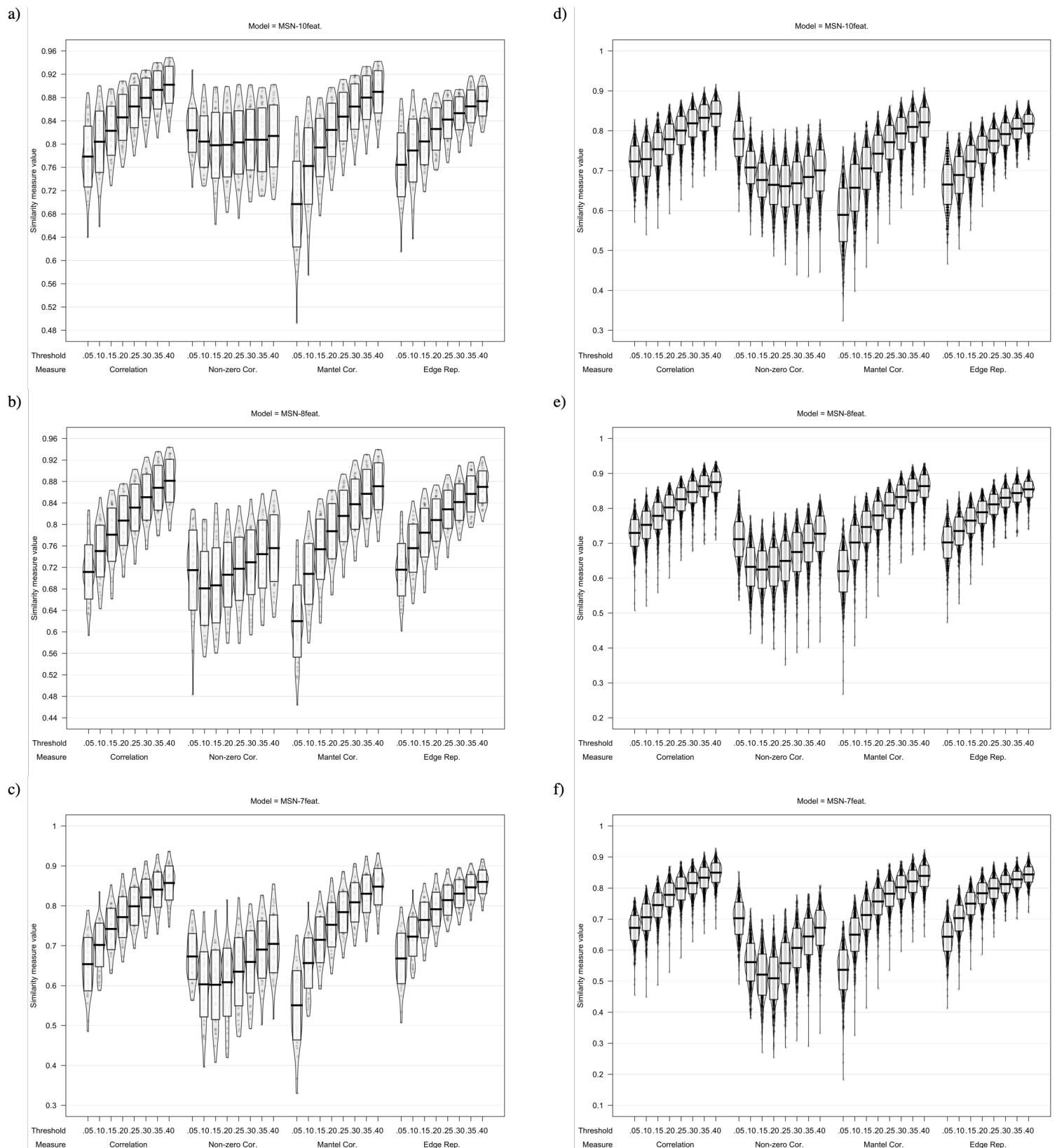


Figure 4. Examples of the correlation matrices generated with each MSN model. Column A and B represent MSNs from two random subjects from the HCP dataset, whilst the final column represents the group average MSN for each MSN model. All correlation matrices visualised here represent the MSN thresholded at a density of 40% and are sorted by lobe assignment as defined by the desikan-killany atlas (Frontal, Parietal, Temporal, Occipital, Insula, Cingulate).



386 Figure 5 Plots showing MSN similarity (across thresholds, with multiple similarity measures) between
 387 a,b,c) individual MSNs generated with test-retest MRI scans and d,e,f) individual-level MSNs and the
 388 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

The current study is the first to formally investigate the potential for generation of MSNs based upon a reduced number of neuroanatomical features, dependant on the complexity of the MRI acquisition sequence. We found that the weighted networks generated from these models are highly congruent, across a number of similarity measures. We expanded previous assessments (e.g. Seidlitz et al., (2018) compared five morphometric features to their full $MSN_{10-feat}$ model) include multiple MSN models. Irrespective of the statistical assessments used here, the between-model similarity was nearly always hierarchical, with greater similarity seen between $MSN_{10-feat}$ and MSN_{8-feat} compared to that between $MSN_{10-feat}$ and MSN_{7-feat} . Weaker similarity was found for sparser networks at a much lower density (i.e. 0.05). When edges were binarised, replication rates remained relatively high, suggesting that the models are sensitive to the specific edges within the network. Each model displayed high-levels of congruence with the group average network, suggesting that these methods index individual differences from a relatively consistent meso-scale-phenotype of brain structure. Li et al. (2017) found high levels

of test-retest reliability of the T1w MSN, and we replicated this with each of the reduced-feature MSNs seemingly had similar reproducibility in terms of test-retest MRI. This is maybe unsurprising given that the HCP scan acquisitions are designed for high, within-subject reproducibility (Van Essen et al., 2013). However, in terms of average nodal strength, a measure of the magnitude of morphometric similarity, significant between-model differences with large effect sizes were found. Specifically, as myelo-architectural features were added to the MSN model (T1w/T2w ratio, FA, MD), the magnitude of morphometric similarity was reduced and regions appeared less similar and more differentiated, hence the lower average nodal strength. This suggests that, despite edge-level congruence, the weighted-topology of individual nodes is different between models. Given that the current, and previous studies (i.e. Seidlitz et al. (2018), utilise these nodal-level metrics to predict functional outcomes, this difference in topology means that between-study comparisons of these predictive models, when generated from different feature sets, is not justified. These differences were only investigated with a single metric of network topology. Additional metrics (i.e. efficiency/clustering) were less attractive candidates given that the MSNs do not adhere to typical assumptions of networks (such as edges representing definitive real connections). Hence, by using strength as a simpler metric we made fewer assumptions about the underlying neurophysiology of the network.

Interestingly, none of the models tested here showed perfect or even near-perfect concordance across the statistical measures employed. This may be because these models were generated with fewer features, rather than being specific to the modality of feature being dropped. Future research could examine this systematically by generating MSNs with 10, 8 and 7 randomly selected features, irrespective of modality of MRI sequence used to derive said feature. .

Each model may also index a different network phenotype. Each MRI acquisition assesses different neuroanatomical features. (e.g. T1w/T2w ratio and DWI specifically index myelo-architecture of regions whilst the T1w MRI assesses macroscopic morphology of the cortex).

Thus, when features are systematically removed in this way by removing an imaging modality, the resultant MSN is likely to represent a different imaging phenotype. Therefore, although each MSN may be substantially statistically similar, in cases where multimodal imaging is available or is feasible, the most appropriate MSN model may depend on the research question to be answered or the clinical population/pathological mechanism under investigation.

The main benefit of reduced MR-acquisition approaches (specifically the MSN_{7-feat.} model) is the applicability to those populations where multiple MR sequence acquisition is more challenging or difficult (Batalle et al., 2018). 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 at a tolerable level for children to ensure compliance with the full MRI protocol and reduce the likelihood of movement across each of the scans. Another potential application of these models is in legacy, clinical imaging. Routine clinical imaging generates vast quantities of MR data that are not typically assessed using quantitative methods. Although expert reporting yields the information needed to inform acute diagnostic requirements, the ability of those scans to predict later outcomes is largely untested or unsatisfactory. The majority of hospital settings will not have access to high resolution, so called ‘advanced’ MRI sequences, or the expertise to analyse such data quantitatively. Overall, this therefore positions MSNs as a useful in-vivo imaging phenotype for studying both clinical and developmental populations, with the T1w-only model holding greatest potential to become a widely adopted, automated approach in clinical neurosciences.

A common assumption is that the topological organization of the brain networks (across multiple MR modalities), as quantified within a graph theoretic framework, captures physiological information relevant to individual differences in cognitive functioning (Bullmore & Sporns, 2009; Fornito, Zalesky, & Breakspear, 2013; Hahn, Lanzenberger, & Kasper, 2019).

We assessed the brain-behaviour relationships using the MSN models by comparing the predictive validity of the three MSN models in relation to general intelligence, with previous research showing that the organization of the MSN network (modelled similarly to the MSN_{10-feat.}) predicted ~40% variance in WASI IQ (Seidlitz et al., 2018). The current study did not find a relationship with either a measure of general cognitive functioning or with a previously untested domain, executive functioning. Using 9- fold cross-validation, no model (at any density) recommended retention of any PLS components. An important strength of the current study is our quantitative approach to cross-validation to confirm the retained number of components; previous studies retained either the single or two components that explained 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, explaining our inability to replicate findings of Seidlitz et al. (2018).

A number of other factors may explain our results. Developmental differences between our sample (healthy young adult population between the 3rd and 4th decades of life) and that of Seidlitz et al. (2018) may account for the lack of predictive ability in our work. Adolescence is a peak period for brain maturation (Gogtay et al., 2004; Sowell et al., 2004), including establishment of cognitive skill such as executive functions, as studied here. This is reflected in data from the NIH-toolbox, in which the total cognition composite highlights shows a greater magnitude of age effects in childhood compared to adulthood (Akshoomoff et al., 2013; Heaton et al., 2014). 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 subtle age-related effects) and so there is likely less between-individual

variance in the MSN. 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)) supports this contention. Therefore, the limited variance in the MSN within the age group we studied may mean that there is not enough variance to relate to cognitive functioning.

Whilst we were not able to replicate previous brain-behaviour relationships with the MSN, given the evidence above, there is an open hypothesis as to whether the MSN is a valuable tool in independent populations. 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).

Our analytic approach may also explain the different findings. Seidlitz et al. (2018) used the Weschler Abbreviated Scale of Intelligence (WASI; (Wechler, 1999)), whereas we used the NIH Toolbox Cognition composite scores (Heaton et al., 2014). 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), and therefore core data elements will enable future studies to clarify whether the application of MSNs is relevant to specific measures. We also calculated the MSN

at a much lower spatial scale (68 ROIs) compared to 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. Future research should replicate the current findings in independent datasets, across different atlases and at different spatial resolutions.

Conclusion

We have demonstrated that, when we generate the MSN based on a reduced/limited number of MR features, we produce correlation matrices that are highly similar to those generated with multi-modal imaging. However, the nodal-level topology differed based on the number of features. In contradiction of previous research, we found 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, this study provides tentative evidence that, in situations where multi-modal imaging is not available or clinically/developmentally inappropriate, T1w-restricted MSN construction may be an appropriate proxy for multi-modal MSNs. However, nodal-level topology is likely to be biased based upon the neuroanatomical feature sets utilised to construct these networks, which will limit generalizability across studies.

References

- Akshoomoff, N., Beaumont, J. L., Bauer, P. J., Dikmen, S. S., Gershon, R. C., Mungas, D., . . . Heaton, R. K. (2013). VIII. NIH Toolbox Cognition Battery (CB): composite scores of crystallized, fluid, and overall cognition. *Monogr Soc Res Child Dev*, 78(4), 119-132. doi:10.1111/mono.12038
- Alexander-Bloch, A., Giedd, J. N., & Bullmore, E. (2013). Imaging structural co-variance between human brain regions. *Nat Rev Neurosci*, 14(5), 322-336. doi:10.1038/nrn3465
- Alexander-Bloch, A., Raznahan, A., Bullmore, E., & Giedd, J. (2013). The convergence of maturational change and structural covariance in human cortical networks. *J Neurosci*, 33(7), 2889-2899. doi:10.1523/JNEUROSCI.3554-12.2013
- Andersson, J. L. R., Skare, S., & Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage*, 20(2), 870-888. doi:10.1016/S1053-8119(03)00336-7
- Andersson, J. L. R., & Sotiropoulos, S. N. (2015). Non-parametric representation and prediction of single- and multi-shell diffusion-weighted MRI data using Gaussian processes. *Neuroimage*, 122, 166-176. doi:10.1016/j.neuroimage.2015.07.067
- Andersson, J. L. R., & Sotiropoulos, S. N. (2016). An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *Neuroimage*, 125, 1063-1078. doi:10.1016/j.neuroimage.2015.10.019
- Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., . . . Consortium, W.-M. H. (2013). Function in the human connectome: Task-fMRI and individual differences in behavior. *Neuroimage*, 80, 169-189. doi:10.1016/j.neuroimage.2013.05.033
- Bastien, P., Vinzi, V. E., & Tenenhaus, M. (2005). PLS generalised linear regression. *Computational Statistics & Data Analysis*, 48(1), 17-46. doi:10.1016/j.csda.2004.02.005

- 567 Bataille, D., Edwards, A. D., & O'Muircheartaigh, J. (2018). Annual Research Review: Not just a small
568 adult brain: understanding later neurodevelopment through imaging the neonatal brain. *J*
569 *Child Psychol Psychiatry*, 59(4), 350-371. doi:10.1111/jcpp.12838
- 570 Bertrand, F., & Maumy-Bertrand, M. (2018). plsRglm: Partial least squares linear and generalized linear
571 regression for processing incomplete datasets by cross-validation and bootstrap techniques
572 with R. *Arxiv*. doi:arXiv:1810.01005.
- 573 Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and
574 functional systems. *Nat Rev Neurosci*, 10(3), 186-198. doi:10.1038/nrn2575
- 575 Consonni, V., Ballabio, D., & Todeschini, R. (2010). Evaluation of model predictive ability by external
576 validation techniques. *Journal of Chemometrics*, 24(3-4), 194-201. doi:10.1002/cem.1290
- 577 Desikan, R. S., Segonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., . . . Killiany, R. J. (2006).
578 An automated labeling system for subdividing the human cerebral cortex on MRI scans into
579 gyral based regions of interest. *Neuroimage*, 31(3), 968-980.
580 doi:10.1016/j.neuroimage.2006.01.021
- 581 Evans, A. C. (2013). Networks of anatomical covariance. *Neuroimage*, 80, 489-504.
582 doi:10.1016/j.neuroimage.2013.05.054
- 583 Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., . . . Dale, A. M. (2002). Whole
584 brain segmentation: automated labeling of neuroanatomical structures in the human brain.
585 *Neuron*, 33(3), 341-355.
- 586 Fornito, A., Zalesky, A., & Breakspear, M. (2013). Graph analysis of the human connectome: promise,
587 progress, and pitfalls. *Neuroimage*, 80, 426-444. doi:10.1016/j.neuroimage.2013.04.087
- 588 Fornito, A., Zalesky, A., & Bullmore, E. (2016). *Fundamentals of Brain Network Analysis*: Academic
589 Press.
- 590 Galdi, P., Blesa, M., Sullivan, G., Lamb, G. J., Stoye, D. Q., Quigley, A. J., . . . Boardman, J. P. (2018).
591 Neonatal Morphometric Similarity Networks Predict Atypical Brain Development Associated
592 with Preterm Birth. *10842*, 47-57. doi:10.1007/978-3-030-00755-3_6

- 593 Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., . . .
 594 Consortium, W.-M. H. (2013). The minimal preprocessing pipelines for the Human
 595 Connectome Project. *Neuroimage*, 80, 105-124. doi:10.1016/j.neuroimage.2013.04.127
- 596 Glasser, M. F., & Van Essen, D. C. (2011). Mapping Human Cortical Areas In Vivo Based on Myelin
 597 Content as Revealed by T1- and T2-Weighted MRI. *Journal of Neuroscience*, 31(32), 11597-
 598 11616. doi:10.1523/Jneurosci.2180-11.2011
- 599 Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., . . . Thompson, P. M.
 600 (2004). Dynamic mapping of human cortical development during childhood through early
 601 adulthood. *Proc Natl Acad Sci U S A*, 101(21), 8174-8179. doi:DOI 10.1073/pnas.0402680101
- 602 Goulas, A., Uylings, H. B., & Hilgetag, C. C. (2017). Principles of ipsilateral and contralateral cortico-
 603 cortical connectivity in the mouse. *Brain Struct Funct*, 222(3), 1281-1295. doi:10.1007/s00429-
 604 016-1277-y
- 605 Hahn, A., Lanzenberger, R., & Kasper, S. (2019). Making Sense of Connectivity. *International Journal of*
 606 *Neuropsychopharmacology*, 22(3), 194-207. doi:10.1093/ijnp/pyy100
- 607 Heaton, R. K., Akshoomoff, N., Tulsky, D., Mungas, D., Weintraub, S., Dikmen, S., . . . Gershon, R. (2014).
 608 Reliability and Validity of Composite Scores from the NIH Toolbox Cognition Battery in Adults.
 609 *Journal of the International Neuropsychological Society*, 20(6), 588-598.
 610 doi:10.1017/S1355617714000241
- 611 Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and
 612 accurate linear registration and motion correction of brain images. *Neuroimage*, 17(2), 825-
 613 841. doi:10.1006/nimg.2002.1132
- 614 Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). Fsl.
 615 *Neuroimage*, 62(2), 782-790. doi:10.1016/j.neuroimage.2011.09.015
- 616 Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The
 617 Unity and Diversity of Executive Functions: A Systematic Review and Re-Analysis of Latent
 618 Variable Studies. *Psychol Bull*, 144(11), 1147-1185. doi:10.1037/bul0000160

- 619 Krishnan, A., Williams, L. J., McIntosh, A. R., & Abdi, H. (2011). Partial Least Squares (PLS) methods for
620 neuroimaging: A tutorial and review. *Neuroimage*, 56(2), 455-475.
621 doi:10.1016/j.neuroimage.2010.07.034
- 622 Lerman-Sinkoff, D. B., Sui, J., Rachakonda, S., Kandala, S., Calhoun, V. D., & Barch, D. M. (2017).
623 Multimodal neural correlates of cognitive control in the Human Connectome Project.
624 *Neuroimage*, 163, 41-54. doi:10.1016/j.neuroimage.2017.08.081
- 625 Li, W., Yang, C., Shi, F., Wu, S., Wang, Q., Nie, Y., & Zhang, X. (2017). Construction of Individual
626 Morphological Brain Networks with Multiple Morphometric Features. *Front Neuroanat*, 11,
627 34. doi:10.3389/fnana.2017.00034
- 628 Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer*
629 *research*, 27(2 Part 1), 209-220.
- 630 Marcus, D. S., Harwell, J., Olsen, T., Hodge, M., Glasser, M. F., Prior, F., . . . Van Essen, D. C. (2011).
631 Informatics and data mining tools and strategies for the human connectome project. *Front*
632 *Neuroinform*, 5, 4. doi:10.3389/fninf.2011.00004
- 633 Morgan, S. E., Seidlitz, J., Whitaker, K., Romero-Garcia, R., Clifton, N. E., Scarpazza, C., . . . Bullmore, E.
634 T. (2018). Cortical patterning of abnormal morphometric similarity in psychosis is associated
635 with brain expression of schizophrenia related genes. *bioRxiv*, 501494. doi:10.1101/501494
- 636 Nomi, J. S., Vij, S. G., Dajani, D. R., Steimke, R., Damaraju, E., Rachakonda, S., . . . Uddin, L. Q. (2017).
637 Chronnectomic patterns and neural flexibility underlie executive function. *Neuroimage*, 147,
638 861-871. doi:10.1016/j.neuroimage.2016.10.026
- 639 Pagani, M., Bifone, A., & Gozzi, A. (2016). Structural covariance networks in the mouse brain.
640 *Neuroimage*, 129, 55-63. doi:10.1016/j.neuroimage.2016.01.025
- 641 R Core Team. (2016). R: A language and environment for statistical computing (Version 3.3.2). Vienna,
642 Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

- 643 Rosen, A. F. G., Roalf, D. R., Ruparel, K., Blake, J., Seelaus, K., Villa, L. P., . . . Satterthwaite, T. D. (2018).
644 Quantitative assessment of structural image quality. *Neuroimage*, 169, 407-418.
645 doi:10.1016/j.neuroimage.2017.12.059
- 646 Seidlitz, J., Nadig, A., Liu, S., Bethlehem, R. A. I., Vértés, P. E., Morgan, S. E., . . . Raznahan, A. (2019).
647 Transcriptomic and Cellular Decoding of Regional Brain Vulnerability to Neurodevelopmental
648 Disorders. *bioRxiv*. doi:10.1101/573279
- 649 Seidlitz, J., Vasa, F., Shinn, M., Romero-Garcia, R., Whitaker, K. J., Vertes, P. E., . . . Bullmore, E. T.
650 (2018). Morphometric Similarity Networks Detect Microscale Cortical Organization and
651 Predict Inter-Individual Cognitive Variation. *Neuron*, 97(1), 231-247 e237.
652 doi:10.1016/j.neuron.2017.11.039
- 653 Sowell, E. R., Thompson, P. M., Leonard, C. M., Welcome, S. E., Kan, E., & Toga, A. W. (2004).
654 Longitudinal mapping of cortical thickness and brain growth in normal children. *Journal of*
655 *Neuroscience*, 24(38), 8223-8231. doi:10.1523/Jneurosci.1798-04.2004
- 656 Stone, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the*
657 *Royal Statistical Society Series B-Statistical Methodology*, 36(2), 111-147.
- 658 Torgo, L., Branco, P., Ribeiro, R. P., & Pfahringer, B. (2015). Resampling strategies for regression. *Expert*
659 *Systems*, 32(3), 465-476. doi:10.1111/exsy.12081
- 660 Tulsky, D. S., Carlozzi, N. E., Chevalier, N., Espy, K. A., Beaumont, J. L., & Mungas, D. (2013). V. NIH
661 Toolbox Cognition Battery (CB): measuring working memory. *Monogr Soc Res Child Dev*, 78(4),
662 70-87. doi:10.1111/mono.12035
- 663 Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., & Consortium, W.
664 U.-M. H. (2013). The WU-Minn Human Connectome Project: an overview. *Neuroimage*, 80,
665 62-79. doi:10.1016/j.neuroimage.2013.05.041
- 666 Wechsler, D. (1999). *Manual for the Wechsler abbreviated intelligence scale (WASI)*. San Antonio, TX:
667 The Psychological Corporation.

- Wechsler, 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
- 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

