Supplementary Material

1. Supplementary Materials and Methods

1.1 MR Data Processing

MR data analysis was performed by two authors (JY and DL, with 2 and 12 years of experience in neuroimaging, respectively). First, the DICOM file from the original scan was converted into Neuroimaging Informatics Technology Initiative (NIfTI) format; the quality of each image was manually checked, and any with obvious artifacts were discarded. Next, using SPM 12 software segmentation options with the default tissue probability map as an *a priori* template, the 3D-T1 structural MR images were segmented into gray matter, white matter, cerebrospinal fluid, bone, soft tissue and air/background [1]; the quality of the segmented images was manually checked. Using the DARTEL option in the SPM package, a template was created to perform a nonlinear transformation of the original image to the normalized image. Then the DARTEL tool was used to convert the segmented original image to the Montreal Neurological Institute (MNI) coordinate space. Following this, the gray matter image was resampled into 2 mm^3 voxels and spatially smoothed (Gaussian smoothing, 6 mm full width at half maximum).

1.2 Brain Network Construction

For each participant, we first extracted gray matter volume values for all the voxels within each region of interest. The probability density function of these values was estimated using kernel density estimation (KDE) [2,3] with bandwidths chosen automatically [4], and used to calculate the probability distribution function (PDF). Kullback-Leibler divergence measures the difference between two probability distributions, which is equivalently the information lost when one probability

distribution is used to approximate another [5]; Kullback-Leibler divergence-based similarity (KLS) was calculated between all pairs of ROIs using their PDF.

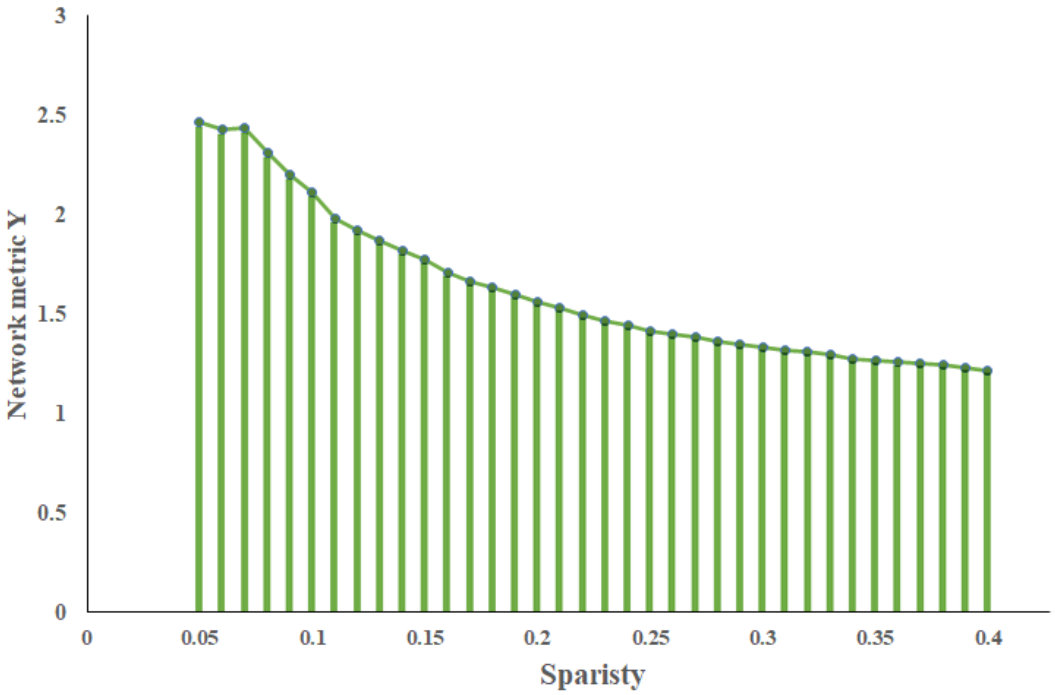
For each participant, this yields a KLS-based morphological connectivity matrix with entries that range from 0-1, where self-connections are set to 0 [6], and 1 signifies two identical distributions. These were used to quantify morphological similarity (considered as connectivity in this application of the graph theory approach) between two regions to estimate interregional network edges. Note that we define this result as a "connection" to refer to brain network edges that have statistically similar morphological connectivity between paired brain regions; this can exist in the absence of direct axonal connectivity or anatomic proximity.

1.3 Support Vector Machine Classification

First, to compare ET and HC, we used Support Vector Machine (SVM) to apply a single-subject classification [7]. SVM uses a set of mathematical functions called kernels to map input vectors to an eigenspace and the linear kernel was preferred to the nonlinear kernel, to minimize the risk of overfitting. In this feature space, the model finds the optimum separation surface which maximizes the margin between different classes within a training dataset. Once the separation surface is determined, an independent testing dataset can be used to predict the class of new observations. An independent set of individuals was used to perform an unbiased assessment of the model's performance, and statistical significance was estimated by permutation (1000 permutations). Specifically, a 10-fold stratified cross-validation scheme was used to separate the original samples (of each dataset) into 10 non-overlapping partitions. In each iteration, one partition was treated as a separate set of tests (used to calculate the performance) and the remaining subjects were defined as training samples. Next, in the training set, we performed an internal cross-validation (i.e. 10-fold stratified nested

cross-validation) to select the optimal set of hyperparameters. Linear SVM has only one hyperparameter (the soft margin parameter C), which is used to control the trade-off between reducing training errors and increasing the separation margin. We optimized this by performing a grid search on the following: C = 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}, 10^{4}. After this internal cross-validation, we could obtain the optimum C value for each input measurement. Next, a second nested cross-validation was performed to optimize the coefficient of each specific measure for the soft voting as we began to perform a multiple measure metric analysis. Each coefficient was evaluated using a grid search, which assumed an integer value between 1 and 10 for the search space. This second nested cross-validation was also performed using a 10-fold hierarchical cross-validation scheme. In these two nested cross-validations, the highest average balanced accuracy of the model was used to find the best hyperparameter value. Finally, after these nested cross-validation steps, the entire training set was used to train the SVM model with the optimal set of hyperparameter values, and the balanced accuracy, specificity and sensitivity of its performance were evaluated on the test set. The final reported balanced accuracy, specificity, and sensitivity were the mean values of the metrics calculated on each partition of the cross-validation scheme. The code used is available at (http://github.com/Warvito/integrating-multi-modal-neuroimaging).

2. Supplementary Figure



Supplementary Fig 1. The area under curve (AUC) for network metrics. The AUC for a network metric Y, which is calculated over the sparsity threshold range of S1 to Sn with interval of ΔS, is computed as $Y^{AUC} = \sum_{k=1}^{n-1}[Y(S_k) + Y(S_{K+1})] * \Delta S/2$. In the current study, we use S1= 0.05, Sn =0.40 and ΔS = 0.01.

REFERENCES

1. Ashburner J, Friston KJ (2005) Unified segmentation. Neuroimage 26 (3):839-851. https://doi.org/10.1016/j.neuroimage.2005.02.018
2. Rosenblatt M (1956) Remarks on Some Nonparametric Estimates of a Density Function. Annals of Mathematical Statistics 27 (3):832-837. https://doi.org/10.1214/aoms/1177728190
3. Parzen E (1962) On Estimation of a Probability Density Function and Mode. Annals of Mathematical Statistics 33 (3):1065-&. https://doi.org/10.1214/aoms/1177704472
4. Botev ZI, Grotowski JF, Kroese DP (2010) Kernel density estimation via diffusion. Annals of Statistics 38 (5):2916-2957. https://doi.org/10.1214/10-aos799
5. Wagenmakers EJ (2003) Model selection and multimodel inference: A practical information-theoretic approach. Journal of Mathematical Psychology 47 (5-6):580-586. https://doi.org/10.1016/s0022-2496(03)00064-6
6. Wang H, Jin X, Zhang Y, Wang J (2016) Single-subject morphological brain networks: connectivity mapping, topological characterization and test-retest reliability. Brain Behav 6 (4):e00448. https://doi.org/10.1002/brb3.448
7. Cortes C, Vapnik V (1995) Support Vector Network. Machine Learning 20 (3):273-297.