Similarity in resting-state functional brain connectivity is not influenced by social closeness in schoolchildren

[1]Carolyn Beth McNabb*, [1]Laura Grace Burgess, [2]Amy Fancourt, [2]Nancy Mulligan, [1]Lily FitzGibbon, [1]Patricia Riddell, [1,3]Kou Murayama

[1]School of Psychology and Clinical Language Sciences, University of Reading, Reading, United Kingdom RG6 7BE

[2]BrainCanDo, Queen Anne's School, Reading, United Kingdom RG4 6DX

[3]Research Institute, Kochi University of Technology, Kami, Kochi, Japan 782-8502

*Corresponding author: Carolyn Beth McNabb, School of Psychology and Clinical Language Sciences, University of Reading, Harry Pitt building, Earley Gate, Reading, United Kingdom RG6 7BE, c.b.mcnabb@reading.ac.uk

Acknowledgements and funding: This research was supported by the JSPS KAKENHI (Grant Number 16H06406 to Kou Murayama), F. J. McGuigan Early Career Investigator Prize from American Psychological Foundation (to Kou Murayama); and the Leverhulme Trust (Grant Number RL-2016-030 to Kou Murayama). This research is also supported by BrainCanDo charitable company. The authors would like to thank the students who gave their time to take part in the social network and imaging components of this study, as well as the parents, teachers and staff of Queen Anne's School who made data collection possible. We also wish to thank the Centre for Integrative Neuroimaging and Neurodynamics at the University of Reading for assistance with scanning and timetabling. Many thanks also to the members of the Motivation Science Lab, University of Reading, who gave their time to assist with social network and imaging data collection.

Abstract

Previous research suggests that the proximity of individuals in a social network predicts how similarly their brains respond to naturalistic stimuli. However, the relationship between social connectedness and brain connectivity in the absence of external stimuli has not been examined. To investigate whether neural homophily between friends exists at rest we collected resting-state functional magnetic resonance imaging (fMRI) data from 68 school-aged girls, along with social network information from all pupils in their year groups (total 5,066 social dyads). Participants were asked to rate the amount of time they voluntarily spent with each person in their year group, and directed social network matrices and community structure were then determined from these data. No statistically significant relationships between social distance, community homogeneity and similarity of global-level resting-state connectivity were observed. Nor were we able to predict social distance using a machine learning technique (i.e. elastic net regression based on the local-level similarities in resting-state whole-brain connectivity between participants). Although neural homophily between friends exists when viewing naturalistic stimuli, this finding did not extend to functional connectivity at rest in our population. Instead, resting-state connectivity may be less susceptible to the influences of a person's social environment.

1. Introduction

Homophily is the tendency of individuals to attract and interact with those who share similar traits. Homophilic selection is observed for broad categorical traits such as gender, ethnicity and sexual orientation [1, 2, 3] but also for personal traits such as motivation [4], personality and cognitive ability [5], and academic achievement [6]. High school and university students have been found to rearrange their local social networks to form ties and clusters with students who have similar performance levels [6] and this type of homophily has been observed even in polygenic scores for academic achievement [7].

Given the predominance of social network homophily for behavioural, personality and cognitive traits, we can reasonably expect that this extends to similarities in brain function. In fact, neural responses observed during unconstrained viewing of naturalistic stimuli (movie clips) were found to be significantly more similar among friends compared with those farther removed in a real-world social network [8]. This effect persisted, even after controlling for inter-subject similarities in demographic variables, such as age, gender, nationality and ethnicity. Social closeness also provides opportunities for behavioural contagion – researchers have shown that social contagion modulates neural representations of risk in reward-related areas and that functional connectivity between the caudate and prefrontal cortex accounts for individual differences in susceptibility to risk-taking contagion [9].

Previously, neural similarity was assessed using intersubject correlation of blood oxygenation level-dependent (BOLD) timeseries across functionally derived regions of the brain. This method of intersubject correlation evaluates the externally generated (extrinsic) stimulus-locked BOLD activation associated with the task but ignores the internally generated (intrinsic) component of BOLD activity, which is cancelled out when correlating across participants [10]. Therefore, it remains unclear whether internally-generated brain activity similarly exhibits neural homophily between friends.

Patterns of brain connectivity elicited from internally generated resting-state BOLD activation are mirrored by activation networks found under explicit task-based activation [11]. For example, resting-state sub-networks have been shown to correspond with externally generated activation from attention, speech, reasoning, emotion, memory and social cognition tasks [11, 12]. Resting-state connectivity is also associated with non-cognitive measures of motivation. Grit and growth mind-set were found to be associated with functional connectivity between ventral striatal and bilateral prefrontal networks important for cognitive-behavioural control [13]. Connectivity at rest also predicts personality. Connectome-based predictive modelling has been used to successfully predict trait-level measures of personality, including openness to experience [14], neuroticism and extraversion [15]. Others have found that global connectivity of the left prefrontal cortex predicts individual differences in fluid intelligence and cognitive control [16] and a clinical measure of attention can be predicted from resting-state connectivity in a network associated with sustained attention [17]. These findings highlight the utility of resting-state connectivity for identifying individual differences in cognition, behaviour and personality, all of which have exhibited homophily within social networks.

Researchers have also linked internally generated brain connectivity with a number of social behaviours. For example, resting-state sub-networks for motor, visual, speech and other language functions have been associated with the quality and quantity of social networks in older adults [18]. Others have demonstrated positive associations between functional connectivity and social network size and embeddedness [19]. There is also evidence for stronger amygdalar connectivity with brain networks subserving perceptual and affiliative behaviours in healthy adults who foster and maintain larger and more complex social networks [20]. Social network size may also dictate the degree of connectivity within the default mode network (DMN). The DMN overlaps considerably with regions important for theory of mind and social cognition [12, 21] and has been found to exhibit greater coupling with anterior cingulate and dorsolateral prefrontal cortex in those with larger social network size [22]. This striking overlap between the DMN and regions involved in social cognition infers a tendency for entertaining thoughts about oneself and others during rest [23]. Despite

investigations into the relationship between internally generated connectivity patterns and social behaviour, however, no study has investigated whether close social relationships are associated with similarities in resting-state connectivity.

In the current study, we collected resting-state fMRI data of pupils attending a single school, along with social network (friendship) information between them. This unique data allow us to test the hypothesis that friends exhibit greater similarity in internally generated connectivity compared with those farther removed in a school-based social network. We tested this hypothesis with a variety of methods based on resting-state network connectivity — whole-brain correlation, graph-theoretic metrics, and machine learning. We evaluated similarities in whole-brain connectivity as well as resting-state subnetworks associated with motivational and social behaviours, which are relevant to schoolchildren.

2. Methods

2.1. Participants

For the social network component of the study, individuals 12-14 years of age in years 8 (cohort 1: n=59; cohort 2: n=51) and 9 (cohort 3: n=65) were recruited from a private girls' day and boarding school in the United Kingdom (as part of a larger study). Participants were recruited during 2017 and 2018 from total year group pools of 62 (cohort 1), 53 (cohort 2) and 75 (cohort 3) students, corresponding to inclusion rates of 95%, 96% and 87%, respectively.

The study was approved by the University Research Ethics Committee. All students gave informed written assent to take part in the study and consent was obtained from parents.

Individuals from cohorts 1, 2 and 3 were also invited to take part in the functional magnetic resonance imaging (fMRI) component of the study. Twenty-eight students from cohort 1, 17 students from cohort 2 and 34 students from cohort 3 were recruited into the fMRI study (cohorts 1-

fMRI, 2-fMRI and 3-fMRI, respectively); of these, 5 participants from cohort 1-fMRI and 6 from cohort 3-fMRI had unusable data due to artefacts caused by dental braces. Children with braces were excluded from cohort 2-fMRI at the time of screening. This resulted in the final inclusion of 68 students in the fMRI component of the study, consisting of 23 (cohort 1-fMRI, 12-13 years of age), 17 (2-fMRI, 12-13 years of age) and 28 (3-fMRI, 13-14 years of age) students in respective cohorts.

The fMRI sample affords a total of 5,066 dyads to examine potential neural homophily. Currently, there is no software available to perform a statistical power analysis with the mixed-effects model we specified to test our hypothesis with the dyadic data (see section 2.6. for linear mixed effects model specification); as such, we created an in-house simulation code to perform a power analysis. The simulation results showed that our sample size is sufficient to detect very small effects of social distance on similarity. Specifically, the study was powered to detect a change in z standardised similarity score of at least .08 with the one unit change in social distance at a power of 80%, and a change in z standardized similarity score of .10 at a power of 95%. As previous neuroimaging work identified a difference of -.2 to -.23 in z standardized similarity scores between social distances 1 to 2 and 2 to 3, respectively [8], we consider our sample size to be sufficiently powered to detect meaningful changes in similarity score between our social distance units.

Exclusion criteria pertaining to all groups consisted of standard safety-related contraindications for MRI. All students gave informed written assent and one parent per child gave informed written consent for their child to participate.

2.2. Data acquisition

The social network data were acquired in class using an online survey administered with SurveyMonkey (SurveyMonkey Inc., San Mateo, California, USA). Social network data were acquired in October and November of 2017 (cohorts 1 and 3) and 2018 (cohort 2), one-to-two months after the start of the academic year. Students had been registered at the school for a maximum of one (cohorts 1 and 2) or two (cohort 3) years when the survey took place. A roster-and-rating method

was used to assess the social networks of students. Specifically, participants were provided with a list of all students in their year group and asked to consider the question: "How much time do you spend interacting with this student?". Students answered on a five-point Likert scale, which included options: "None", "A rare amount", "Some", "More than some" and "Most". Participants were told to consider time spent voluntarily interacting with other students but not time spent in planned seating situations, allocated group work or in classes without opportunities to talk amongst themselves. Investigators were blinded to the identity of students.

In addition, participants were asked to nominate up to five individuals from their cohort with whom they considered themselves to be "close with" [24]. Peer nominations and roster-and-rating methods measure different aspects of peer interactions. Whereas rating assesses general acceptance of peers, nomination is thought to encourage the naming of "best friends" [25]. In the main text, we report results from roster-and-rating data only. However, nomination data provided the same conclusion as roster-and-rating data (see supplementary figures S1 and S2). Participants were asked to report on social interactions within their own year group only.

Structural and resting-state functional MRI data were acquired using a Siemens Magnetom Prisma_fit 3T scanner. All participants were imaged using a 32-channel head coil. A structural T1-weighted image was acquired using an MPRAGE sequence [26] with repetition time (TR) 2300 ms; echo time (TE) 2.29 ms; inversion time 900 ms; flip angle 8°; in-plane acceleration (GRAPPA) factor of 2; field of view (FOV) 240 mm; voxel size 0.9 x 0.9 x 0.9 mm.

Resting-state functional images were acquired over 10 minutes using the Siemens two-dimensional multiband gradient-echo echo planar image (EPI) sequence with TR 1500 ms, TE 30 ms; multiband slice acceleration factor 4; GRAPPA 2; flip angle 66°; echo spacing 0.93 ms; EPI factor 96; phase-encode direction posterior >> anterior; slices 68; volumes 400; FOV 192 mm; voxel size 2 x 2 x 2 mm.

During resting-state data acquisition, participants were asked to lie still with eyes open and look at a blank screen in front of the scanner. Instructions were given to relax and think of nothing in particular.

2.3. Social network characterisation

Social network analysis was performed using the igraph package in R [27, 28]. Data were included only for individuals who took part in the social network survey; any outgoing social ties with those not participating in the survey were removed. Mutually-reported (reciprocal) social ties are deemed to be more robust indicators of friendship than unreciprocated ties. Therefore, an unweighted, undirected graph containing only mutual social ties was used to determine social distance and community affiliation.

Social network data were processed as follows: Roster-and-rating data (5-point scale) were binarised using a threshold of 4 (i.e. only instances in which students spent "more than some" or "most" of the time with another student were included). This threshold was selected to mitigate central tendency bias often reported with Likert-type questionnaires [29]. Any non-mutual connections were then removed (i.e. if $subj_i$ gave $subj_j$ a rating of 4 or greater but $subj_j$ gave $subj_i$ a rating below 4, the connection would be lost). This yielded an unweighted (binary), undirected (reciprocal) adjacency matrix for each cohort, from which social networks graphs were derived. Adjusting the distance threshold to 5 (I spend "most of my time" with this person) or using non-reciprocated (directed) social ties (rather than mutual ties) did not change the overall results of the analyses (see supplementary figures S3-S6).

The distance between two individuals in a social network is an important predictor of behavioural tendencies [30] and may relate to shared patterns in brain function [8]. Social distances in the current study were relatively short compared with previous literature (e.g. Parkinson et al. (2018) reported a

network diameter of 6 using only mutually reported social ties compared to a maximum diameter of 4 in the current study). The highly interconnected nature of our networks may affect how social distance and brain function similarities are associated with one another. Therefore, we also evaluated dyadic similarities as a function of community affiliation. By splitting social networks into smaller friendship communities and evaluating the differences in brain similarity between those inside and outside each community, we measured relationships between neural homophily and social behaviour at a friendship group level.

Measures of social proximity (social distance and community affiliation) were determined separately for each cohort. Social distance was calculated as the shortest path length between each pair of nodes (students) in the network. A dyad (student pair) with a social distance (shortest path length) of 1 represented a relationship in which both students had said they spent “more than some” or “most” of their time with the other student (i.e. they were friends). Social distances of 2 and 3 represented dyad pairs in which students did not possess a reciprocal friendship (i.e. did not have a mutual rating of “more than some” or “most” of the time) but reported a mutual friend or friend of a friend, respectively.

Community structure was ascertained using the Louvain method [31]. This method implements multi-level modularity optimisation to subdivide the network into non-overlapping groups of nodes that maximise the number of within-group (within-module) friendships (edges) and minimised the number of between-group friendships for each cohort.

Each cohort was described in terms of its network characteristics, in particular, its network diameter, modularity, mean path length, reciprocity and density. Network diameter is the length of the longest geodesic distance between two nodes in the network, i.e. the number of edges between $subj_i$ and $subj_j$ when these individuals are the farthest from each other in the network.

Modularity (Q) is a measure of how easily a network segregates into smaller subnetworks and is defined by the equation:

$$Q = \frac{1}{2m}\sum_{ij}\left(A_{ij} - \frac{k_i * k_j}{2m}\right)\delta(c_i, c_j)$$

where m is the number of edges in the network, $A_{ij}$is the element of the A adjacency matrix in row i and column j, k_i is the degree (number of edges associated with a node) of i, c_i is the community to which i belongs and $\delta(c_i, c_j)$ is 1 if $c_i = c_j$ and 0 otherwise. Nonzero values of Q represent deviations from randomness; a value above 0.3 is an indicator of significant community structure in a network [32].

Mean path length is the mean shortest path length (number of edges separating a pair of nodes) between all nodes in the network. Reciprocity defines the proportion of connections in a directed graph that are mutual connections. It is otherwise defined as the probability that the counterpart (j to i) of a directed edge (i to j) is included in the graph. Graph density is the ratio of actual connections (edges) to possible connections in the graph; larger values denote more densely connected networks.

2.4. Functional MRI data analysis

FMRI data processing was carried out using FEAT (FMRI Expert Analysis Tool, version 6.0 [33]) part of the FMRIB Software Library (FSL; Oxford, United Kingdom) [34, 35]. Registration of functional images to high resolution structural and Montreal Neurological Institute (MNI-152) standard space images was carried out using FLIRT [36, 37]. Registration from high resolution structural to standard space was then further refined using FNIRT nonlinear registration [38, 39].

The following pre-statistics processing was applied; motion correction using MCFLIRT [37]; non-brain removal using BET [40]; multiplicative mean intensity normalization of the volume at each time point and high pass temporal filtering (Gaussian-weighted least-squares straight line fitting, with

sigma=50.0s). Independent components analysis (ICA)-based exploratory data analysis was carried out using MELODIC [41], in order to investigate the possible presence of unexpected artefacts. FIX - FMRIB's ICA-based Xnoiseifier [42, 43] was used to auto-classify ICA components into "good" and "bad" components following hand-classification and training using a sample of 10 subjects' data (4 from cohort 1- fMRI and 3 each from cohorts 2-fMRI and 3-fMRI, all randomly selected). "Bad" components were removed from the data and clean data were registered to standard space using warping parameters determined by FEAT.

Analysis of pre-processed data is illustrated in Figure 1. First, motion-corrected fMRI data were divided into 272 parcels using a whole-brain parcellation scheme [44]. This parcellation scheme combines parcels from the Human Brainnetome Atlas [45], a modification of the Desikan-Killiany (DK) atlas [46] and the probabilistic MR atlas of the human cerebellum [47]. Mean blood oxygen level dependent (BOLD) timeseries were extracted from each of the 272 parcellated regions (nodes). Next, the Pearson's correlation coefficient (connection strength) between BOLD timeseries was determined for each pair of nodes in Matlab 2016a (MathWorks, USA). This produced a weighted, undirected whole-brain matrix of functional connectivity for each participant. All self-self nodal connections were removed prior to further analysis.

Z transformations of weighted, undirected whole-brain matrices were then used for inter-subject correlations. Data from each participant were compared with every other participant from the fMRI cohort using a Pearson's correlation. For example, the Z-transformed whole-brain matrix of $subj_i$ was vectorised and correlated with the (vectorised) Z-transformed whole-brain matrix of $subj_j$, then $subj_k$, then $subj_l$, and so on. This method provides a measure of the similarity of connectivity strength in whole-brain networks within participant pairs. Prior to further analysis, correlation strengths were standardised within each cohort to have a mean of 0 and a standard deviation of 1.

Evaluating similarities in resting-state connectivity at a whole-brain level is informative but the inclusion of brain regions unlikely to impact or be impacted by social connectedness could lead to

losses in signal to noise. Focusing on smaller groups of nodes, related to cognitive performance and social interaction would therefore be expected to increase functional similarities associated with friendship. Thus, in addition to whole-brain resting-state connectivity, we evaluated functional resting-state connectivity for subnetworks relevant for social and motivational processing. These included the DMN [48], salience network [49], and left and right frontoparietal networks (lFPN and rFPN, respectively) [50]. DMN and FPNs were extracted from [11]. Z statistic images of ICA maps were then binarised using a threshold of z=4 and warped to 1 mm MNI-152 standard space. Our threshold of z=4 is larger than that reported by [11], who adopted a threshold of z=3 for presentation of resting-state subnetworks. Our larger z threshold was chosen to increase the probability that brain nodes overlapping DMN and FPN spatial maps were truly part of these networks. The salience network was derived using the online meta-analysis tool Neurosynth.org [51] (accessed January 2019; seed positioned in the anterior insular [x=36, y=18, z=4]). The salience network was binarised using a threshold of 0.3, based on results from previous research [52]. The overlap between each resting-state subnetwork and the whole-brain parcellation was determined by applying a binarised mask of the resting-state subnetwork over whole-brain parcellation data. Nodes from the whole-brain parcellation that overlapped with at least 50 voxels from the resting-state subnetwork were used to partition the whole-brain node-to-node matrix. A graphical depiction of the analysis is provided in the supplementary materials (figure S7).

2.5. Brain network characterisation

Graph metrics of functional connectivity were derived from whole-brain weighted, undirected matrices using the Brain Connectivity Toolbox [53] in Matlab (2016a). Analysis steps are provided in Figure 2. Modularity and community structure were calculated using the Louvain method [31], assigning higher values to positively, compared with negatively weighted connections. Modularity was defined as:

$$Q^* = Q^+ + \frac{v^-}{v^+ + v^-} Q^-$$

$$= \frac{1}{v^+} \sum_{ij} \left(w_{ij}^+ - e_{ij}^+\right) \delta\left(M_i, M_j\right) - \frac{1}{v^+ + v^-} \sum_{ij} \left(w_{ij}^- - e_{ij}^-\right) \delta\left(M_i, M_j\right)$$

where Q^+ is modularity from positively weighted connections, Q^- is modularity from negatively weighted connections, v^+ is the sum of all positive connection weights of i, v^- is the sum of all negative connection weights of i, w_{ij}^+ is the present within-module positive connection weights, w_{ij}^- is the present within-module negative connection weights, e_{ij}^+ is the chance expected within-module positive connection weights, e_{ij}^- is the chance expected within-module negative connection weights and $\delta\left(M_i, M_j\right)$ is 1 if i and j are in the same module and 0 otherwise.

The absolute difference in brain modularity within a dyad pair (i.e. between $subj_i$ and $subj_j$) was calculated for every pair of dyads in a social network. Differences were then standardised within each cohort to have a mean of 0 and standard deviation of 1.

Nodal strength (a node-level measure of centrality, the importance of a node in its network) and diversity (a node-level measure of integration that takes into account the strength of a node within its own module) were also calculated using asymmetric values for positively and negatively weighted connections, as previously described by Rubinov & Sporns (2011). Strength was defined as:

$$S_i'^* = s_i'^+ - \left(\frac{s_i^-}{s_i^+ + s_i^-}\right) s_i'^-$$

where $s_i'^+$ is the normalised sum of positive connection weights associated with node i, and s_i^+ and s_i^- are the raw sums of positive and negative connections weights, respectively, associated with node i.

Diversity was defined as:

$$h_i^* = h_i^+ - \left(\frac{s_i^-}{s_i^+ + s_i^-}\right) h_i^-$$

and

$$h_i^{\pm} = \frac{1}{\log m} \sum_{u \in M} p_i^{\pm}(u) \log p_i^{\pm}(u)$$

where $p_i^{\pm}(u) = \frac{s_i^{\pm}(u)}{s_i^{\pm}}$, $s_i^{\pm}(u)$ is the strength of node i within module u (the total weight of connections of i to all nodes in u) and m is the number of modules in modularity partition M.

Nodal strength and diversity data from each participant were compared with every other participant from the fMRI cohort using a Pearson's correlation. Prior to further analysis, correlation strengths were standardised within each cohort to have a mean of 0 and a standard deviation of 1.

2.6. Dyadic similarities in functional connectivity as a function of social proximity

To test our hypothesis, we examined whether similarities in resting-state connectivity and brain network characterization were explained by social proximity. In this set of analyses, each student dyad (pair of students) in the fMRI cohort was described by two independent variables: social distance (i.e. shortest path length between the pair within the full year group cohort) and community affiliation (i.e. whether or not two students belonged to the same community, determined using the Louvain community detection method). Dependent variables were similarities in whole-brain connectivity, resting-state subnetwork (DMN, salience network, lFPN and rFPN) connectivity, nodal strength, nodal diversity (each determined by the Pearson's correlation coefficient), and modularity (determined by the absolute difference between members in the dyad).

Using pairs of students (dyads) as the unit of analysis creates dependence in the data, caused by the involvement of every student in multiple dyads (cross-nesting) [54 55]. Using ordinary least-square methods in such data potentially increases Type-1 error rates. We accounted for this dependence structure by including each dyad member (student) as a random variable in a linear mixed effects

(LME) model with crossed random effects (see [56], for a similar model specification). Any subject-specific effects on the dyadic outcomes are therefore accounted for in the model [57].

More specifically, the LME model is specified as:

$$Similarity_{ij} = \beta_{00} + \beta_{01} * social\ distance_{ij} + subj_i + subj_j + e_{ij}$$

$Similarity_{ij}$ is the similarity in brain function between students i and j, where β_{00} is the intercept, and β_{01} represents the relationship between social distance and brain similarity; $subj_i$ and $subj_j$ are crossed random effects (i.e. student-specific effects) of students i and j, respectively, and e_{ij} are residuals. We posited a Gaussian distribution for the random effects and residuals. No distribution assumption was made for social distance. Where social network data were described using community affiliation, $social\ distance_{ij}$ in the above equation was replaced with $community\ similarity_{ij}$ (i.e. whether or not two students belong to the same friendship community, as determined by the Louvain method).

Data were analysed using the lme4 package [58] in R. Dependent variables included similarities in whole-brain, resting-state subnetwork and graph theory measures of connectivity between students (dyads) in each cohort. For each dependent variable, we tested a model with social distance as the independent variable, and tested another model with community affiliation as the independent variable.

We also tested a model with random slopes (i.e. random slopes of the subjects were added to the model above). This model failed to converge in most of the analyses. The omission of random slopes could make the statistical test less conservative [59, 60]. However, as our results all showed non-significant effects in the model without random slopes, we decided not to pursue the model with random slopes any further.

Effects of demographic data as well as their interactions with social distance or community affiliation were also included in a second series of models to determine whether boarding status (i.e. whether

students matched in their boarding status [boarding or day student]) and ethnicity (i.e. whether or not students reported belonging to the same ethnic group) affected the relationship between social proximity and similarity in resting state connectivity between students. As only one student from each fMRI cohort was left-handed, we elected not to include handedness as a covariate.

Cohorts were analysed separately to preserve the independence of the three social networks. Then, we integrated the regression coefficients of the LME models from three cohorts using a random-effect meta-analysis with the Metafor package [61] in R.

2.7. Data-driven predictive model of social proximity from neural similarity

In the previous analysis, we assessed the overall similarity of the whole brain network or subnetworks between pairs of students and examined whether social distance is related to the overall similarity. Although this analysis gives us the most straightforward test of our hypothesis, the analysis does not address the possibility that social distance is represented by the collection of local-level similarities (i.e. similarity between a specific pair of nodes). To examine whether any local-level similarity in the brain functional connectivity encodes social distance, we employed regularised regression techniques to predict social distance between two students based on similarities in their functional brain connectivity of all pairs of nodes. Specifically, we computed the absolute difference in connection strength for each edge in the 272 node brain network (i.e. each node-to-node connection) for every student dyad in the fMRI cohort. This yielded $Nnodes * (Nnodes - 1)/2 = 36{,}856$ similarity (absolute difference) measures for each student dyad, where $Nnodes$ is the number of nodes in the whole-brain parcellation (i.e. 272). We then employed elastic-net regularised linear regression to predict social distance of student dyads from the 36,856 similarity measures of resting-state functional connectivity. This regression technique has been used to successfully predict a variety of outcomes from MRI data, including openness to experience (a Big Five personality trait) [14], psychosis [62], and progression to Alzheimer's disease in people with mild cognitive impairment [63], demonstrating its ability to cope with the high dimensionality of MRI data.

Elastic-net regression combines penalty features of lasso and ridge regression techniques to shrink the coefficients of some regressors toward zero to deal with high dimensional data. Whereas lasso regression can shrink unnecessary regressors to zero and thereby reduce the number of predictors, ridge regression retains all regressors for inclusion in the model. Both lasso and ridge regression techniques have been shown to perform well when dealing with high-dimensional data under various conditions [64]. Combining the two approaches, elastic-net regression allows for adjustment of the lasso-to-ridge ratio (α), providing greater opportunity for better model fits [65].

Elastic-net regressions of connectivity and social distance data were conducted in R using the glmnet package [66]. Three regression models were trained using data from two fMRI cohorts each (see Table 1). One fMRI cohort's data were withheld during training so that the performance of the regression model could be evaluated using a previously unseen set of data.

The best-performing regression model for each training set was determined by optimising the tuning parameters λ and α (see Table 1). λ is a tuning parameter for the shrinkage penalty used to adjust the regression coefficients in the elastic-net regression. When λ = 0, the penalty term has no effect but as λ tends toward infinity, the shrinkage penalty grows, and the regression coefficient estimates approach zero. The optimal value of λ for each regression model was determined using a 10-fold nested cross-validation within the training data. The largest value of λ such that the cross validation error was within one standard error of the minimum was selected and the model was re-fit using all available observations.

The tuning parameter α dictates the ratio of ridge to lasso in the elastic-net regression. α values between 0 and 1 (with iterations of 0.1) were evaluated for each regression model. The best performing α value was chosen based on the smallest root mean squared error (RMSE) between predicted and observed values of social distance in the training data. Best-performing α values were α = 0 (pure ridge regression) for models 1 & 2 and α = 1 (pure lasso regression) for model 3. The

performance of the optimised regression model was then evaluated by predicting the social distance of dyads in the previously unseen testing dataset.

After obtaining the predicted social distance scores from the elastic-net regression models, we evaluated the accuracy of predictive models by examining the relationship between the predicted social distance and the observed (actual) social distance. As before, there is a dependency of the data structure, owing to the involvement of each student in multiple dyads, potentially inflating the test statistics assessing statistical significance of the accuracy scores. To account for the dyadic nature of the data, we again used an LME model to obtain *p* value using the observed social distance as the dependent variable and the predicted social distance as the independent variable, with $subj_i$ and $subj_j$ included as crossed random effects.

Finally, to evaluate the overall predictability of social distance using similarity in resting-state connectivity, we conducted a meta-analysis of LME models from the three elastic-net regressions (models 1, 2 and 3). A positive beta weight with 95% confidence interval excluding zero would indicate that prediction of social distance from neural similarity is feasible.

3. Data availability

The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

4. Results

4.1. Participant and network characteristics

Demographic data for each cohort are presented in Table 2. fMRI cohorts were relatively well matched to full year group cohorts for most demographic characteristics. Network characteristics for each cohort are presented in Table 3. Relatively higher modularity and mean path length in cohort 3 suggest that this network was more segregated and less well integrated compared with the younger two cohorts (cohorts 1 and 2).

4.2. Social network characterisation

The social networks for each cohort are depicted in Figure 3; fMRI cohorts were well distributed within whole year group samples and include both highly influential and less influential students (determined by Eigenvector centrality). Community detection analysis identified four communities (modules) in cohorts 1 and 3 and three communities in cohort 2 (Figure 4).

4.3. Dyadic similarities in functional connectivity as a function of social distance

Similarities in functional connectivity were assessed using LMEs on a whole-brain network level as well as within individual resting-state subnetworks for each fMRI cohort. No statistically significant relationship between social distance and similarity in functional brain connectivity was observed for any fMRI cohort at any resting-state network level (see Figure 5 for example plot). Beta weights (slopes) and 95% confidence intervals for LME models are presented in Figure 6a. Corresponding t and p values for individual tests are provided in the supplementary materials (table S1). Meta-analyses of LME models did not reveal any significant effects of social distance on the degree of functional brain similarity between students. Likewise, there was no relationship between the similarity in strength, diversity or modularity of students' brains and their distance from one another in the social network (Figure 7a). These results indicate that the minimum path length between two individuals in a social network is not associated with similarities in brain function at rest, either at a whole-brain network or subnetwork level.

4.4. Dyadic similarities in functional connectivity as a function of community affiliation

Students' relationships were then defined by their affiliation to different communities (friendship groups), derived using the Louvain method for community detection. Dyads including students from

the same community were scored 1 and dyads including students from different communities were scored 0. LME models predicting brain similarity as a function of community affiliation did not support an effect of friendship grouping as a predictor of connectivity similarity. This was true for whole-brain and resting-state subnetwork connectivity (Figure 6b), as well as for graph measures of connectivity (Figure 7b). These results support those of the social distance analysis. They indicate that the similarity in resting brain function is no greater between individuals in the same social community (which we used here to estimate friendship groups) than those spanning different communities.

Inclusion of demographic variables (i.e. similarities in ethnicity and boarding status) in LME models did not affect the statistical significance of the overall findings (see table S2 for *t* and *p* values of individual tests).

4.5. Prediction of social distance based on node-to-node neural similarities

Using elastic-net regression, we sought to determine whether similarities in node-to-node connectivity within the brain could predict social distance between students. Data were split into training and testing sets such that each cohort was used to train two models and test the predictive validity of a third model. Assignments of training and testing sets for the three models are provided in Table 1. Optimal parameters (i.e. α and λ values) for each model were determined using training data; these parameters were then used to predict social distance in the testing set (to which the model was naïve).

RMSE quantifies how much a set of predicted values differ from their observed counterparts by measuring the standard deviations of the prediction errors. Lower values indicate smaller errors. RMSEs were 0.60, 0.65 and 0.82 for models 1, 2 and 3, respectively. Correlations for the observed vs predicted distance with the p values of the beta weights obtained from the LMEs between each pair

of participants are presented in Figure 8a. Meta-analysis of LME models showed poor predictive power of models to classify social distance of dyads based on whole-brain functional connectivity (Figure 8b). Our meta-analysis results suggest that our predictive models will not extrapolate well to predict social distance in previously unseen social networks.

5. Discussion

Our results provide little evidence for homophily of internally generated (resting-state) functional brain connectivity in school-based social networks. Neither whole-brain nor subnetwork-based analysis (i.e. subnetworks relevant to social and motivational processing) of resting-state connectivity resulted in significant differences in similarity between friends and those farther removed in their social network. Likewise, graph theoretical measures of brain connectivity, including modularity, diversity and strength, were no more similar among friends than other more distantly connected pairs of students. Results from elastic-net regression, using the whole collection of local-level connectivity to predict social distance, also provide minimal evidence for similarity in resting-state functional connectivity among friends. Our findings were robust across individual cohorts of students and demonstrated a consistent non-significant result for homophily of resting-state connectivity. The inclusion of a data-driven approach to analysis (elastic-net regression) suggests our lack of evidence for the hypothesis is not due to noise from irrelevant variables or poor a priori selection of resting-state subnetworks for dimensionality reduction. Likewise, results are not due to poor sampling from the overall population. Students participating in the fMRI component of the study exhibited demographic characteristics representative of their original class cohorts, from which 87% - 96% of enrolled students provided social network data.

Current literature supports a role for social closeness in synchronisation of neural activation. For example, students who report higher social closeness to one another who also engage in silent eye contact prior to class exhibit stronger pairwise brain-to-brain synchrony during class activities

compared with those less close who engage in eye contact [67]. This increased synchronisation between friends is evident even when friends are in the absence of one another; using functional MRI of individual students in a real-world social network, Parkinson and colleagues (2018) investigated synchronisation of neural activation during video clip viewing and found evidence for homophily at the neural response level. Brain regions where response similarity was associated with social network proximity included areas implicated in motivation, learning, affective processing, memory, attention, theory of mind and language processing [8], some behavioural traits of which have exhibited homophily in previous non-imaging studies [4, 5, 68, 69]. These results suggest that, at least in terms of cognitive processing, similarities in behaviour relate to similarities in brain function.

In contrast, the current study examined neural homophily during a resting state scan. Our findings suggest that neural homophily observed in previous work may be specific to stimulus-evoked activation, and may not extend to stimulus-free intrinsically-generated brain activities. Importantly, stimuli used by Parkinson and colleagues (2018) included video clips of comedy, debates and documentaries, intended to evoke social and emotional responses from participants. The homophily observed in their study may therefore be dependent on cognitive processes important for social interaction. This would also explain how our resting-state experiment, which was relatively devoid of social context, failed to elicit homophilic outcomes.

It should be noted that, in a stimulus-free environment, i.e. during rest, subjects are free to mind wander, providing no time-locked cue with which to directly compare activation between two subjects. Instead, simultaneous activation of spatially disparate brain regions is evaluated within each subject to identify networks of brain regions that exhibit highly correlated patterns of activity. Therefore, in the current study, we evaluated the correlational strength between every node in the brain for each individual participant and correlated the whole-brain or subnetwork connectivity pattern between every pair of participants in the social network. This was a powerful approach to evaluating dyadic similarities in resting-state brain function in a social network, but at the same

time, this novel approach makes it difficult to directly compare the current findings with the previous one [8], which focused on the similarities of the activation pattern rather than the pattern of brain functional connectivity. Given the evidence that the architecture of task-based networks closely resembles networks seen at rest [70], it may be an interesting future inquiry to examine whether neural homophily is observed in the brain functional network connectivity triggered by external cues and stimuli.

A notable difference between our sample population and that of Parkinson et al. is the age at which social network and imaging data were collected (schoolchildren vs undergraduate students). Differences in DMN connectivity have been reported between adults and children [71] as well as within individuals throughout early adolescence [72]. In addition, researchers have found associations between pubertal development and strength of intrinsic functional connectivity [73]. Younger samples may therefore exhibit less intrinsic network homophily than more mature samples due to greater brain variability between subjects.

The current study benefits from several strengths. Most notably, we evaluated homophily of resting-state connectivity in three different social networks comprising students of the same gender and similar age and education level, eliminating by design these demographic variables as possible sources of confound for neural homophily. Cohorts were evaluated as independent samples and then as individual studies in meta-analyses, ensuring sufficient statistical power of the overall analysis to find neural homophily. Homophily based on cognitive ability has been reported at a higher rate among girls compared with boys [5] and polygenic scores for educational achievement are more homogenous in women's female social networks compared with men's male social networks [7]. As intelligence and cognitive control are reflected in the resting-state connectivity [16], we therefore anticipated that an all-female sample would exhibit stronger homophily of resting-state connectivity than a mixed-gender or all-male sample. Based on these attributes, our study was well designed to identify homophily at the resting-state connectivity level.

These results contribute to the homophily literature by suggesting that homophily at the neural level may require some external stimulus that engages individuals in social or cognitive thoughts before synchronisation or similarities in connectivity are evident. To further our understanding of how our brain functioning is shaped by social factors, future research should examine the exact conditions under which neural homophily can be observed.

1. Shrum W, Cheek NH, Saundra Mac DH. Friendship in School: Gender and Racial Homophily. *Sociology of Education* **61**, 227-239 (1988).

2. Tuma NB, Hallinan MT. The Effects of Sex, Race, and Achievement on Schoolchildren's Friendships*. *Soc Forces* **57**, 1265-1285 (1979).

3. Logan LS. Status Homophily, Sexual Identity, and Lesbian Social Ties. *J Homosex* **60**, 1494-1519 (2013).

4. Shin H, Ryan AM. Friendship networks and achievement goals: an examination of selection and influence processes and variations by gender. *J Youth Adolesc* **43**, 1453-1464 (2014).

5. Ilmarinen V-J, Vainikainen M-P, Verkasalo MJ, Lönnqvist J-E. Homophilous Friendship Assortment Based on Personality Traits and Cognitive Ability in Middle Childhood: The Moderating Effect of Peer Network Size. *European Journal of Personality* **31**, 208-219 (2017).

6. Smirnov I, Thurner S. Formation of homophily in academic performance: Students change their friends rather than performance. *PLoS One* **12**, e0183473 (2017).

7. Domingue BW, Belsky DW, Fletcher JM, Conley D, Boardman JD, Harris KM. The social genome of friends and schoolmates in the National Longitudinal Study of Adolescent to Adult Health. *Proc Natl Acad Sci U S A* **115**, 702-707 (2018).

8. Parkinson C, Kleinbaum AM, Wheatley T. Similar neural responses predict friendship. *Nature Communications* **9**, 332 (2018).

9. Suzuki S, Jensen ELS, Bossaerts P, O'Doherty JP. Behavioral contagion during learning about another agent's risk-preferences acts on the neural representation of decision-risk. *Proc Natl Acad Sci U S A* **113**, 3755-3760 (2016).

10. Nummenmaa L, Lahnakoski JM, Glerean E. Sharing the social world via intersubject neural synchronisation. *Current Opinion in Psychology* **24**, 7-14 (2018).

11. Smith SM*, et al.* Correspondence of the brain's functional architecture during activation and rest. *Proc Natl Acad Sci U S A* **106**, 13040-13045 (2009).

12. Laird AR*, et al.* Behavioral interpretations of intrinsic connectivity networks. *J Cogn Neurosci* **23**, 4022-4037 (2011).

13. Myers CA, Wang C, Black JM, Bugescu N, Hoeft F. The matter of motivation: Striatal resting-state connectivity is dissociable between grit and growth mindset. *Soc Cogn Affect Neurosci* **11**, 1521-1527 (2016).

14. Dubois J, Galdi P, Han Y, Paul LK, Adolphs R. Resting-state functional brain connectivity best predicts the personality dimension of openness to experience. *Personality neuroscience* **1**, (2018).

15. Hsu W-T, Rosenberg MD, Scheinost D, Constable RT, Chun MM. Resting-state functional connectivity predicts neuroticism and extraversion in novel individuals. *Soc Cogn Affect Neurosci* **13**, 224-232 (2018).

16. Cole MW, Yarkoni T, Repovš G, Anticevic A, Braver TS. Global connectivity of prefrontal cortex predicts cognitive control and intelligence. *J Neurosci* **32**, 8988-8999 (2012).

17. Rosenberg MD*, et al.* A neuromarker of sustained attention from whole-brain functional connectivity. *Nat Neurosci* **19**, 165 (2015).

18. Pillemer S, Holtzer R, Blumen HM. Functional connectivity associated with social networks in older adults: A resting-state fMRI study. *Soc Neurosci* **12**, 242-252 (2017).

19. Joo W-t, Kwak S, Youm Y, Chey J. Brain functional connectivity difference in the complete network of an entire village: the role of social network size and embeddedness. *Sci Rep* **7**, 4465 (2017).

20. Bickart KC, Hollenbeck MC, Barrett LF, Dickerson BC. Intrinsic Amygdala–Cortical Functional Connectivity Predicts Social Network Size in Humans. *The Journal of Neuroscience* **32**, 14729-14741 (2012).

21. Mars RB, Neubert F-X, Noonan MP, Sallet J, Toni I, Rushworth MF. On the relationship between the “default mode network” and the “social brain”. *Front Hum Neurosci* **6**, 189 (2012).

22. Noonan M, Mars R, Sallet J, Dunbar R, Fellows L. The structural and functional brain networks that support human social networks. *Behav Brain Res* **355**, 12-23 (2018).

23. Schilbach L, Eickhoff SB, Rotarska-Jagiela A, Fink GR, Vogeley K. Minds at rest? Social cognition as the default mode of cognizing and its putative relationship to the "default system" of the brain. *Conscious Cogn* **17**, 457-467 (2008).

24. Coie JD, Dodge KA, Coppotelli H. Dimensions and types of social status: A cross-age perspective. *Dev Psychol* **18**, 557 (1982).

25. Schofield JW, Whitley Jr BE. Peer Nomination versus Rating Scale Measurement of Children's Peer Preferences in Desegregated Schools. *Soc Psychol Q* **46**, 242-251 (1982).

26. Brant-Zawadzki M, Gillan GD, Nitz WR. MP RAGE: a three-dimensional, T1-weighted, gradient-echo sequence--initial experience in the brain. *Radiology* **182**, 769-775 (1992).

27. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal, Complex Systems* **1695**, 1-9 (2006).

28. R Core Development Team. R: A Language and Environment for Statistical Computing.). 1.1.4 edn. R Foundation for Statistical Computing (2012).

29. Stevens SS. Issues in psychophysical measurement. *Psychol Rev* **78**, 426 (1971).

30. Apicella CL, Marlowe FW, Fowler JH, Christakis NA. Social networks and cooperation in hunter-gatherers. *Nature* **481**, 497 (2012).

31. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* **2008**, P10008 (2008).

32. Clauset A, Newman ME, Moore C. Finding community structure in very large networks. *Physical review E* **70**, 066111 (2004).

33. Woolrich MW, Ripley BD, Brady M, Smith SM. Temporal autocorrelation in univariate linear modeling of FMRI data. *Neuroimage* **14**, 1370-1386 (2001).

34. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. FSL. *Neuroimage* **62**, 782-790 (2012).

35. Smith SM*, et al.* Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* **23**, S208-S219 (2004).

36. Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. *Med Image Anal* **5**, 143-156 (2001).

37. Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17**, 825-841 (2002).

38. Andersson JLR, Jenkinson M, Smith S. TR07JA1 : Non-linear optimisation.). FMRIB Centre, Oxford, United Kingdom (2007).

39. Andersson JLR, Jenkinson M, Smith S. TR07JA2 : Non-linear registration, aka Spatial normalisation.). FMRIB Centre, Oxford, United Kingdom (2007).

40. Smith SM. Fast robust automated brain extraction. *Hum Brain Mapp* **17**, 143-155 (2002).

41. Beckmann CF, Smith SM. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans Med Imaging* **23**, 137-152 (2004).

42. Griffanti L*, et al.* ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *Neuroimage* **95**, 232-247 (2014).

43. Salimi-Khorshidi G, Douaud G, Beckmann CF, Glasser MF, Griffanti L, Smith SM. Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *Neuroimage* **90**, 449-468 (2014).

44. McNabb CB*, et al.* Functional network dysconnectivity as a biomarker of treatment resistance in schizophrenia. *Schizophr Research* **195**, 160-167 (2018).

45. Fan L*, et al.* The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. *Cerebral Cortex (New York, NY)* **26**, 3508-3526 (2016).

46. Desikan RS*, et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968-980 (2006).

47. Diedrichsen J, Balsters JH, Flavell J, Cussans E, Ramnani N. A probabilistic MR atlas of the human cerebellum. *Neuroimage* **46**, 39-46 (2009).

48. Li W, Mai X, Liu C. The default mode network and social understanding of others: what do brain connectivity studies tell us. *Front Hum Neurosci* **8**, (2014).

49. Silverman MH, Jedd K, Luciana M. Neural networks involved in adolescent reward processing: An activation likelihood estimation meta-analysis of functional neuroimaging studies. *Neuroimage* **122**, 427-439 (2015).

50. Meyer ML, Taylor SE, Lieberman MD. Social working memory and its distinctive link to social cognitive ability: an fMRI study. *Soc Cogn Affect Neurosci* **10**, 1338-1347 (2015).

51. Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD. Large-scale automated synthesis of human functional neuroimaging data. *Nature methods* **8**, 665 (2011).

52. Menon V. Salience Network. In: *Brain mapping: An encyclopedic reference* (ed Toga AW). Academic Press (2015).

53. Rubinov M, Sporns O. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* **52**, 1059-1069 (2010).

54. Kenny DA, Kashy DA, Cook WL. *Dyadic data analysis*. Guilford press (2006).

55. Snijders TA, Kenny DA. The social relations model for family data: A multilevel approach. *Personal Relationships* **6**, 471-486 (1999).

56. Chen G, Taylor PA, Shin Y-W, Reynolds RC, Cox RW. Untangling the relatedness among correlations, Part II: Inter-subject correlation group analysis through linear mixed-effects modeling. *Neuroimage* **147**, 825-840 (2017).

57. Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language* **59**, 390-412 (2008).

58. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *2015* **67**, 48 (2015).

59. Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language* **68**, 10.1016/j.jml.2012.1011.1001 (2013).

60. Murayama K, Sakaki M, Yan VX, Smith GM. Type I error inflation in the traditional by-participant analysis to metamemory accuracy: a generalized mixed-effects model perspective. *J Exp Psychol Learn Mem Cogn* **40**, 1287-1306 (2014).

61. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. *2010* **36**, 48 (2010).

62. Salvador R*, et al.* Evaluation of machine learning algorithms and structural features for optimal MRI-based diagnostic prediction in psychosis. *PLoS One* **12**, e0175683 (2017).

63. Teipel SJ, Kurth J, Krause B, Grothe MJ. The relative importance of imaging markers for the prediction of Alzheimer's disease dementia in mild cognitive impairment — Beyond classical regression. *NeuroImage: Clinical* **8**, 583-593 (2015).

64. Sirimongkolkasem T, Drikvandi R. On Regularisation Methods for Analysis of High Dimensional Data. *Annals of Data Science*, 1-27 (2019).

65. James G, Witten D, Hastie T, Tibshirani R. *An introduction to statistical learning*. Springer (2013).

66. Friedman JH, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* **33**, 22 (2010).

67. Dikker S*, et al.* Brain-to-Brain Synchrony Tracks Real-World Dynamic Group Interactions in the Classroom. *Curr Biol* **27**, 1375-1380 (2017).

68. Tesser A, Campbell J, Smith M. Friendship choice and performance: Self-evaluation maintenance in children. *J Pers Soc Psychol* **46**, 561 (1984).

69. Shin H, Ryan AM. Early adolescent friendships and academic adjustment: Examining selection and influence processes with longitudinal social network analysis. *Dev Psychol* **50**, 2462 (2014).

70. Cole Michael W, Bassett Danielle S, Power Jonathan D, Braver Todd S, Petersen Steven E. Intrinsic and Task-Evoked Network Architectures of the Human Brain. *Neuron* **83**, 238-251 (2014).

71. Fair DA*, et al.* The maturing architecture of the brain's default network. *Proc Natl Acad Sci U S A* **105**, 4028-4032 (2008).

72. Sherman LE, Rudie JD, Pfeifer JH, Masten CL, McNealy K, Dapretto M. Development of the Default Mode and Central Executive Networks across early adolescence: A longitudinal study. *Dev Cogn Neurosci* **10**, 148-159 (2014).

73. Ernst M*, et al.* Pubertal maturation and sex effects on the default-mode network connectivity implicated in mood dysregulation. *Translational psychiatry* **9**, 103-103 (2019).

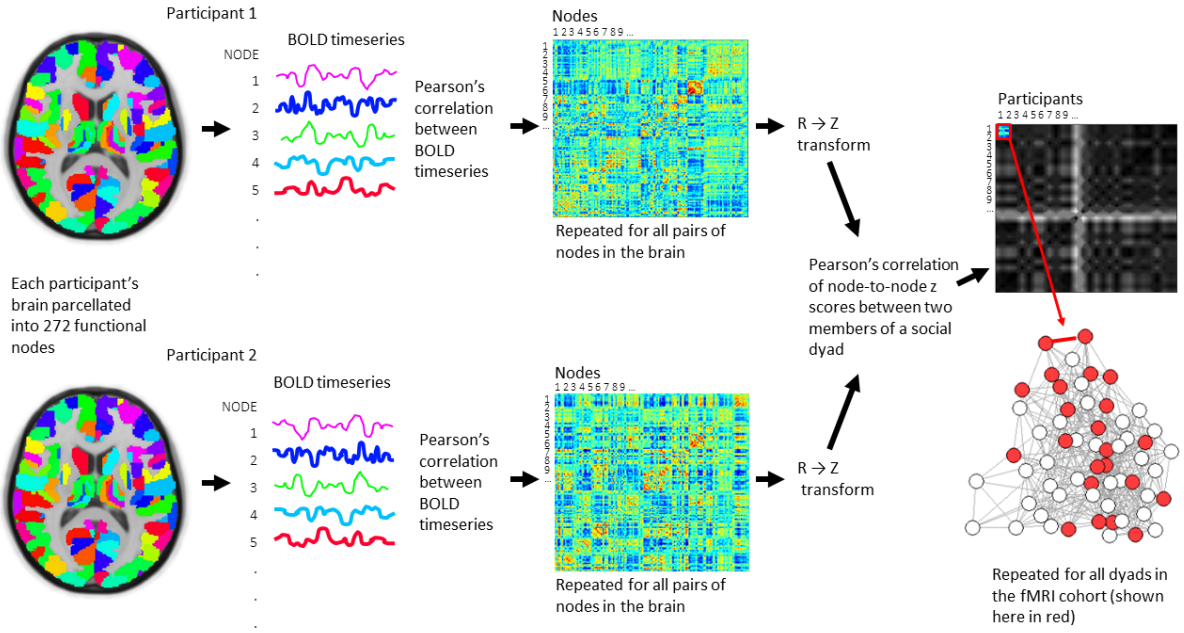

Figure 1. Processing pipeline for determining similarity of resting-state functional connectivity between participants. Mean blood oxygen-dependent (BOLD) timeseries data for each brain parcel were correlated within participants and then each participant's and then whole-brain correlation matrices were compared between participants.

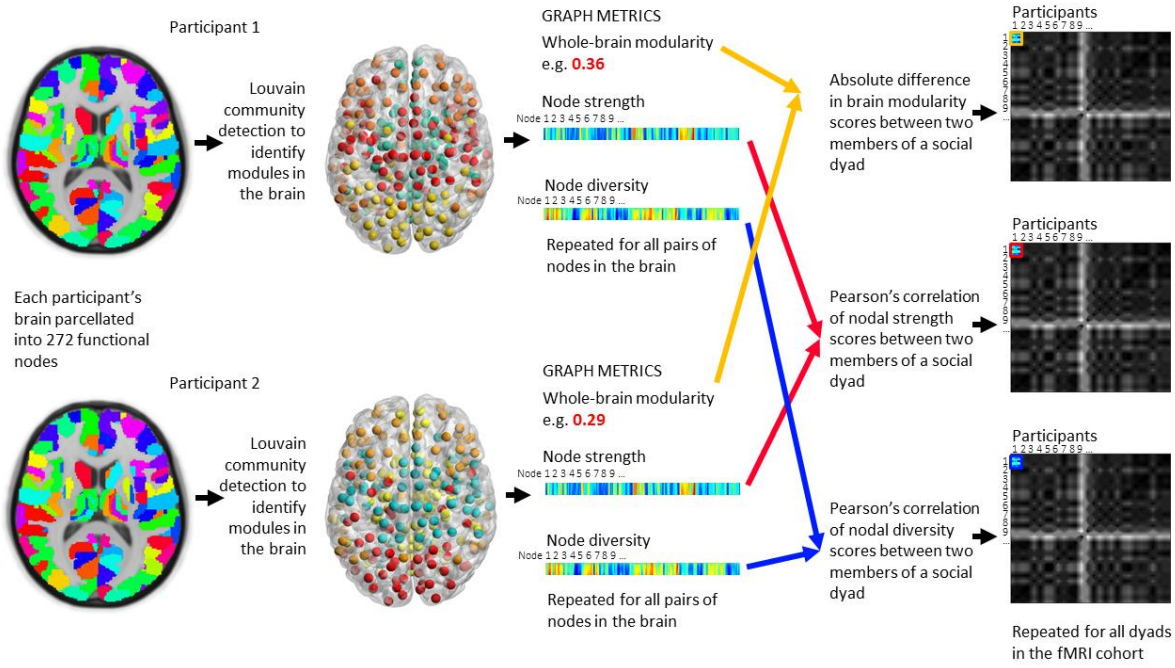

Figure 2. Processing pipeline for graph metric analysis of resting-state data. Community detection was performed on each participant's brain data to determine modules (communities) from which further graph metrics could be analysed. Brain modularity, nodal strength and nodal diversity were determined for each participant and compared between all participants in the MRI cohorts.

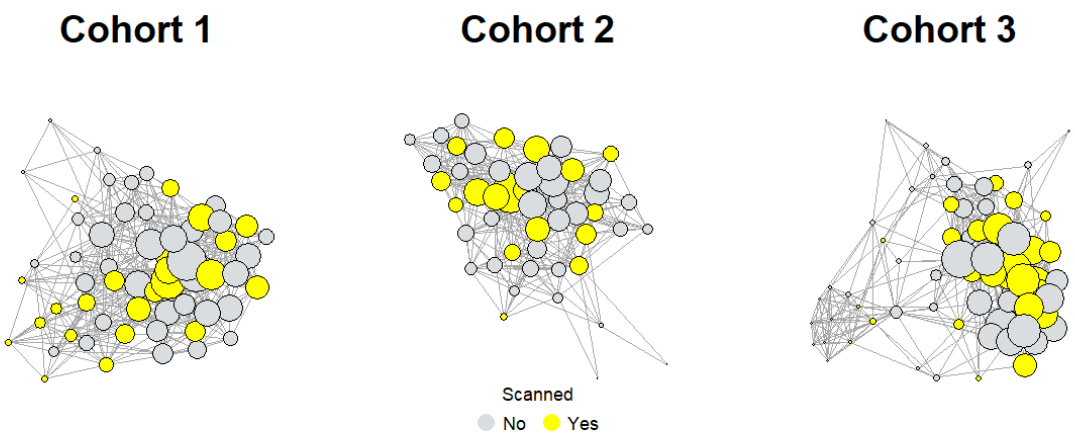

Figure 3. Social networks of cohorts 1, 2 and 3, depicting reciprocal friendships. Nodes represent students; lines (edges) represent mutually reported social ties where students rated the amount of time they spent with each other as "more than some" or "most" of the time. FMRI cohorts are depicted in green; students who provided information about their social interactions but were not included in the fMRI cohort are shown in grey. The size of each node depicts the Eigenvector centrality of that student. Eigenvector centrality is a measure of the relative importance/influence of a node in the network. Nodes with high importance (those who are themselves well-connected and are connected to others who are well-connected) have higher centrality (these are the largest nodes in the network), those with low importance have low centrality (these are the smallest nodes in the network).

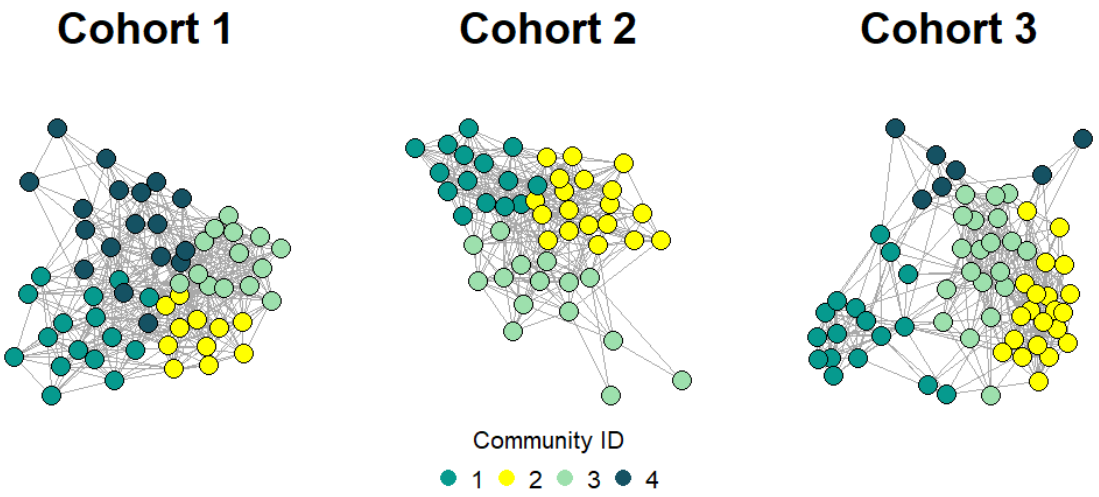

Figure 4. Community structure of each cohort, determined using the Louvain method [31]. Community detection was used to estimate "friendship" groups within each cohort. The fMRI study included students from all communities.

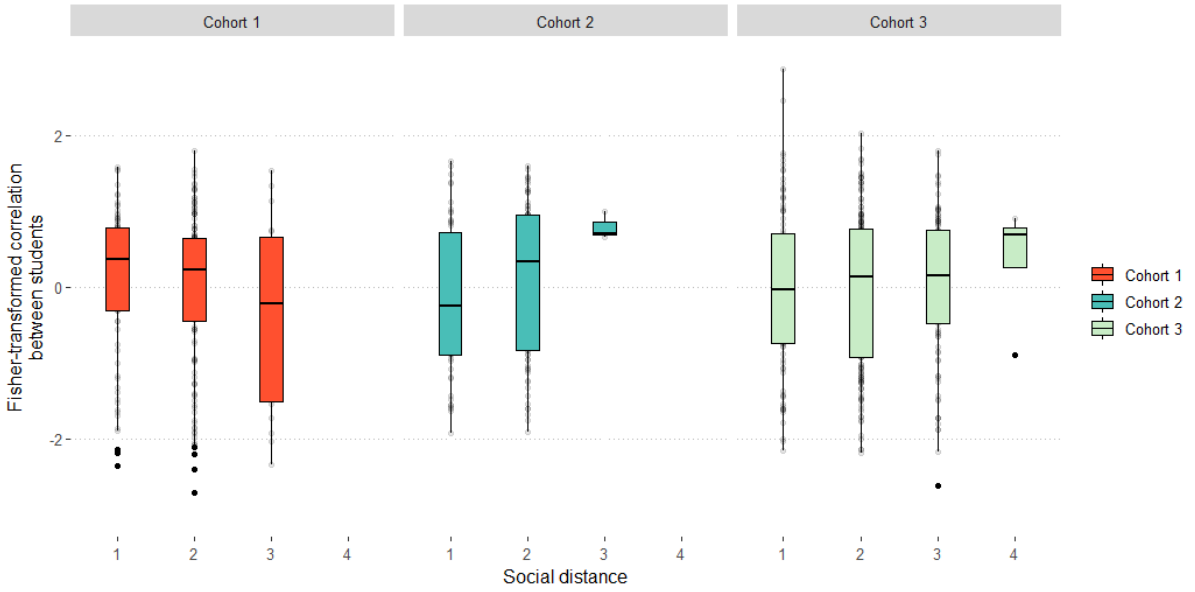

Figure 5. Example plot of standardised correlation strength of whole-brain connectivity between dyad members (y axis), as a function of social distance (x axis). Data here are for whole-brain connectivity similarities for all three cohorts; threshold was set at a distance of 4 (I spend “more than some” of my time with this person).

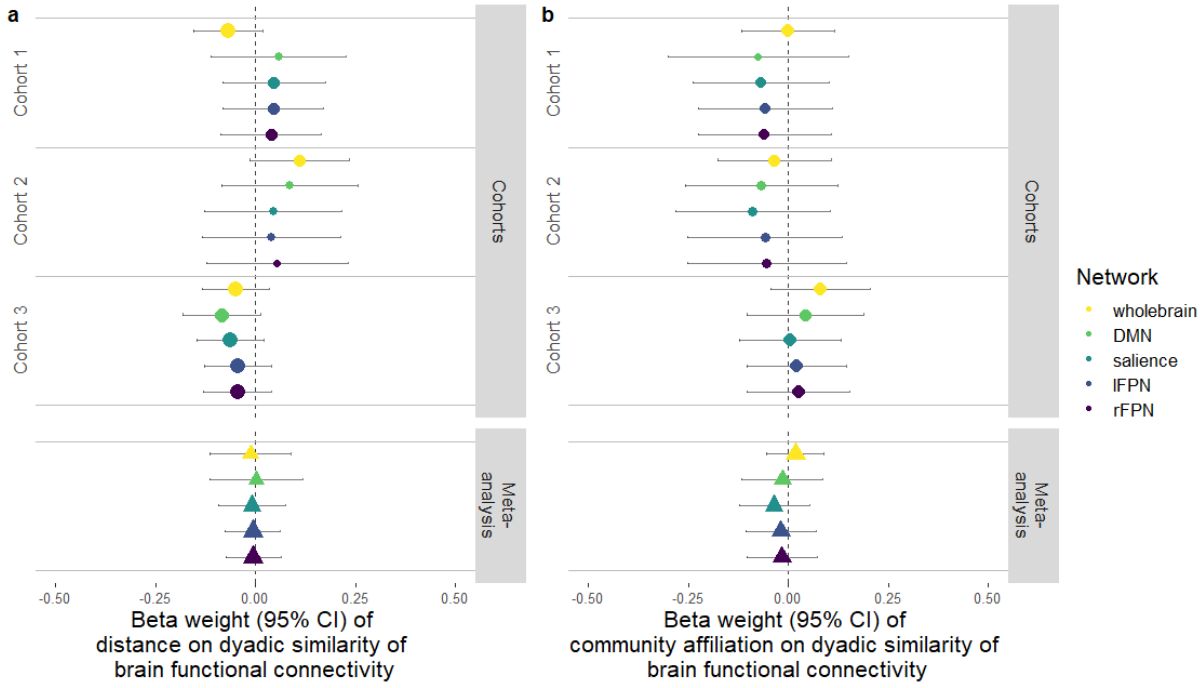

Figure 6. a) LME model outcomes and meta-analyses of functional connectivity similarity as a function of social distance. b) LME model outcomes and meta-analyses of functional connectivity similarity as a function of community affiliation. Whole-brain and RESTING STATE NETWORK data are shown for all three cohorts. Circles and bars represent beta weights (slopes) and 95% confidence intervals (95% CI), respectively, for individual cohorts. Triangles represent beta weights for meta-analyses. Relative confidence in the effect is represented by the size of the circle/triangle; colours represent brain networks.

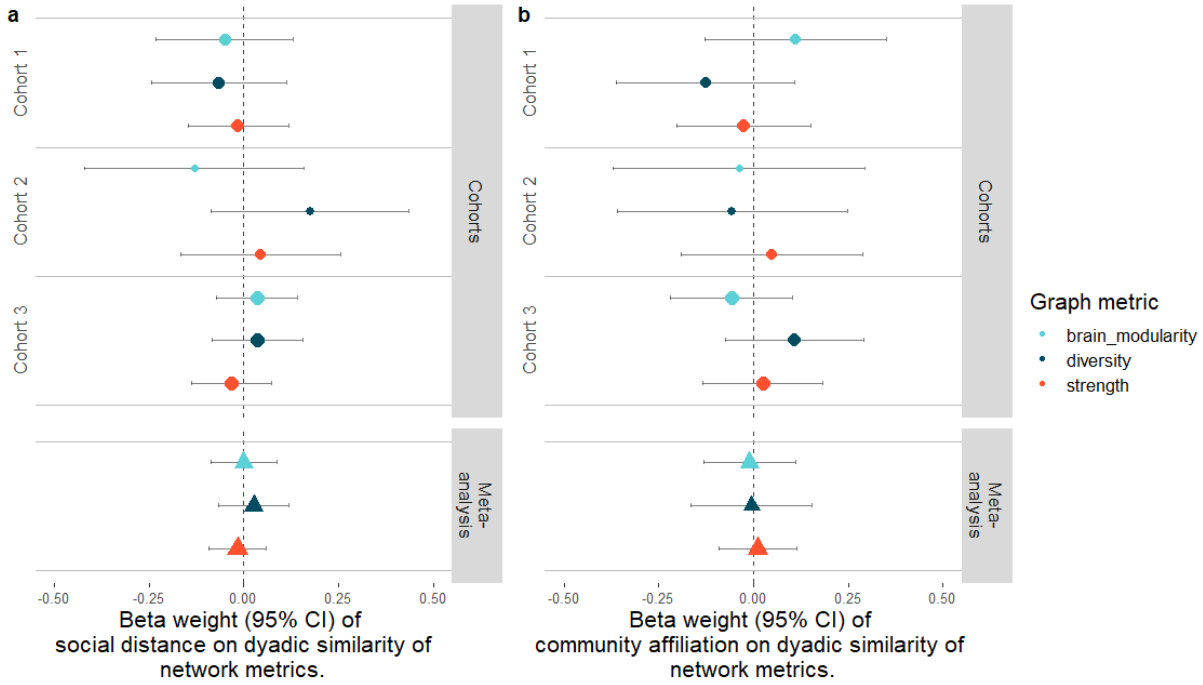

Figure 7. a) LME model outcomes and meta-analyses of similarities in graph theory measures of nodal strength, nodal diversity and brain modularity as a function of social distance. b) LME model outcomes and meta-analyses of similarities in graph theory measures of nodal strength, nodal diversity and brain modularity as a function of community affiliation. Data are shown for all three cohorts. Circles and bars represent beta weights (slopes) and 95% CI, respectively, for individual cohorts. Triangles represent beta weights for meta-analyses. Relative confidence in the effect is represented by the size of the circle/triangle; colours represent graph metrics.

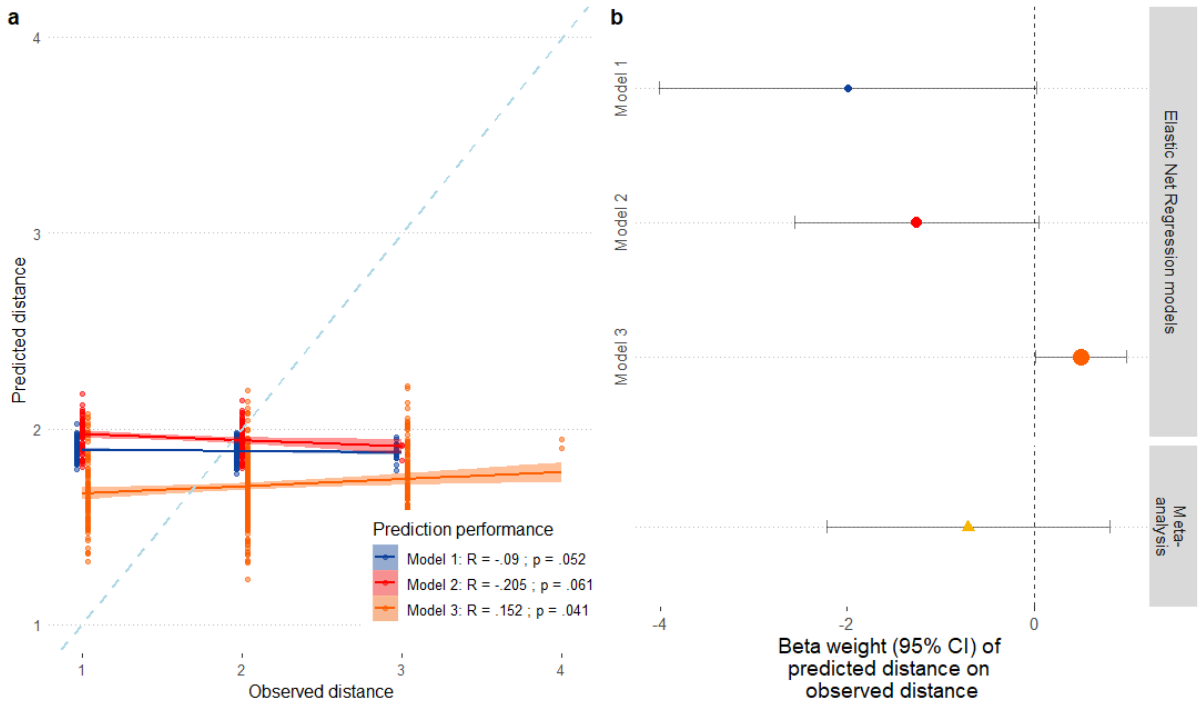

Figure 8. Performance of regression models predicting social distance from similarity in whole-brain functional connectivity. a) Predicted vs. observed social distance in testing data sets for models 1, 2 and 3. Blue dotted line represents the line of perfect accuracy (predicted = observed). Pearson's correlations were negative for models 1 and 2, indicating that the models performed worse than chance. P values are from LME with crossed random effects, taking into account clustering of dyad members. b) Meta-analysis of LME models for regression model of predictive performance.

Table 1. Training and testing sets for each predictive model. Each model was trained using data from two of the MRI cohorts and trained on unseen data from the third MRI cohort. Tuning parameters were determined separately for each model during the training phase.

	Model 1	Model 2	Model 3
Training set	Cohort 2-fMRI Cohort 3-fMRI	Cohort 1-fMRI Cohort 3-fMRI	Cohort 1-fMRI Cohort 2-fMRI
Testing set	Cohort 1-fMRI	Cohort 2-fMRI	Cohort 3-fMRI
Tuning Parameters:			
λ	209.085	176.565	.068
α	0	0	1

Table 2. Demographic data for full and fMRI cohorts.

Full cohort [fMRI cohort]	Cohort 1	Cohort 2	Cohort 3
N	59 [23]	51 [17]	65 [28]
Ethnicity (%)			
White	86.4 [91.3]	78.4 [76.5]	75.4 [75]
Asian	8.5 [4.3]	7.8 [5.9]	7.7 [3.6]
Black	3.4 [4.3]	7.8 [5.9]	9.2 [14.3]
Mixed	1.7 [0]	5.9 [11.8]	6.2 [7.1]
Other	0 [0]	0 [0]	1.5 [0]
Boarding status (% boarders)	27.1 [17.4]	33.3 [35.3]	41.5 [39.3]
Handedness (% left handed)	[4.4]	[5.9]	[3.6]

Table 3. Network characteristics of cohorts 1, 2 and 3, using threshold 4 social network data. Network diameter is the length of the longest geodesic distance between two nodes in the network. Modularity is a measure of how easily a network segregates into smaller subnetworks; large values represent networks that segregate easily into smaller communities. Mean path length is the mean geodesic distance between any two nodes in the network; smaller values are representative of more "tight-knit" networks. Reciprocity defines the proportion of connections in a directed graph that are mutual connections. Graph density gives the ratio of the number of connections (edges) and the number of possible connections in the network; higher values indicate that a larger number of possible connections have been made.

Network characteristics	Cohort 1	Cohort 2	Cohort 3
Network diameter	3	3	4
Modularity	0.304	0.286	0.426
Mean path length	1.769	1.750	2.033
Reciprocity	0.618	0.630	0.668
Graph density	0.285	0.320	0.220