NeuroImage 238 (2021) 118204

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Subject identification using edge-centric functional connectivity

Youngheun Jo[a,b], Joshua Faskowitz[a,b], Farnaz Zamani Esfahlani[a], Olaf Sporns[a,b,c,d], Richard F. Betzel[a,b,c,d,*]

[a] Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405, USA
[b] Cognitive Science Program, Indiana University, Bloomington, IN 47405, USA
[c] Program in Neuroscience, Indiana University, Bloomington, IN 47405, USA
[d] Network Science Institute, Indiana University, Bloomington, IN 47405, USA

A B S T R A C T

Group-level studies do not capture individual differences in network organization, an important prerequisite for understanding neural substrates shaping behavior and for developing interventions in clinical conditions. Recent studies have employed 'fingerprinting' analyses on functional connectivity to identify subjects' idiosyncratic features. Here, we develop a complementary approach based on an edge-centric model of functional connectivity, which focuses on the co-fluctuations of edges. We first show whole-brain edge functional connectivity (eFC) to be a robust substrate that improves identifiability over nodal FC (nFC) across different datasets and parcellations. Next, we characterize subjects' identifiability at different spatial scales, from single nodes to the level of functional systems and clusters using k-means clustering. Across spatial scales, we find that heteromodal brain regions exhibit consistently greater identifiability than unimodal, sensorimotor, and limbic regions. Lastly, we show that identifiability can be further improved by reconstructing eFC using specific subsets of its principal components. In summary, our results highlight the utility of the edge-centric network model for capturing meaningful subject-specific features and sets the stage for future investigations into individual differences using edge-centric models.

Introduction

Over the past several decades, the field of neuroimaging has leveraged powerful computational methods to develop standardized, population-level descriptions of brain morphology and functional organization (Glasser et al., 2016; Holmes et al., 1998). These approaches have facilitated group-level and cross-sectional comparisons of brains (Smith et al., 2015; Xia et al., 2018), enhancing our knowledge of the functional and neuroanatomical underpinnings of cognition (Cole et al., 2014; Crossley et al., 2013), brain development (Cao et al., 2014; Fair et al., 2009), and neuropsychiatric disease (Fornito et al., 2015). However, these efforts have emphasized group-level effects at the expense of individual brains, whose organization is personalized and idiosyncratic (Cui et al., 2020; Laumann et al., 2015; Poldrack et al., 2015).

Recently, several important studies have begun to shift focus away from group-level analyses and onto single subjects (Dubois and Adolphs, 2016; Finn et al., 2015). The aim of this line of research is to build comprehensive maps of individuals' brains (Gordon et al., 2017; 2018; Laumann et al., 2015; Poldrack et al., 2015), with the hope that inclusion of personalized details will add clarity to brain organization, brain-behavior relationships (Betzel et al., 2017; Huth et al., 2016; 2012; Mirchi et al., 2019; Seitzman et al., 2019), and inform treatment of neuropsychiatric disorders by helping design more efficacious and targeted interventions (Demeter et al., 2020; Gratton et al., 2019).

One particular strand of this research focuses on mapping the features of brain networks that are idiosyncratic to individuals. Like fingerprints, these features are capable of distinguishing one person's brain from that of another (Finn et al., 2015; Horien et al., 2019a). Brain network fingerprints serve as reliable substrates of individuals (Finn et al., 2015; Kliemann et al., 2019) that are stable over time (Gratton et al., 2018; Horien et al., 2019b; Jalbrzikowski et al., 2020a), across subsets of FC data (Byrge and Kennedy, 2019), and across acquisition sites (Bari et al., 2019). In addition, identifiable characteristics have shown clinical diagnostic potential (Svaldi et al., 2019) and have proven useful for the classification of individual behaviors and cognitive states (Salehi et al., 2020; Yoo et al., 2019).

To date, most fingerprinting analyses have focused on features derived from brain networks in which nodes represent positioned electrodes (Cox et al., 2018) or brain regions (Amico and Goñi, 2018b; Finn et al., 2015). Recently, we proposed an alternative model of brain connectivity that focuses on interactivity among a network's connections (or edges) (Betzel et al., 2021; Esfahlani et al., 2020; Faskowitz et al., 2020; Jo et al., 2020; Sporns et al., 2020). We refer to these patterns as edge functional connectivity (eFC). Adopting an edge-centric perspective has been fruitful in other scientific disciplines, e.g. in uncovering the overlapping community structure of complex biological and social networks (Ahn et al., 2010; Evans and Lambiotte, 2009). Recent studies have applied edge-centric methods in the form of "line graphs" (Evans and Lambiotte, 2009) to anatomical networks of interregional white matter

* Corresponding author at: Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405, USA.
E-mail address: rbetzel@indiana.edu (R.F. Betzel).

https://doi.org/10.1016/j.neuroimage.2021.118204.
Received 22 September 2020; Received in revised form 23 May 2021; Accepted 24 May 2021
Available online 1 June 2021.

tracts (de Reus et al., 2014). Similarly, eFC has provided a new window into studying the organization of brains, including the overlapping community structure (Faskowitz et al., 2020) and how participation of functional systems vary across communities (Jo et al., 2020). However, it remains unclear how eFC compares with traditional nodal functional connectivity (nFC) in terms of its ability to convey individual-specific information.

Here, we apply a novel edge-centric framework for investigating functional brain networks in subject identification. To calculate eFC, we use a procedure similar to that of calculating nFC. We start by standardizing (z-scoring) regional activity time series. nFC is usually computed as a Pearson correlation - the temporal average of the element-wise product of two standardized time series. Here, we omit the temporal averaging procedure, resulting in an "edge time series" (Esfahlani et al., 2020; Sporns et al., 2020a). The elements of an edge time series index instantaneous co-fluctuations between pairs of nodes (edges); they express positive values when their activity deflects in the same direction and negative when activity deflects oppositely. To calculate eFC, we compute the correlation between pairs of edge time series, resulting in an edge-by-edge matrix. Intuitively, if nFC reflects communication between two brain regions, the edge time series represent communication patterns across time, and eFC represents similarity between different patterns of communication (Faskowitz et al., 2020; Uddin, 2020).

In this study, we extend functional connectivity-based fingerprinting to edge-centric networks. Using functional imaging data from two independently acquired datasets (the Midnight Scan Club (Gordon et al., 2017) and the Human Connectome Project (Van Essen et al., 2013); MSC and HCP, respectively), we compare the performance of whole-brain nFC and eFC on subject identification, demonstrating that with sufficient amounts of data, eFC enables greater and more robust identifiability than nFC. Next, we investigated the system- and node-level drivers of the improved identifiability in eFC, focusing on system-specificity using a "leave-one-node-out" approach. We found nodes and edges associated with heteromodal brain systems to be the primary drivers of subject identification. Finally, we tested whether it was possible to optimize identifiability by reconstructing eFC and nFC using a restricted set of principal components. We found that reconstructed eFC significantly outperformed that of reconstructed nFC in terms of its optimized identifiability. Our work sets the stage for future studies to use eFC to develop network-based biomarkers for tracking inter-individual differences in behavior, development, and disease diagnosis.

Results

In this report, we systematically evaluate eFC and nFC in terms of differential identifiability and discuss their similarities and differences at regional (node) and subsystems (group of nodes) levels. Throughout this section, we analyze data from two high-quality independently acquired datasets: the Midnight Scan Club (MSC; (Gordon et al., 2017; Gratton et al., 2018)), which consists of ten participants scanned ten times each, and 100 unrelated subjects scanned two times each from the Human Connectome Project (HCP; (Van Essen et al., 2013)).

Identifiability using edge functional connectivity

Subject identification can be quantified using the measure "differential identifiability", or I_{diff}, which is calculated as the mean within-subject similarity minus the mean between-subject similarity of connectivity matrices (Amico and Goñi, 2018b). Existing subject identification applications have relied on connectivity patterns derived from nFC and thus the idiosyncratic characteristics of eFC remains unknown. In this section we compare the identifiability of cortex-wide nFC and eFC and its dependence on the amount of data available.

First, we compared cortex-wide eFC and nFC in terms of subject identifiability. Briefly, this entailed estimating nFC and eFC separately for each of the 100 resting-state fMRI scans (10 subjects; 10 scans each) in the Midnight Scan Club dataset, and generating similarity matrices for each connectivity modality as the Pearson correlation between the upper triangle elements of subjects' nFC or eFC matrices (Fig. 1*e*). We then estimated differential identifiability from these similarity matrices (Fig. 1*f*).

We found that eFC outperformed nFC, yielding greater values of I_{diff} when using functional networks derived from the entire cerebral cortex (Fig. 2*a*). Similar results were also found when using different parcellations and in an independent dataset (Fig. S1). To visualize the within- and between-subject similarity, we used multi-dimensional scaling to project subjects and their scans into a two-dimensional space that approximately preserves the pairwise distance relationships encoded in the nFC and eFC similarity matrices (Fig. 2*c-d*). Therefore, the distance between two scans on the MDS plot indicates the similarity measured *via* non-linear dimensionality reduction. We calculated the accuracy of each scan's subject label against it's K-nearest neighbors' subject labels using standardized Euclidean distance. We found no significant differences between mean accuracies using nFC and eFC ($t-test$, $p = 0.9786$). The subject label accuracy was calculated as the fraction of subjects in the nine nearest points in the MDS space that match the subject label of a scan.

In the previous analysis, we calculated eFC and nFC using approximately 30 minutes of data (the duration of scan session in the MSC dataset). Next, we tested whether subject identifiability was modulated by scan length, i.e. whether the value of I_{diff} varied as a function of the amount of data available (Amico and Goñi, 2018a; Bari et al., 2019). To test this, we created shorter or longer "sessions" by either dividing the existing runs into shorter, contiguous segments, or by concatenating data from multiple scans to form longer sessions. We varied the duration of artificial scan sessions in increments of 100 samples, starting with 100 and ending with 4000. This entire sampling procedure was repeated 100 times. We found that with fewer than 500 time points (approximately 20 minutes) I_{diff} was greater for nFC than eFC ($p < 10^{-6}$, $t-test$); Fig. 2*a*. However, at 800 time points (approximately 30 minutes), eFC began to significantly outperform nFC ($p < 10^{-4}$; $t-test$); Fig. 2*b*. We report similar results using different parcellations and datasets (Fig. S1). Our results using eFC are in line with previous research where identifiability was increased with extended scan length using conventional nFC (Amico and Goñi, 2018b; Bari et al., 2019). Collectively, our findings indicate that, given sufficient amounts of data (approximately 30 minutes), eFC enables a more robust identification of subjects across sessions than nFC.

Regional drivers of cortex-wide eFC identifiability

In the previous section, we found that eFC improved cortex-wide identifiability over nFC given a sufficient number of samples. In this section, we wanted to pinpoint the brain regions that contributed to this improvement. To do this, we use a "leave-one-node-out" method to measure the relative impact that each brain region had on subject identification. We then summarized these results by grouping nodes according to canonical brain systems and assessing, statistically, the contribution of each system to the overall identifiability. eFC measures interactions between pairs of edges, each of which corresponds to a pair of brain regions. To determine which brain regions drive these effects, we iteratively removed each of the 200 brain regions and recalculated eFC and I_{diff} using the remaining 199 regions. We then compared this I_{diff} value with the the value obtained using the intact brain (all 200 regions). Here, regions were defined based on a functional atlas (Schaefer et al., 2017), where resting-state nFC was used to create functionally homogeneous parcels of approximately equal sized. Information about the parcellation and atlas can be found here: https://github.com/ThomasYeoLab/CBIG/tree/master/stable_projects/brain_parcellation/Schaefer2018_LocalGlobal. We also performed a similar analysis using edges instead of nodes, the results of which are reported in Fig. S2. We found that, when removed, cortical areas located in the control and temporoparietal network regions yielded

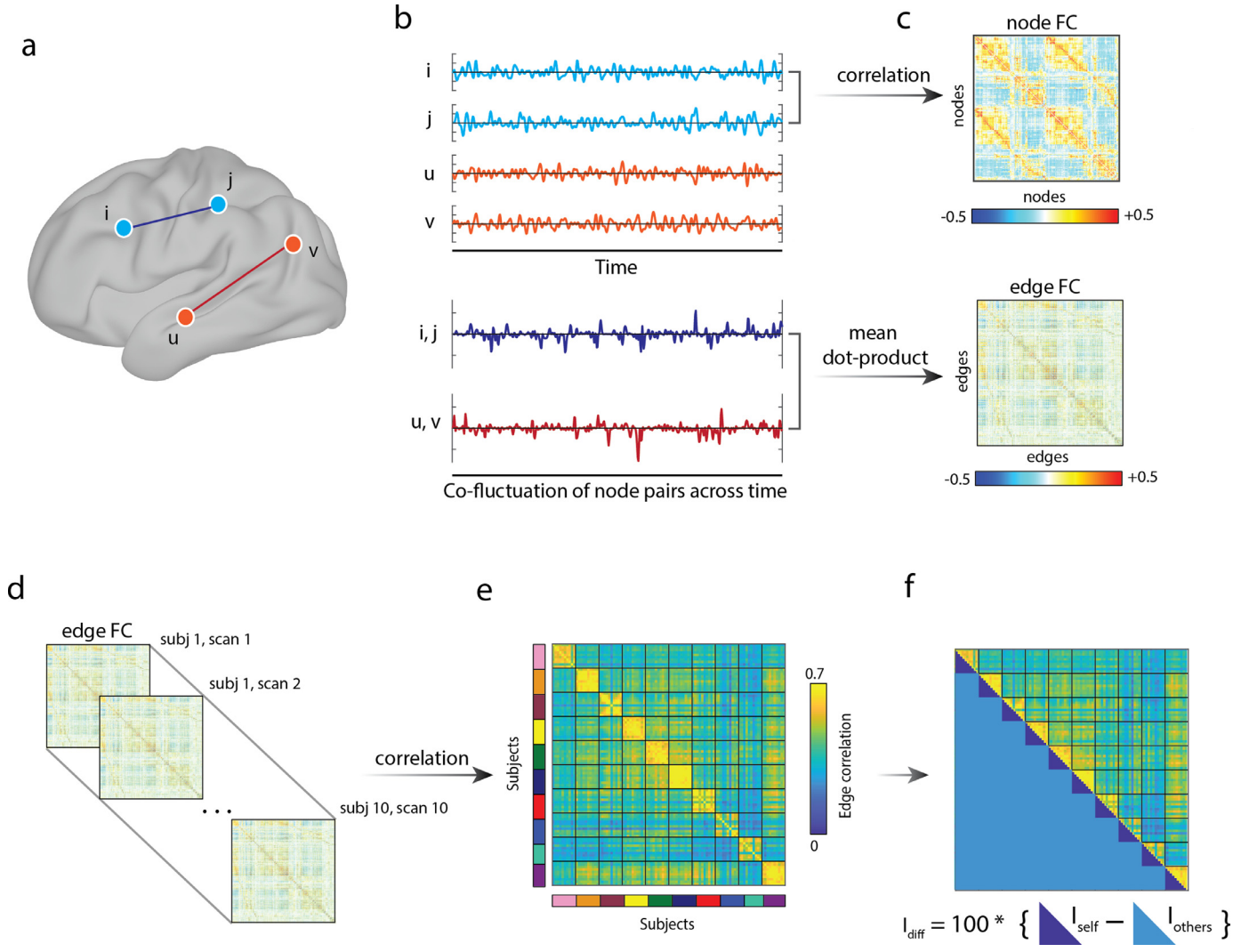

Fig. 1. Schematic of differential identifiability in eFC. All panels from this figure were generated using data from the Midnight Scan Club dataset. (*a*) To illustrate the calculation of node and edge FC (nFC and eFC, respectively), consider four nodes: *i*, *j*, *u*, and *v*. nFC is defined as the pairwise correlation of regional activity. For nodes *i* and *j*, nFC is calculated by first z-scoring each nodes' time series, computing the element-wise product, and averaging these values (*b* and *c; top*). The same operation could be carried out for nodes *u* and *v*. eFC is calculated by first generating co-fluctuation time series for pairs of nodes. This involves computing their element-wise product, but omitting the averaging step (*b; bottom*). Each co-fluctuation time series is defined for a pair of nodes. eFC is calculated as the temporal similarity (e.g. correlation, cosine similarity, etc.) of pairs of co-fluctuation time series (*c; bottom*). (*d*) To calculate differential identifiability, (*e*) we extract the upper triangle elements of subjects' eFC matrices and compute the spatial correlation of those elements, resulting in subject-by-subject similarity matrix. (*f*) Differential identifiability, I_{diff} is calculated as the mean within-subject similarity minus the mean between-subject similarity.

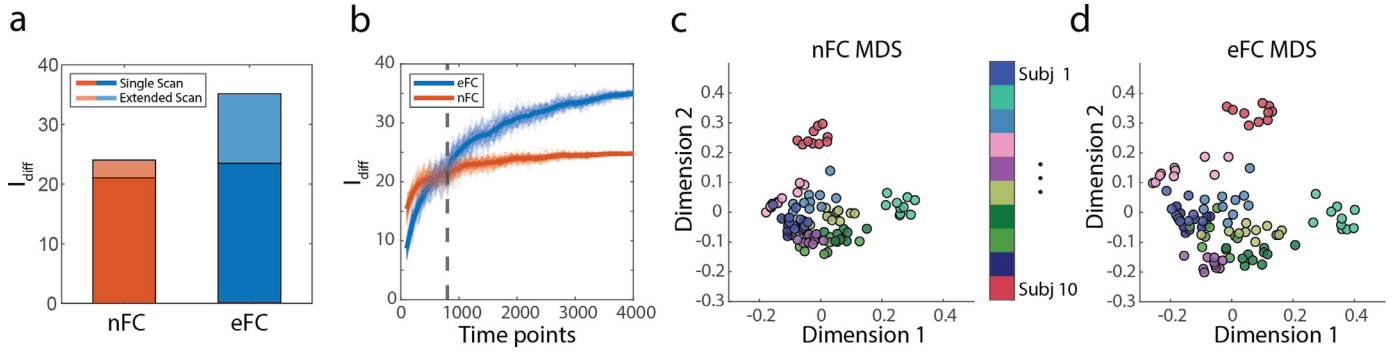

Fig. 2. Subject identification of eFC and nFC and effect of scan length. All panels from this figure were generated using data from the Midnight Scan Club dataset with 200 node Schaefer parcellations. (*a*) I_{diff} of nFC and eFC for single scans (darker color) and for maximally concatenated scans (lighter color). (*b*) I_{diff} of eFC and nFC by scan length. The black dotted line (800 time points) indicates the length of the time series in which eFC significantly outperforms nFC. Thick blue and orange lines are the average I_{diff}s for 100 iterations and thin lines indicate each iteration's I_{diff} by scan length. (*c*) Subjects' scans plotted using multi-dimensional scaling for nFC. (*d*) Subjects' scans plotted using multi-dimensional scaling for eFC. Each subject's scan corresponds to a color on the colorbar between panels *c* - *d*.

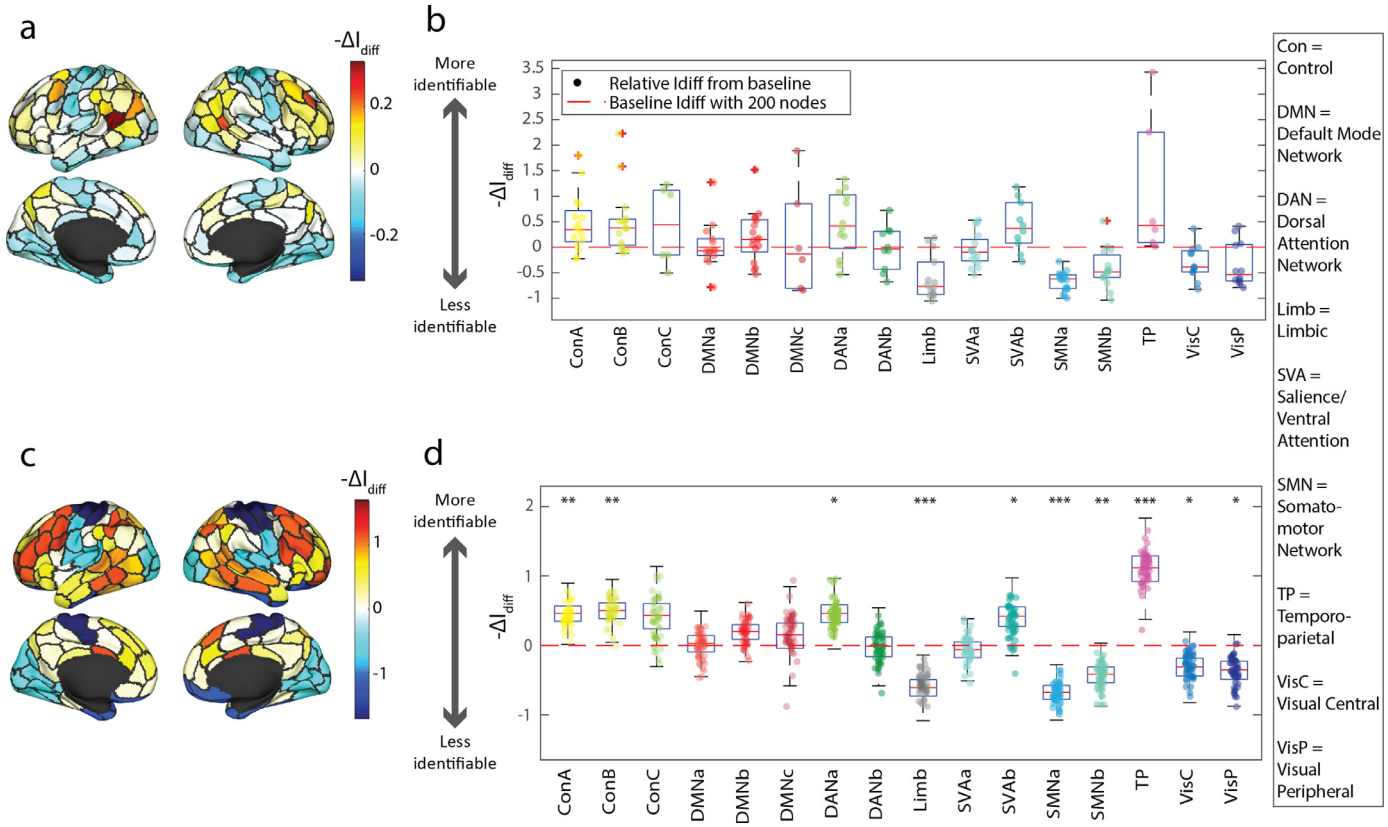

Fig. 3. Nodal contribution to identifiability in eFC. (*a*) To assess regional contributions to I_{diff}, we calculated the change in I_{diff} after removing all edges incident upon each of the $N = 200$ nodes. Here, we show the change in I_{diff} projected onto the cortical surface. (*b*) We plotted the relative impact of each node's removal prior to eFC calculation on I_{diff} by brain systems. (*c*) Excluding entire system's nodes prior to eFC construction reveals that specific systems contribute positively or negatively to I_{diff} and that (*d*) compared to random removal of matching numbers of nodes, have significantly different results on I_{diff} (Bonferroni-corrected *p*-values; $^{*} = p < 0.003$, $^{**} = p < 0.0006$, $^{***} = p < 0.00006$). For visualization we inverted the ΔI_{diff} to $-\Delta I_{diff}$, which quantifies the relative effect of a node's removal on the overall I_{diff}. If a removal of a node results in reduced I_{diff}, this node's contribution to identifiability would be considered as 'increasing I_{diff}'.

decreased differential identifiability while removal of regions associated with somatomotor, limbic, and visual networks increased identifiability (Fig. 3*a*; note that for the sake of visualization, we invert the sign of ΔI_{diff}). All nodes were also grouped according to their systems visualizing each node's contribution to subject identifiability (Fig. 3*b*). Each node's contribution was measured as the I_{diff} without a particular node minus the I_{diff} calculated from all 200 nodes. For example, in Fig. 3*b*, removing a single node from the control A system (represented as a single yellow dot) prior to eFC construction would result in a reduction in I_{diff}, which we plot here as having a "positive" effect on I_{diff}. In contrast, removing a single node from the limbic system yields an increase in I_{diff}, hence we plot it as a "negative" effect on I_{diff}.

Next, we tested which system's nodes significantly influenced identifiability compared to random removal of matching numbers of nodes (Fig. 3*c* - *d*). Nodes were randomly reassigned to systems by randomly permuting system labels (10,000 iterations). We found that when removing nodes from the control A, control B, dorsal attention A, salience ventral network B, and temporoparietal networks significantly decrease identifiability than when randomly removing matching numbers of nodes (thus, the relatively positive effect of these systems in panel Fig. 3*d*). Also, we found that excluding nodes from limbic, somatomotor A and B, and central and peripheral visual networks significantly increase identifiability compared to randomly removing matching numbers of nodes (thus, the relative negative effect of these networks to I_{diff} in panel Fig. 3*d*).

In summary, we used a *leave-one-node-out* approach to uncover the regional drivers of whole-brain identifiability. We found that the inclusion of frontoparietal and superior temporal regions help increase identifiability, while somatomotor, limbic, and visual regions lead to reductions in identifiability. These results on system-level identifiability largely agree with prior research using conventional, node-centric functional connectivity (Finn et al., 2015; Mueller et al., 2013; Peña-Gómez et al., 2018), localizing idiosyncrasies of brain network organization to a specific subset of systems.

Identifiability of systems and clusters in eFC

In the previous sections, we demonstrated that given an fMRI scan of sufficient duration eFC outperforms nFC in subject identification and that heteromodal brain regions compared to unimodal, contribute to higher identifiability. Here, we continued our investigation into the drivers of cortex-wide identifiability, focusing on the contributions of each functional brain system in eFC to identifiability. In this section, we aim to answer the questions: How do edges from single systems contribute to identifiability? How do clusters of edges perform in identifiability?

To address these questions, we first estimated I_{diff} using only connections associated with specific brain systems. In the case of nFC, this means calculating identifiability using only edges whose stub nodes, the two brain regions that constitute the end of an edge, are assigned to the same brain system (Schaefer et al., 2017). We performed a similar operation using eFC's edge pairs subjected to the requirement that all four nodes associated with the two edges that comprise an eFC entry were assigned to the same system. In general, we found that system-specific I_{diff} for eFC and nFC was highly correlated ($R = 0.9578$, $p < 10^{-8}$; Fig. 4*b*). We report the top and bottom five systems ordered by their

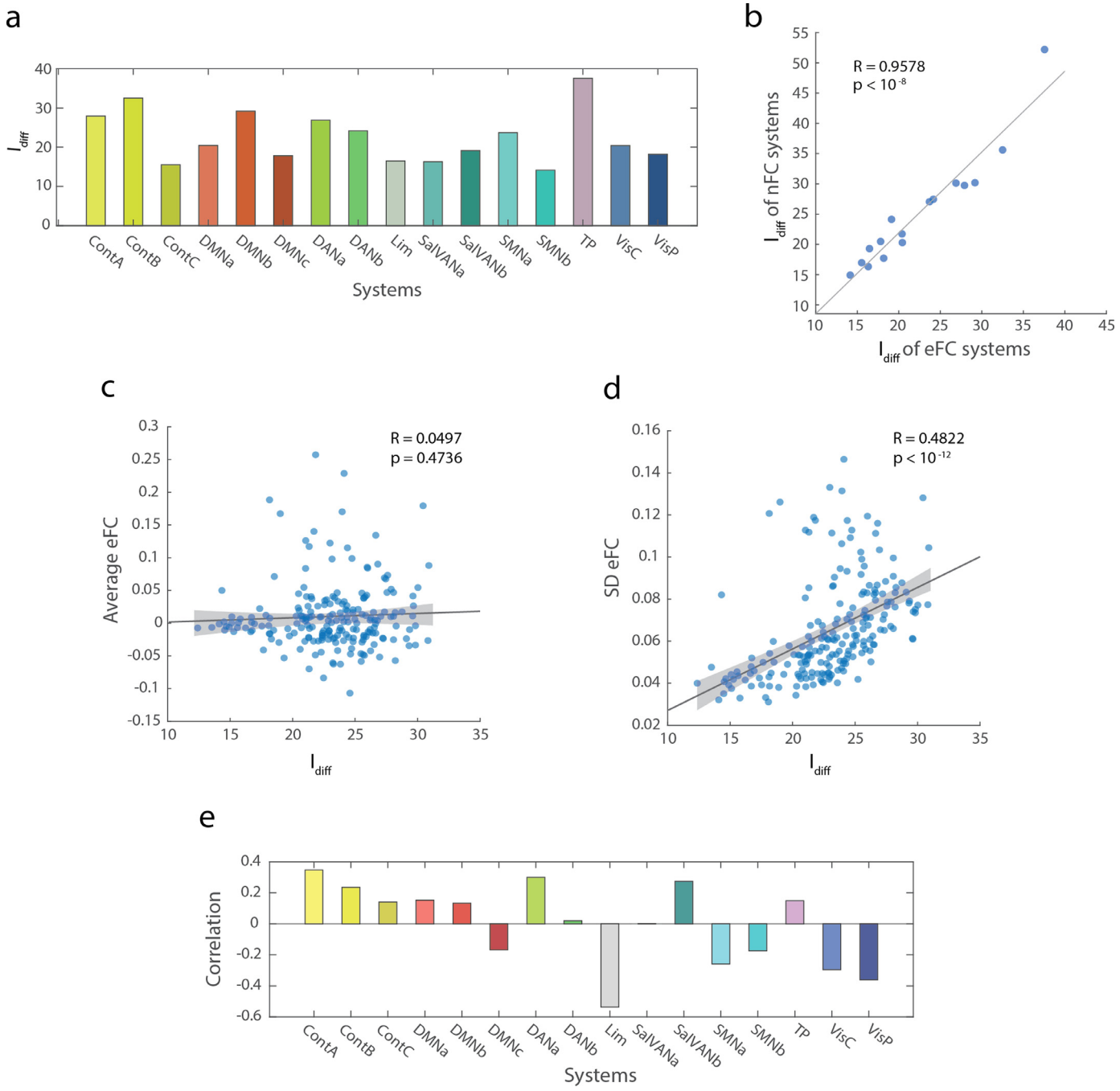

Fig. 4. System- and cluster-level characteristics of edge functional connectivity (eFC). The within- versus between-subject similarity of eFC in the MSC dataset shows variance of identifiability across systems and high correlation to the identifiability of systems using FC. Panel (*a*) shows the within subject minus between subject similarity of eFC for systems for 10 rsfMRI scans from 10 subjects. Panel (*b*) shows the high correlation between identifiability in eFC and nFC by systems. (*c*) I_{diff} was not correlated with the average values of eFC in each cluster but in (*d*) the standard deviation of eFC values of eFC in each cluster significantly correlated with I_{diff}. The bootstrapped 95% confidence interval is shaded in grey. (*e*) The correlation of 16 canonical brain systems to I_{diff} of eFC clusters. All analyses were performed using the MSC 200 node parcellation.

I_{diff}. We found that when both nodes originated in temporoparietal, control A, control B, default mode network B, and dorsal attention network B, I_{diff} tended to be larger (Fig. 4*a*). In contrast, we found that when both nodes came from somatomotor B, salience ventral attention A, control C, visual peripheral, and limbic networks, I_{diff} tended to be smaller (Fig. 4*a*). We note that these single-system edge pairs (four nodes originating from a single system) represent only a small fraction (< 1%) of the eFC matrix and therefore, these results are limited to the subset of matrix elements. Since the eFC incorporates a broader repertoire of edge pairs, some of which involve nodes originating in up to four distinct systems, we further analyze eFC at its cluster-level.

To investigate I_{diff} of edge pairs whose nodes originate in different systems or clusters, we clustered the eFC matrices using a standard k-means algorithm using Euclidean distance, varying the number of clusters from $k = 2$ to 20. To do so, we first performed an eigendecomposition of a group-averaged eFC matrix (the mean across all scans and subjects; Fig. S4*a-b*). We retained the first 50 components, which were then used as input to the k-means clustering algorithm (Fig. S4*c*). For a given

number of clusters, k, we repeated the algorithm 250 times (Fig. S4*d*). From these 250 estimates, we retained a single representative partition as the one that was, on average, most similar to the others (Fig. S4*e*). The result of this procedure was a partition of edges into non-overlapping communities for every k. Note that this clustering procedure was carried out using the group-average eFC matrix, which allows the cluster correspondence across subjects.

We then related properties of the eFC matrix to identifiability. Specifically, we used the group-representative clusters to partition edge-edge connections into "blocks" that fall between pairs of communities (Fig. S4*f*). Then, for each block we calculated the mean and variance of those connections from all scans by expressing those weights as vectors. Next, we computed the similarity between all pairs of subjects and scans using these weights, and from these similarity values calculated the identifiability of that block. These procedures left us with three measures for each block: 1) The mean weight of edge-edge connections of each scan's matrix, 2) The standard deviation of edge-edge connections of each scan's matrix, 3) Identifiability using matrices using the Pearson correlation of edge-edge connections across all scans. Finally, for each of the k clusters, we calculated the mean of these three values over all blocks in which that cluster participated (effectively the rows or columns of the block matrix). We then plotted the relationship of identifiability with the mean and standard deviation values in Fig. 4*c-d*. At each k, we calculated the I_{diff} for the constituent "blocks" of clusters (Fig. 4).

The average value of the mean and standard deviation of I_{diff} for all blocks involving each cluster was calculated as the representative value for each cluster. This process resulted in a total of 209 points ($k = 2 - 20$) in Fig. 4*c-d*. Specifically, we found that the mean eFC of a cluster was not correlated with its I_{diff} ($R = 0.0497$, $p = 0.4736$; Fig. 4*c*). On the other hand, the mean variability of eFC weights associated with a given cluster was positively correlated with I_{diff} ($R = 0.4822$, $p < 10^{-12}$; Fig. 4*d*). These blocks refer to the edge–edge weights within a cluster when the eFC is reordered by the cluster labels (Fig. S4*f*).

Every block corresponds to a set of edge-edge connections (eFC) between edges assigned to different clusters. For each block, we calculated the mean and standard deviation of edge-edge connections, which summarizes the central tendency and variability of the distribution. We found that the mean weight of edge-edge connections did not have a significant relationship with identifiability. However, the variability of edge-edge connections within a block was significantly correlated with higher subject identifiability. We also add the results from clusters of nFC ($k = 2$ to 20) and their relationship with I_{diff} in Fig. S5. Unlike the results derived using eFC (mean: $R = 0.0497$, $p = 0.4736$; standard deviation: $R = 0.4822$, $p < 10^{-}12$), we found only weak correlation patterns when applied to nFC (mean: $R = 0.1392$, $p = 0.0444$; standard deviation: $R = 0.1494$, $p = 0.0309$).

What systems might be responsible for driving high levels of I_{diff}? To address this question, we calculated how frequently each system was represented within a given cluster, and, separately for each brain system, calculated the correlation of this frequency with I_{diff}. We found that the presence of control, default mode A and B, dorsal attention, salience ventral attention, and temporoparietal network nodes in a cluster is positively correlated with the cluster's I_{diff}. In contrast, the presence of nodes from limbic, default mode C, and sensorimotor systems (somatomotor and visual) networks in clulsters is associated with reduced I_{diff} (Fig. 4*e*).

Collectively, these results suggest that in eFC, higher order cognitive systems, e.g. control, attention, and default mode networks, contribute to enhance subject identifiability, while sensorimotor and limbic networks reduce identifiability. As in the previous section, these results are in line with previous analyses using nFC demonstrating that similar systems and regions promote enhanced identifiability (Finn et al., 2015; Mueller et al., 2013). Furthermore, our results suggest that the intrinsic heterogeneity and variability of connection weights may be an underlying factor explaining why certain systems are associated with higher or lower levels of identifiability.

Reconstructed eFC using PCA improves subject identifiability

Throughout this manuscript, we have focused on calculating I_{diff} using the full eFC matrix or specific subsets of its edge-edge connections. As a final analysis, we wanted to test whether we could improve differential identifiability by optimally reconstructing eFC using a relatively small number of its principal components.

Previous research has used principal component analysis (PCA) of nFC to enhance identifiability (Amico and Goñi, 2018b; Bari et al., 2019). Briefly, this procedure entails concatenating nFC (or in this case, eFC) from all subjects and scans into a single matrix of scans-by-elements, decomposing this matrix into its principal components (PCs), and reconstructing eFC by gradually including more and more of its PCs (in descending order of their eigenvalues). Prior-to and after each reconstruction, we calculated the I_{diff} with added PCs for each scan's eFC. Here, we apply this technique to both nFC and eFC from the MSC dataset.

Using this reconstruction method, we found that I_{diff} could be improved for both nFC and eFC. In both nFC and eFC, I_{diff} peaked at $k = 10$ components (corresponding to the number of subjects) and eFC outperformed nFC in subject identification (peak value of $I_{diff} = 35.27$ compared to $I_{diff} = 21.17$; Fig. 5*b,e*). These results were also found in test-retest comparisons (Fig. S8). Why, then, do the number of PCs match the number of subjects when optimizing for identifiability? First, we tested the effect of scans per subject on the number of PCs for maximizing identifiability. When testing for two scans per subject with 100 subjects Fig. S7*c* and with randomly selected two scans per subject with 10 subjects Fig. S7*a*, the number of PCs required for optimizing identifiability matched the number of subjects in the dataset. These results were replicated when we tested on the entire HCP dataset (N = 100, two scans per subject; Fig. S7*e*) and when matching scan lengths of the HCP dataset (2400 time points) to the MSC dataset (800 time points; Fig. S7*f*).

Next, we investigated the PC coefficients that improve (PC = 1 − 10) or reduce (PC = 11 − 100) identifiability. The first principal component (PC1), mathematically, explains the largest variance of eFC values across scans and subjects. PC1 was the only component whose coefficients were uniformly positive (Fig. S9). The next nine coefficients (PC = 2 − 10) expressed "blocky" patterns that correspond to single subjects (Fig. S9), while this pattern was absent in PC = 11 − 100 (Fig. S10, Fig. S11, Fig. S12).

In agreement with previous reports (Amico and Goñi, 2018b; Rajapandian et al., 2020), our results demonstrate that subject identification can be improved by selectively retaining a subset of components that match the number of subjects in the dataset. We show that the magnitude of improvement is considerably greater using eFC compared to nFC, suggesting that eFC may better capture personalized and idiosyncratic features compared to nFC (Svaldi et al., 2019; 2018).

Discussion

Here, we applied subject identification to a novel, edge-centric network representation of the human cerebral cortex. We found that given sufficient scan length, eFC exhibits greater levels of differential identifiability than nFC, an improvement that we linked to contributions made by brain regions in association cortex. Finally, we used a dimension-reduction and reconstruction method to show that the relative improvement in identifiability enjoyed by eFC could be further enhanced, highlighting the potential for eFC to be used in future studies.

Edge functional connectivity enhances subject identifiability

Central to this paper is the observation that eFC results in improved subject identification relative to conventional node-based connectivity, nFC. Whereas nFC measures the similarity of activity between two brain

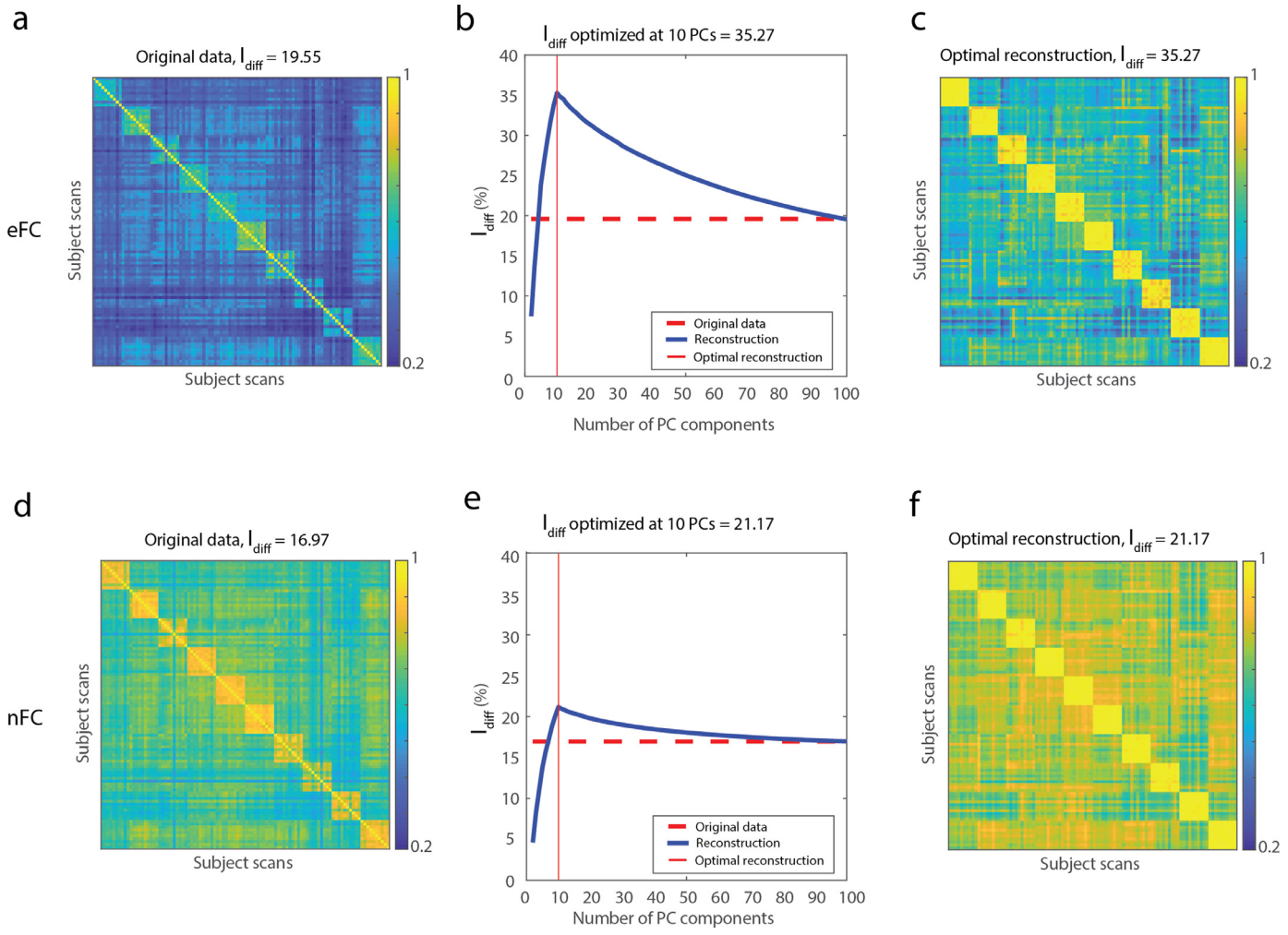

Fig. 5. PCA reconstruction for differential identifiability optimization with eFC and FC (*a*) Scan-to-scan similarity (Pearson correlation) matrix of eFC prior to PCA (original data). (*b*) Optimal I_{diff} measured when PC = 10; red dotted line = I_{diff} orig. (*c*) Reconstructed output eFC matrix for maximal I_{diff} using 10 PCs. (*d*) Scan-to-scan similarity (Pearson correlation) matrix of nFC prior to PCA (original data). (*e*) Optimal I_{diff} measured when PC = 10, red dotted line = I_{diff} orig. (*f*) Reconstructed output nFC matrix for maximal I_{diff} using 10 PCs.

regions – a first-order correlation (Owen et al., 2019) – eFC measures the similarity of co-fluctuations between edge pairs – a higher-order correlation (Faskowitz et al., 2020). Understanding the higher-order organization of networks has proven useful in other disciplines (Ahn et al., 2010; Evans and Lambiotte, 2009; Nepusz and Vicsek, 2012; Trinh and Kwon, 2016). Here, adopting an edge-centric perspective allows us to link higher-order brain network organization with subject specific features.

We asked whether the eFC – a higher-order reconstruction – leads to improved identification of subjects. That is, if we were to examine the identifiability of whole brain nFC and eFC, would eFC allow us to more accurately identify a single subject based on their connectivity data? We also asked whether identification using eFC would be impacted by scan length and amount of data. In general, we found that given sufficient amounts of data (approximately 800 time points or 30 min), eFC outperformed nFC in terms of identifiability. Also, we note that I_{diff} using eFC plateaus at a greater I_{diff} with concatenated scans. This result may reflect the higher-order nature of eFC. Effectively, eFC measures correlations of correlations and requires more data than nFC to obtain a stable estimate (Bourin and Bondon, 1998). Our results suggest that as this estimate stabilizes, eFC may offer advantages over nFC in terms of the extent to which it encodes individualized and idiosyncratic features.

Next, we adopted the measure differential identifiability (Amico and Goñi, 2018b) which is a summary metric of within-subject scan similarity compared to between-subject scan similarities. While other self-similarity metrics such as the Pearson correlation (Finn et al., 2015) and I2C2 Shou et al. (2013) have robustly identified subjects, I_{diff} also takes into consideration the between-subject scan similarities (as does Discriminability (Bridgeford et al., 2020)). Therefore, a scan that is highly identifiable will not only be self-similar, but also be different from other subjects' scans. We note that based on how I_{diff} is calculated, I_{diff} may be biased by the number of scans per subject relative to the size of the cohort. This renders direct comparisons and interpretations of I_{diff} values across different datasets difficult. We add that benchmark tests across I2C2, Discriminability, and I_{diff} in the MSC 100 and HCP 100 node datasets have shown I_{diff} to show the highest subject identification (Fig. S13).

Our results were replicated across two datasets and two different parcellations. These results suggest that higher-order network structure carries important subject-specific information and that idiosyncratic features of networks are hence better captured *via* an edge-centric viewpoint. These observations both challenge and extend current knowledge on subject identification and precision network mapping (Amico and Goñi, 2018b; Finn et al., 2015; Gordon et al., 2017; Gratton et al., 2018; Svaldi et al., 2018). Our results further advocate for deep phe-

notyping that requires more data from individuals. Future investigation is required for understanding the relationship between scan length (amount of data) and higher-order brain representations, such as eFC (Owen et al., 2019).

Heteromodal regions drive subject identification

Which brain regions drive the subject specificity of eFC? Are there particular regions of the brain that make a subject more or less identifiable? To answer these questions, we analyzed subject identifiability on three different scales. First, we analyzed each node's contribution to I_{diff} using a "leave-one-node-out" approach. We found that when removing certain nodes prior to constructing the eFC matrices, resulted in significantly reduced or increased I_{diff}. In particular, nodes and edges from higher-order systems lead to significant reductions in I_{diff}, while those associated with sensorimotor and limbic regions yielded increased identifiability when removed prior to eFC construction. These findings support earlier results with nFC in which brain regions of higher-order, heteromodal function - including frontal, parietal, and temporal areas - drive subject identification compared to unimodal, primary sensory brain regions (Finn et al., 2015).

Why might heteromodal regions help enhance identification? One possible explanation has to do with their polyfunctionality. Regions in control and temporoparietal networks are thought to support a wide variety of cognitive and psychological operations which are known to vary across individuals. These same regions have been found to be highly variable across subjects in terms of their functional connectivity patterns (Mueller et al., 2013), the most evolutionarily recent parts of cortex to have expanded (Sepulcre et al., 2010), and are among those that develop latest in life and therefore likely to have been shaped by early-life experiences (Elston et al., 2009). Combined, the observation that heteromodal brain regions drive individuals' functional fingerprints is likely a combined effect of cognitive, developmental, and evolutionary processes. The precise mechanisms underlying these processes, however, are unclear and should be the target of future investigations.

We then asked whether higher and lower I_{diff} is attributable to edge pairs from a single canonical brain system (Schaefer et al., 2017). To address this question, we estimated subjects' eFC matrices separately from edge pairs constructed using only one of sixteen brain systems. We found that identifiability from single-system eFC and nFC were significantly positively correlated and that edges forming stubs from heteromodal brain regions tended to have higher I_{diff}. In other words, the cohesive edge pairs within single systems are not likely to be driving eFC's improved I_{diff} which is observed globally. Rather, it suggests that edges falling between different brain networks may be driving the improved I_{diff} in eFC compared to that of nFC.

Results from the "leave-one-node-out" and single-system I_{diff} suggest specific brain regions, namely those involved in higher-order brain function, drive subject identification. These higher-order association regions have been found to show the highest inter-subject variance in functional connectivity studies (Miranda-Dominguez et al., 2014; Mueller et al., 2013). The functional connectivity of associative brain regions has been found to be evolutionarily recent (Zilles et al., 1988), involved in intelligence (Choi et al., 2008; Cole et al., 2012), and when disturbed, lead to neuropsychiatric disorders (Fornito and Harrison, 2012; Greicius, 2008). Also, the identifiability of individuals in these higher-order brain regions using nFC has reported to be predictive of fluid intelligence (Amico and Goñi, 2018b; Finn et al., 2015). Combined, our results show that subject identifiability and higher-order functionalities both originate from heteromodal brain regions suggesting idiosyncrasy may arise from such brain functions.

The number of vertices and connections in eFC and nFC matrices diverge by orders of magnitude. To match their dimensionality and ensure a fairer comparison, we clustered nFC and eFC matrices into the same number of communities. We found that the variance of edge–edge connections was positively correlated with subject identification, while the average weight of those connections was not. However, we found relatively weak relationships using group-averaged nFC matrices. Thus, our results further suggest that variability of edge–edge connections in eFC are an important feature for subject identifiability. We also found that clusters containing edges whose stubs originated in heteromodal systems resulted in greater levels of identifiability compared to clusters composed of edges associated with unimodal brain systems. These results are in agreement with findings from nFC studies (Finn et al., 2015; Peña-Gómez et al., 2018) and with more recent studies involving clusters of eFC (Jo et al., 2020). Our findings suggest that variability of connection weights within systems and clusters may be an important feature driving the efficacy of connectome fingerprinting and identifiability.

For our analyses, we used the Yeo 17 functional brain networks (Thomas Yeo et al., 2011) from the 100 and 200 node Schaefer parcellations (Schaefer et al., 2017). While the Schaefer atlas provides an incrementally scalable parcellation from 100 to 1000 nodes, due to computational limitations, the differential identifiability and it's subsequent analyses were not performed on the finer Schaefer parcellations (300, 400,..., 1000). Based on the results in Fig. S1, in both MSC and HCP datasets, eFC was found more identifiable over nFC in the 100 and 200 node parcellations. However, the absolute difference in identifiability diminished as the number of nodes increased from 100 to 200. Further analysis is required to test subject identifiability in even finer parcellations, which may ultimately lead to a plateau in I_{diff} for both eFC and nFC. Potentially, a multimodal approach that includes both modalities of functional brain network representations may be deployed for maximum subject identifiability. Further investigation is required to determine the effect of parcellations and the number of elements in measuring identifiability.

Collectively, our findings suggest that subject identification is driven by heteromodal regions from higher-order brain systems. These observations have clear implications for generating robust network biomarkers with subject-specific information, while reducing the amount of required data to its subset. More importantly, our results demonstrate that clusters of eFC with high variance, which maybe undermined in group-level analyses, may be useful in determining subject-specific characteristics or in personalized medical treatment. Also, this study leaves clues for future research on task-specific biomarkers to further specify the effect of tasks during fMRI acquisition on identifiability and to maximize subject-specific characteristics without requiring scans of prohibitive length, especially for vulnerable and clinical populations (Laumann et al., 2015). Lastly, clustering eFC matrices showed potential as a method of dimensionality reduction that is robust across subjects. Future work is necessary to clarify clusters or subsets of edges that are robust for group-level versus subject-level brain networks and answer whether the inter-subject idiosyncrasies and brain's eFC are modulated by task states.

Principal Component Analysis highlights idiosyncrasies in eFC

Here, we followed recent nFC studies and applied PCA to eFC data, effectively reducing eFC to a small set of principal components (Amico and Goñi, 2018b; Bari et al., 2019). We found that selectively reconstructing eFC using only those components that explained the greatest variance resulted in improved subject identifiability, far beyond the improvements when an identical procedure was applied to nFC data. This suggests that given the exactly same fMRI BOLD data, we can extract enhanced subject-level fingerprints from eFC data. We note that other methods such as z-score normalization of cross-subject correlation matrices also increase the absolute difference between within- versus between-subject scan similarities (Finn et al., 2015). Further research is necessary for determining strategies that maximize subject-specific information from a given dataset.

Interestingly, the number of PCs required to optimize eFC's subject identifiability matched the number of subjects in two independently

acquired datasets. These results parallel previous research using nFC for subject identification (Amico and Goñi, 2018b). Our results additionally show that this is robust despite a single subject having more than one test-retest pair of rsfMRI scans. Why, then, do the number of PCs match the number of subjects when optimizing for identifiability? To address this question, we analyzed the coefficients of PCs that increase (PC = 1 − 10) versus those that decrease (PC = 11 − 100) identifiability. The first principal component, mathematically, should explain the largest variance of eFC values across scans. This was the only PC out of 100 that had a consistently positive value. Also, only PCs 2 to 10 showed significant 'blockiness' for specific subject's scans. One possibility is that the first PC explains group-level eFC variance, the underlying group-level eFC features, whereas PCs 2 to 10 tend to explain subject-level eFC variance. We speculate that the number of blocky PCs are $N-1$ (N = number of subjects) since every subject can be identified with $N-1$ PCs *via* the process of elimination. These results provide further evidence that eFC can be a valuable framework when investigating and improving subject identifiability with linear transformation algorithms such as PCA. However, future work is necessary to disseminate the precise features of these PCs and alternative dimensionality reduction methods such as factor analysis or CCA (Child, 1990; Thompson, 1984) should be explored for I_{diff} optimization.

Future directions

Our results present exciting possibilities for future studies. Here, we used a previously-defined measure of identifiability. However, this measure can be misleading in some cases. For instance, I_{diff} can still take on high values if most subjects exhibit high levels of self-similarity, even if the remaining subjects exhibit poor self-similarity (Abbas et al., 2020a; 2020b; Amico and Goñi, 2018b; Jalbrzikowski et al., 2020a). Future work should investigate alternative measures for quantifying the performance of subject identification procedures including image intra-class correlation (I2C2) coefficients (Shou et al., 2013) and Discriminability (Bridgeford et al., 2020). Additionally, future work should explore multivariate approaches for subject identification (Yoo et al., 2019).

Questions on the origins of identifiability by tracking its changes across development and phylogeny should also be uncovered. Regions with high levels of identifiability exhibit pronounced cross-species differences (Xu et al., 2020), are consistently idiosyncratic across the human life span (Jalbrzikowski et al., 2020b), genetically influenced (Demeter et al., 2020), and highly identifiable in rat animal models (Bergmann et al., 2020). Understanding the origins of identifiability and connectivity-based idiosyncrasy could help inform our understanding of psychopathology, brain evolution, and both development and aging.

Applications of machine learning techniques (Demeter et al., 2020), combining various imaging modalities (Kumar et al., 2018), and increasing the granularity of fMRI parcellations (Tipnis et al., 2020) have been found effective in improving subject identifiability. Focusing on only those neural elements that maximize subject idiosyncrasies (Sripada et al., 2020) or selectively reconstructing eFC using high amplitude "event" time frames, which in previous work improved identifiability of nFC (Esfahlani et al., 2020), may help reduce the amount of data and computational resources required for improving identifiability with eFC. Future investigation is required for maximizing the uncovered idiosyncratic features from limited fMRI scan durations and computational resources using eFC.

Limitations

An overarching limitation associated with eFC is that linking it back to individual brain regions is challenging. Each connection in the eFC matrix always involves four nodes (two edge–edge connections), and allocating its properties to any one brain region or cognitive system is, in most cases, not possible. Here, we circumvented this complication by measuring the effect of a node's removal (equivalent to 199 edges in a 200 node parcellation) from the resulting eFC's identifiability and only accounting for edge pairs derived from single systems. While these approaches attempt to locate edges to brain regions, future work is necessary to determine a robust method in tracing eFC elements back to specific locations in brain.

A second limitation concerns the measure of "differential identifiability", which accounts for both within-subject and between-subject similarity. However, we acknowledge alternative measures of identifiability such as calculating the subject identification accuracy of FC-FC correlations (Finn et al., 2015), ROC (receiver operating characteristic) curve accuracies (Jalbrzikowski et al., 2020a), and using a Nearest Centroid Classification for each subject in graph embedding representations(Abbas et al., 2020a). Nonetheless, we focused our analyses using I_{diff} since this approach methodologically extends a popular identification approach (Finn et al., 2015) while accounting for "common features" shared across subjects (Amico and Goñi, 2018b). However, other methods for measuring identifiability and comparing fMRI data should also be investigated and developed to further understand idiosyncrasies in edge-centric FC.

A final limitation concerns the procedure for estimating edge clusters and utilizing its results. Here, we use a k-means algorithm to partition edges into a group-wise fixed number of clusters based on the eFC similarity. The benefit of this approach to estimating edge clusters is that this algorithm is computationally efficient and can be calculated using a distance metric. Given the extensive list of alternative methods for clustering (Fortunato, 2010; Porter et al., 2009; Sporns and Betzel, 2016), other algorithms must be investigated for detecting edge clusters. Also, between-cluster edges may act as bridges between functional systems that further explain idiosyncrasies and behavioral associations (Jo et al., 2020). However, we add that our analyses exclude between-cluster edges and focus on within-cluster edges. Future studies should investigate the effects of varying clustering algorithms and study the effect of within-cluster versus between-cluster edges on individual differences.

Materials and methods

Datasets

The Midnight Scan Club (MSC) dataset (Gordon et al., 2017) included rsfMRI from 10 adults (50% female, age = 29.1 ± 3.3). The study was approved by the Washington University School of Medicine Human Studies Committee and Institutional Review Board with the informed consent from all subjects. 12 scans from each subject were acquired on separate days starting from midnight. Per each subject, 10 rsfMRI scans were collected with a gradient-echo EPI sequence(run duration = 30 min, TR = 2200 ms, TE = 27 ms, flip angle = 90°) with the participant's eyes open while recording eyetracking to monitor for prolonged eye closure (for assessing drowsiness). Images were collected on a 3T Siemens Trio.

The Human Connectome Project (HCP) dataset (Van Essen et al., 2012) included resting state functional data (rsfMRI) from 100 unrelated adult subjects (54% female, mean age = 29.11 ± 3.67, age range = 22–36). These subjects were selected as they comprised the "100 Unrelated Subjects (U100)" released by the Human Connectome Project. The study was approved by the Washington University Institutional Review Board and informed consent was obtained from all subjects. Subjects underwent four 15 minute rsfMRI scans over a two day span. A full description of the imaging parameters and image preprocessing can be found in Glasser et al. (2013). The rsfMRI data was acquired with a gradient-echo EPI sequence (run duration = 14:33 min, TR = 720 ms, TE = 33.1ms, flip angle = 52°, 2 mm isotropic voxel resolution, multiband factor = 8) with eyes open and instructions to fixate on a cross. Images were collected on a 3T Siemens Connectome Skyra with a 32-channel head coil.

Image preprocessing

MSC functional preprocessing

Functional images in the MSC dataset were preprocessed using *fMRIPrep* 1.3.2 (Esteban et al., 2017), which is based on Nipype 1.1.9 (Gorgolewski et al., 2011). The following description of *fMRIPrep*'s preprocessing is based on boilerplate distributed with the software covered by a 'no rights reserved' (CCO) license. Internal operations of *fMRIPrep* use Nilearn 0.5.0 (Abraham et al., 2014), ANTs 2.2.0, FreeSurfer 6.0.1, FSL 5.0.9, and AFNI v16.2.07. For more details about the pipeline, see the section corresponding to workflows in fMRIPrep's documentation.

The T1-weighted (T1w) image was corrected for intensity non-uniformity with N4BiasFieldCorrection (Avants et al., 2008; Tustison et al., 2010), distributed with ANTS, and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a Nipype implementation of the antsBrainExtraction.sh workflow, using NKI as the target template. Brain surfaces were reconstructed using recon-all (Dale et al., 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter using Mindboggle (Klein et al., 2017). Spatial normalization to the *ICBM 152 Nonlinear Asymmetrical template version 2009c* (Fonov et al., 2009) was performed through nonlinear registration using antsRegistration, using brain-extracted versions of both T1w volume and template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter(GM) was performed on the brain-extracted T1w using FSL's fast (Zhang et al., 2001).

Functional data was slice time corrected using AFNI's `3dTshift` and motion corrected using FSL's `mcflirt`. *Fieldmap-less* distortion correction was performed by co-registering the functional image to the same subject. T1w image with intensity inverted constrained with an average fieldmap template, implemented with `antsRegistration`. This was followed by co-registration to the corresponding T1w using boundary-based registration with 9 degrees of freedom. Motion correcting transformations, field distortion correcting warp, BOLD-to-T1w transformation and T1w-to-template (MNI) warp were concatenated and applied in a single step using `antsApplyTransforms` using Lanczos interpolation. Several confounding timeseries were calculated based on this preprocessed BOLD framewise displacement (FD), DVARS and three region-wise global signals. FD and DVARS are calculated for each functional run, both using their implementations in Nipype. The three global signals are extracted within the CSF, the WM, and the whole-brain masks. The resultant `NIFTI` file for each MSC subject used in this study followed the file naming pattern: `*_spaceT1w_descpreproc_bold.nii.gz`.

HCP functional preprocessing

Functional images in the HCP dataset were minimally preprocessed according to the description provided in Glasser et al. (2013). Briefly, these data were corrected for gradient distortion, and motion, and then aligned to a corresponding T1-weighted (T1w) image with one spline interpolation step. This volume was further corrected for intensity bias and normalized to a mean of 10,000. This volume was then projected to the 32k_fs_LR mesh, excluding outliers, and aligned to a common space using a multi-modal surface registration Robinson et al. (2014). The resultant CIFTI file for each HCP subject used in this study followed the file naming pattern: `*_REST{1,2}_{LR,RL}_Atlas_MSMAll.dtseries.nii`.

Image quality control

All functional images in the MSC and HCP datasets were retained. The quality of functional images in the MSC were assessed using fMRIPrep's visual reports and MRIQC 0.15.1 (Esteban et al., 2017). MSC data was visually inspected for whole brain field of view coverage, signal artifacts, and proper alignment to the corresponding anatomical image. All time series data were visually inspected as well.

Functional networks preprocessing

Parcellation preprocessing

A functional parcellation designed to optimize both local gradient and global similarity measures of the fMRI signal(Schaefer et al., 2017) (Schaefer200) was used to define 200 areas on the cerebral cortex. These nodes are also mapped to the Yeo canonical functional networks (Thomas Yeo et al., 2011). For the HCP dataset, the Schaefer200 is openly available in 32k_fs_LR space as a CIFTI file. For the MSC dataset, a Schaefer200 parcellation was obtained for each subject using a Gaussian classifier surface atlas(Fischl et al., 2004) (trained on 100 unrelated HCP subjects) and FreeSurfer's mris_ca_label function. These tools utilize the surface registrations computed in the recon-all pipeline to transfer a group average atlas to subject space based on individual surface curvature and sulcal patterns. This method rendered a T1w space volume for each subject. For use with functional data, the parcellation was resampled to 2 mm T1w space. This process could be repeated for other resolutions of the parcellation (i.e. Schaefer100).

Functional network preprocessing

Each preprocessed BOLD image was linearly detrended, band-pass filtered (0.008 - 0.08 Hz) (Parkes et al., 2018), confound regressed and standardized using Nilearn's signal.clean, which removes confounds orthogonally to the temporal filters (Lindquist et al., 2019). The confound regression employed (Satterthwaite et al., 2013) included 6 motion estimates, time series of the mean CSF, mean WF, and mean global signal, the derivatives of these nine regressors, and the squares of these 18 terms. Furthermore, a spike regressor was added for each fMRI frame exceeding a motion threshold (MSC = 0.5 mm framewise displacement, HCP = 0.25 mm root mean squared displacement). This confound strategy has been shown to be a relatively effective option for reducing motion-related artifacts (Parkes et al., 2018). Following preprocessing and nuisance regression, residual mean BOLD time series at each node were recovered.

Edge graph construction

eFC can be calculated by acquiring the regional time series data and their z-scores. Next, for all pairs of brain regions, we calculate the element-wise product of their z-scored time series. This returns the 'edge time series' that represent the magnitude of co-fluctuation for pairs of brain regions which can be correlated across time as the Pearson correlation coefficient. Lastly, the scalar product between pairs of edge time seires is calculated and repeated over all pairs of edges to create an edge-by-edge matrix, which are normalized to the interval [−1, 1].

Differential identifiability

The functional connectome's identifiability or fingerprinting is based on the assumption that a single subject's connectivity profile should be, more similar within the same subject across scans and sessions than between different subjects. Previous research using the conventional functional connectome (Finn et al., 2015) showed that, robust identification of an individual is possible using sample FC to find the "target" FC of the subject in a pool of subject FCs with Pearson correlation analyses. Prior research on quantifying individual differences in functional connectivity include calculating the *geodesic distance* (Venkatesh et al., 2020) and *Pearson correlation* across individual's scans (Amico and Goñi, 2018b). While the geodesic distance approach also provides a summary measure of the inter-scan differences, we adopt the quantification metric by Amico and Goñi (2018b), which takes into consideration the covariance and standard deviation of the eFC and FC matrices. This metric is called the differential identifiability (I_{diff}) derived from the "identifiability matrix", i.e. the matrix of correlations (Pearson) between subjects' FCs. The I_{diff} is calculated by quantifying self-identifiability or I_{self} and

substracting between subject similarity or I_{others}, represented as the diagonal and off-diagonal elements of the identifiability matrix (Fig. 1b). Differential identifiability (I_{diff}) of a group of subjects can be summarized as the following:

$$I_{diff} = (I_{self} - I_{others}) * 100 \quad (1)$$

which is the difference of average within-subject similarity and average between-subjects similarity of FCs. A high value of I_{diff} in a dataset will have higher subject identifiability and the optimization of identifiability is reduced to maximizing I_{diff}.

To test for reproducibility and the effect of parcellations, we repeated the analysis with 100 unrelated subjects from the HCP dataset (Van Essen et al., 2012). The HCP dataset includes four scans of each subject collected during two scan sessions on separate dates. For each session, we concatenated both scans (left-to-right and right-to-left phase encoding) into a single session-representative time series. We later used these two time series to measure within-subject similarity I_{self}. Both datasets were tested for effects of numbers of parcellations by measuring *differential identifiability* in 100 and 200 node datasets (Fig. S1).

Tracing identifiability to brain regions by a leave-one-node-out approach

In the previous section, we described the procedure for calculating I_{diff}, a measure known to quantify the relative similarity of scans within versus between subjects(Amico and Goñi, 2018a). In this section, we test which brain regions reduce subject identification when removed from the calculation of eFC, compared to when measuring I_{diff} with the entire eFC matrix. In addition, we test whether removal of specific brain systems significantly reduce I_{diff} compared to using the whole-brain eFC.

The direct connection between eFC's edge-edge pair co-fluctuation and a brain region can somewhat be an arbitrary procedure since there can be up to $\binom{16}{1} + \binom{16}{2} + \binom{16}{3} + \binom{16}{4}$ combinations of brain regions in a single edge pair in eFC. To avoid making assumptions that edge pairs' weights are linearly related to the edge pair strength, we adopt a leave-one-node-out approach prior to eFC matrix construction. The effect of removing a node prior to eFC construction was calculated by subtracting the I_{diff} of the single-node-removed eFC from the whole-brain I_{diff} (Fig. 3*a-b*).

Next, we examined the effect of each functional brain system, and if removed, the difference in I_{diff} measured from eFC. Analogous to the single node removal approach earlier, we removed single systems (i.e. all nodes from a single system) and measured the effect on I_{diff} by subtracting this result from the whole-brain I_{diff} (Fig. 3*c-d*). We did not control for the difference in number of nodes (i.e. entire brain = 200 nodes, single-system-removed ≈ 190 nodes) since the effect of the total number of nodes to I_{diff} is unclear.

Single system edge pairs identifiability

The benefit and caveat of the leave-one-node-out approach is that it removes all edge pairs involving a particular node or system due to eFC's overlapping characteristic(Faskowitz et al., 2020). Therefore, we were still unclear of the effects of a purely single system due to this characteristic. In order to determine the I_{diff} a single-system, we extracted the edge pairs that include nodes only from a particular system Fig. 4*a* and measured the I_{diff}. The analogous approach was applied to nFC, which included node-pairs from the nFC which include nodes from a single system. The single-system level node-pairs were compared to that of edge pairs from eFC (Pearson correlation). Also, subject similarity matrices using edge pairs from single systems are included in (Fig. S3).

K-means clustering for identifiability

The eFC matrices used here have approximately a squared dimensionality of components compared to the conventional nodal FC matrices. While the higher dimensionality of eFC may provide insight into the relationship of it's components (the edge–edge communications) that is not directly shown in FC matrices, clustering the components of the eFC matrices present a computational challenge. This is especially the case if the number of partitions of the matrix is unknown and is left for exploration. To address this issue and cluster the eFC, we have applied a simple two-step clustering procedure that operates on a low-dimensional representation of the eFC matrix.

First, following (Faskowitz et al., 2020), we performed an eigendecomposition of the mean eFC matrix (19900 × 19900) created by averaging across eFC matrices from all scans, retaining the top 50 eigenvectors associated with the largest eigenvalues. These eigenvectors were rescaled to the interval [−1, 1] by dividing each eigenvector by its largest magnitude element. Then we clustered the rescaled eigenvectors using a standard k-means algorithm with Euclidean distance. We varied the number of communities, k, from $k = 2$ to 20, repeating the clustering algorithm 250 times at each value. We retained as a representative partition, the one with the greatest overall similarity to all other partitions.

We used k-means to partition edges into clusters, with the number of clusters ranging from $k = 2$ to $k = 20$ (a total of 209 different clusters). A partition of eFC into k clusters can be used to divide the corresponding eFC matrix into k blocks of within-cluster edge-edge connections and $k(k-1)/2$ blocks of between-cluster edge–edge connections. Our aim was to use features of these blocks to better understand how different clusters and systems were related to I_{diff}. These steps have been illustrated in Fig. S4 for further clarification.

To this end, we performed several analyses. First, for each block, we extracted its elements for every subject and scan and computed the pairwise similarity matrix. Two scans were considered similar to one another if the weights of edge–edge connections within that block were correlated. We then calculated several other features for each block. These included the mean and standard deviation across each block's edge–edge connections as well as the extent to which each of the brain systems were represented among the edges assigned to the block. In short, the blocks were derived from the partitions of the group-averaged eFC matrix into clusters. Each block represents the set of edge–edge connections between pairs of clusters. The eFC elements from each block from all scans and subjects were used for calculating the block's mean and standard deviation of eFC values. Then, for each of the k clusters, we calculated the averaged mean and standard deviation of all blocks involving that cluster. Finally, we counted the elements of each block by their system labels, which we found were correlated with the block's I_{diff}.

Mathematically, these clusters represent non-overlapping clusters of edges. We note that other community detection algorithms such as modularity maximization (Newman and Girvan, 2004) and Infomap (Rosvall and Bergstrom, 2008) could also be used. Here, however, we used k-means due to practicality (i.e. fast runtime).

Principal component analysis

Principal component analysis (PCA) is a widely used statistical method (Jolliffe, 2014) that allows exploration of the underlying structure of the data. PCA transforms a set of observed data with potentially correlated variables into a set of linearly uncorrelated variables called principal components. These principal components are then ranked in a descending order that explains the most to least variance of the data. We adopted principal component analysis to directly compare eFC's identifiability performance to that of nFC, which has been explored by Amico and Goñi (2018b).

First, the number of principal components are matched with the number of functional connectomes of the dataset. This allows for the decomposition from PCA, by definition, to account for 100% of the variance in the data. The PCs from PC = 2 to 20 were ranked by their explained variance in a descending order. Individual's nFC and eFC were then reconstructed as a function of the number of components included

based on the rationale that group-level information is carried in high variance components and subject-level information is conveyed in less higher variance components. In this reconstruction process, we additively included PCs in the descending order of explained variance. At each additive reconstruction, each individual's connectivity matrix was reconstructed based on the mean and linear recombination of the PCs from PC = 1 to N.

Next, we controlled for the effect of number of scans per individual, which affects the total scan duration or amount of data per subject (Fig. 2*a,b*). From the MSC dataset, we randomly selected two out of ten scans per subject as the test-retest scans for the PCA-derived I_{diff} maximization with 100 iterations. In both eFC and nFC, I_{diff} optimized with ten PCs, which match the number of subjects in the dataset. Also, we randomly sampled 10 subjects out of 100 to test if I_{diff} optimizes at the number of subjects regardless of the individuals. From the HCP dataset, we randomly selected ten subjects out of 100 with two scans per subject. In each of the 100 iterations, the I_{diff} for each PC was plotted for visualization (Fig. S7*c-d*). We also tested for the effect of scan duration by matching the number of time points in the HCP dataset (2400 time points) to that of the MSC dataset (800 time points; Fig. S7*f*). For each subject, to create a scan of 800 time points, we concatenated only the mid 400 time points (out of 1200) in each of the left-to-right and right-to-left scans. In other words, the mid section of two scans were connected to create a single scan of 800 time points for each subject. With the concatenated scans, we performed an analysis analogous to Fig. S7*c*. Finally, we calculated the I_{diff} from PC reconstruction using the entire HCP dataset (100 subjects; two scans per subject; Fig. S7*e*).

In both the MSC and HCP dataset, we found the number of subjects to match the number of principal components in which I_{diff} is maximized in the reconstructed eFC and nFC matrices. To determine the potential driver of this result, we decomposed each PC's coefficients for each scan Fig. S9, Fig. S10, Fig. S11, Fig. S12. For PCs 2 to 10, each subjects' coefficients were tested against that of the other subjects ($t-test$ with Bonferroni-correction).

Data Availability

All imaging data come from publicly-available, open-access repositories. Human connectome project data can be accessed *via* https://db.humanconnectome.org/app/template/Login.vm after signing a data use agreement. Midnight scan club data can be accessed *via* OpenNeuro at https://openneuro.org/datasets/dx000224/versions/1.0.1.

Code Availability

All processing and analysis code is available upon reasonable request.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.neuroimage.2021.118204

Credit authorship contribution statement

Youngheun Jo: Conceptualization, Formal analysis, Writing - original draft, Writing - review & editing. **Joshua Faskowitz:** Data curation, Writing - review & editing. **Farnaz Zamani Esfahlani:** Supervision, Writing - review & editing. **Olaf Sporns:** Supervision, Writing - review & editing. **Richard F. Betzel:** Conceptualization, Formal analysis, Writing - original draft.

References

Abbas, K., Amico, E., Svaldi, D. O., Tipnis, U., Duong-Tran, D. A., Liu, M., Rajapandian, M., Harezlak, J., Ances, B. M., Goñi, J., 2020a. Geff: graph embedding for functional fingerprinting. arXiv preprint arXiv:2001.06605.

Abbas, K., Liu, M., Venkatesh, M., Amico, E., Harezlak, J., Kaplan, A. D., Ventresca, M., Pessoa, L., Goñi, J., 2020b. Regularization of functional connectomes and its impact on geodesic distance and fingerprinting. arXiv preprint arXiv:2003.05393.

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., Varoquaux, G., 2014. Machine learning for neuroimaging with scikit-learn. Front. Neuroinform. 8, 14.

Ahn, Y.-Y., Bagrow, J.P., Lehmann, S., 2010. Link communities reveal multiscale complexity in networks. Nature 466 (7307), 761.

Amico, E., Goñi, J., 2018. Mapping hybrid functional-structural connectivity traits in the human connectome. Netw. Neurosci. 2 (3), 306–322.

Amico, E., Goñi, J., 2018. The quest for identifiability in human functional connectomes. Sci. Rep. 8 (1), 8254.

Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. 12 (1), 26–41.

Bari, S., Amico, E., Vike, N., Talavage, T.M., Goñi, J., 2019. Uncovering multi-site identifiability based on resting-state functional connectomes. NeuroImage 202, 115967.

Bergmann, E., Gofman, X., Kavushansky, A., Kahn, I., 2020. Individual variability in functional connectivity architecture of the mouse brain. bioRxiv.

Betzel, Richard, Cutts, Sarah, Greenwell, Sarah, Sporns, Olaf, 2021. Individualized event structure drives individual differences in whole-brain functional connectivity. bioRxiv doi:10.1101/2021.03.12.435168.

Betzel, R.F., Satterthwaite, T.D., Gold, J.I., Bassett, D.S., 2017. Positive affect, surprise, and fatigue are correlates of network flexibility. Sci. Rep. 7 (1), 520.

Bourin, C., Bondon, P., 1998. Efficiency of high-order moment estimates. IEEE Trans. Signal Process. 46 (1), 255–258.

Bridgeford, E.W., Wang, S., Yang, Z., Wang, Z., Xu, T., Craddock, C., Dey, J., Kiar, G., Gray-Roncal, W., Coulantoni, C., et al., 2020. Big data reproducibility: Applications in brain imaging and genomics. bioRxiv 802629.

Byrge, L., Kennedy, D.P., 2019. High-accuracy individual identification using a æthin sliceg of the functional connectome. Netw. Neurosci. 3 (2), 363–383.

Cao, M., Wang, J.-H., Dai, Z.-J., Cao, X.-Y., Jiang, L.-L., Fan, F.-M., Song, X.-W., Xia, M.-R., Shu, N., Dong, Q., et al., 2014. Topological organization of the human brain functional connectome across the lifespan. Dev. Cognit. Neurosci. 7, 76–93.

Child, D., 1990. The Essentials of Factor Analysis. Cassell Educational.

Choi, Y.Y., Shamosh, N.A., Cho, S.H., DeYoung, C.G., Lee, M.J., Lee, J.-M., Kim, S.I., Cho, Z.-H., Kim, K., Gray, J.R., et al., 2008. Multiple bases of human intelligence revealed by cortical thickness and neural activation. J. Neurosci. 28 (41), 10323–10329.

Cole, M.W., Bassett, D.S., Power, J.D., Braver, T.S., Petersen, S.E., 2014. Intrinsic and task-evoked network architectures of the human brain. Neuron 83 (1), 238–251.

Cole, M.W., Yarkoni, T., Repovš, G., Anticevic, A., Braver, T.S., 2012. Global connectivity of prefrontal cortex predicts cognitive control and intelligence. J. Neurosci. 32 (26), 8988–8999.

Cox, R., Schapiro, A.C., Stickgold, R., 2018. Variability and stability of large-scale cortical oscillation patterns. Netw. Neurosci. 2 (4), 481–512.

Crossley, N.A., Mechelli, A., Vértes, P.E., Winton-Brown, T.T., Patel, A.X., Ginestet, C.E., McGuire, P., Bullmore, E.T., 2013. Cognitive relevance of the community structure of the human brain functional coactivation network. Proc. Natl. Acad. Sci. 110 (28), 11583–11588.

Cui, Z., Li, H., Xia, C.H., Larsen, B., Adebimpe, A., Baum, G.L., Cieslak, M., Gur, R.E., Gur, R.C., Moore, T.M., et al., 2020. Individual variation in functional topography of association networks in youth. Neuron.

Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis: I. segmentation and surface reconstruction. Neuroimage 9 (2), 179–194.

Demeter, D.V., Engelhardt, L.E., Mallett, R., Gordon, E.M., Nugiel, T., Harden, K.P., Tucker-Drob, E.M., Lewis-Peacock, J.A., Church, J.A., 2020. Functional connectivity fingerprints at rest are similar across youths and adults and vary with genetic similarity. Iscience 23 (1), 100801.

Dubois, J., Adolphs, R., 2016. Building a science of individual differences from fmri. Trends Cognit. Sci. 20 (6), 425–443.

Elston, G.N., Oga, T., Fujita, I., 2009. Spinogenesis and pruning scales across functional hierarchies. J. Neurosci. 29 (10), 3271–3275.

Esfahlani, F.Z., Jo, Y., Faskowitz, J., Byrge, L., Kennedy, D., Sporns, O., Betzel, R., 2020. High-amplitude co-fluctuations in cortical activity drive functional connectivity. bioRxiv 800045.

Esteban, O., Birman, D., Schaer, M., Koyejo, O.O., Poldrack, R.A., Gorgolewski, K.J., 2017. Mriqc: advancing the automatic prediction of image quality in MRI from unseen sites. PloS One 12 (9), e0184661.

Evans, T., Lambiotte, R., 2009. Line graphs, link partitions, and overlapping communities. Phys. Rev. E 80 (1), 016105.

Fair, D.A., Cohen, A.L., Power, J.D., Dosenbach, N.U., Church, J.A., Miezin, F.M., Schlaggar, B.L., Petersen, S.E., 2009. Functional brain networks develop from a æ̈local to distributedg organization. PLoS Comput. Biol. 5 (5), e1000381.

Faskowitz, J., Esfahlani, F.Z., Jo, Y., Sporns, O., Betzel, R.F., 2020. Edge-centric functional network representations of human cerebral cortex reveal overlapping system-level architecture. Nat. Neurosci. 23 (12), 1644–1654.

Finn, E.S., Shen, X., Scheinost, D., Rosenberg, M.D., Huang, J., Chun, M.M., Papademetris, X., Constable, R.T., 2015. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. Nat. Neurosci. 18 (11), 1664.

Fischl, B., Van Der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., et al., 2004. Automatically parcellating the human cerebral cortex. Cereb. Cortex 14 (1), 11–22.

Fonov, V.S., Evans, A.C., McKinstry, R.C., Almli, C., Collins, D., 2009. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage (47) S102.

Fornito, A., Harrison, B., 2012. Brain connectivity and mental illness. Front. Psychiatry 3, 72.
Fornito, A., Zalesky, A., Breakspear, M., 2015. The connectomics of brain disorders. Nat. Rev. Neurosci. 16 (3), 159–172.
Fortunato, S., 2010. Community detection in graphs. Phys. Rep. 486 (3-5), 75–174.
Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., et al., 2016. A multi-modal parcellation of human cerebral cortex. Nature 536 (7615), 171–178.
Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., et al., 2013. The minimal preprocessing pipelines for the human connectome project. Neuroimage 80, 105–124.
Gordon, E.M., Laumann, T.O., Gilmore, A.W., Newbold, D.J., Greene, D.J., Berg, J.J., Ortega, M., Hoyt-Drazen, C., Gratton, C., Sun, H., et al., 2017. Precision functional mapping of individual human brains. Neuron 95 (4), 791–807.
Gordon, E.M., Lynch, C.J., Gratton, C., Laumann, T.O., Gilmore, A.W., Greene, D.J., Ortega, M., Nguyen, A.L., Schlaggar, B.L., Petersen, S.E., et al., 2018. Three distinct sets of connector hubs integrate human brain function. Cell Rep. 24 (7), 1687–1695.
Gorgolewski, K., Burns, C.D., Madison, C., Clark, D., Halchenko, Y.O., Waskom, M.L., Ghosh, S.S., 2011. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. Front. Neuroinform. 5, 13.
Gratton, C., Kraus, B.T., Greene, D.J., Gordon, E.M., Laumann, T.O., Nelson, S.M., Dosenbach, N.U., Petersen, S.E., 2019. Defining individual-specific functional neuroanatomy for precision psychiatry. Biol. Psychiatry.
Gratton, C., Laumann, T.O., Nielsen, A.N., Greene, D.J., Gordon, E.M., Gilmore, A.W., Nelson, S.M., Coalson, R.S., Snyder, A.Z., Schlaggar, B.L., et al., 2018. Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation. Neuron 98 (2), 439–452.
Greicius, M., 2008. Resting-state functional connectivity in neuropsychiatric disorders. Curr. Opin. Neurol. 21 (4), 424–430.
Holmes, C.J., Hoge, R., Collins, L., Woods, R., Toga, A.W., Evans, A.C., 1998. Enhancement of mr images using registration for signal averaging. J. Comput. Assist. Tomogr. 22 (2), 324–333.
Horien, C., Scheinost, D., Constable, R.T., 2019. The uniqueness of the individual functional connectome. In: Connectomics. Elsevier, pp. 63–81.
Horien, C., Shen, X., Scheinost, D., Constable, R.T., 2019. The individual functional connectome is unique and stable over months to years. Neuroimage 189, 676–687.
Huth, A.G., De Heer, W.A., Griffiths, T.L., Theunissen, F.E., Gallant, J.L., 2016. Natural speech reveals the semantic maps that tile human cerebral cortex. Nature 532 (7600), 453–458.
Huth, A.G., Nishimoto, S., Vu, A.T., Gallant, J.L., 2012. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron 76 (6), 1210–1224.
Jalbrzikowski, M., Liu, F., Foran, W., Klei, L., Calabro, F.J., Roeder, K., Devlin, B., Luna, B., 2020. Functional connectome fingerprinting accuracy in youths and adults is similar when examined on the same day and 1.5 years apart. bioRxiv doi:10.1101/812719.
Jalbrzikowski, M., Liu, F., Foran, W., Klei, L., Calabro, F.J., Roeder, K., Devlin, B., Luna, B., 2020. Functional connectome fingerprinting accuracy in youths and adults is similar when examined on the same day and 1.5 years apart. bioRxiv 812719.
Jo, Y., Esfahlani, F.Z., Faskowitz, J., Chumin, E.J., Sporns, O., Betzel, R.F., 2020. The diversity and multiplexity of edge communities within and between brain systems. bioRxiv doi:10.1101/2020.05.05.067777.
Jolliffe, I., 2014. Principal Component Analysis. American Cancer Society.
Klein, A., Ghosh, S.S., Bao, F.S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Neto, E.C., et al., 2017. Mindboggling morphometry of human brains. PLoS Comput. Biol. 13 (2).
Kliemann, D., Adolphs, R., Tyszka, J.M., Fischl, B., Yeo, B.T., Nair, R., Dubois, J., Paul, L.K., 2019. Intrinsic functional connectivity of the brain in adults with a single cerebral hemisphere. Cell Rep. 29 (8), 2398–2407.
Kumar, K., Toews, M., Chauvin, L., Colliot, O., Desrosiers, C., 2018. Multi-modal brain fingerprinting: a manifold approximation based framework. NeuroImage 183, 212–226.
Laumann, T.O., Gordon, E.M., Adeyemo, B., Snyder, A.Z., Joo, S.J., Chen, M.-Y., Gilmore, A.W., McDermott, K.B., Nelson, S.M., Dosenbach, N.U., et al., 2015. Functional system and areal organization of a highly sampled individual human brain. Neuron 87 (3), 657–670.
Lindquist, M.A., Geuter, S., Wager, T.D., Caffo, B.S., 2019. Modular preprocessing pipelines can reintroduce artifacts into fmri data. Hum. Brain Mapp. 40 (8), 2358–2376.
Miranda-Dominguez, O., Mills, B.D., Carpenter, S.D., Grant, K.A., Kroenke, C.D., Nigg, J.T., Fair, D.A., 2014. Connectotyping: model based fingerprinting of the functional connectome. PloS One 9 (11), e111048.
Mirchi, N., Betzel, R.F., Bernhardt, B.C., Dagher, A., Mišić, B., 2019. Tracking mood fluctuations with functional network patterns. Soc. Cognit. Affect. Neurosci. 14 (1), 47–57.
Mueller, S., Wang, D., Fox, M.D., Yeo, B.T., Sepulcre, J., Sabuncu, M.R., Shafee, R., Lu, J., Liu, H., 2013. Individual variability in functional connectivity architecture of the human brain. Neuron 77 (3), 586–595.
Nepusz, T., Vicsek, T., 2012. Controlling edge dynamics in complex networks. Nat. Phys. 8 (7), 568.
Newman, M.E., Girvan, M., 2004. Finding and evaluating community structure in networks. Phys. Rev. E 69 (2), 026113.
Owen, L.L., Chang, T.H., Manning, J.R., 2019. High-level cognition during story listening is reflected in high-order dynamic correlations in neural activity patterns. bioRxiv 763821.
Parkes, L., Fulcher, B., Yücel, M., Fornito, A., 2018. An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional mri. Neuroimage 171, 415–436.
Peña-Gómez, C., Avena-Koenigsberger, A., Sepulcre, J., Sporns, O., 2018. Spatiotemporal network markers of individual variability in the human functional connectome. Cerebr. Cortex 28 (8), 2922–2934.
Poldrack, R.A., Laumann, T.O., Koyejo, O., Gregory, B., Hover, A., Chen, M.-Y., Gorgolewski, K.J., Luci, J., Joo, S.J., Boyd, R.L., et al., 2015. Long-term neural and physiological phenotyping of a single human. Nat. Commun. 6, 8885.
Porter, M.A., Onnela, J.-P., Mucha, P.J., 2009. Communities in networks. Not. AMS 56 (9), 1082–1097.
Rajapandian, M., Amico, E., Abbas, K., Ventresca, M., Goñi, J., 2020. Uncovering differential identifiability in network properties of human brain functional connectomes. Netw. Neurosci. 4 (3), 698–713.
de Reus, M.A., Saenger, V.M., Kahn, R.S., van den Heuvel, M.P., 2014. An edge-centric perspective on the human connectome: link communities in the brain. Philos. Trans. R. Soc. B: Biol. Sci. 369 (1653), 20130527.
Robinson, E.C., Jbabdi, S., Glasser, M.F., Andersson, J., Burgess, G.C., Harms, M.P., Smith, S.M., Van Essen, D.C., Jenkinson, M., 2014. Msm: a new flexible framework for multimodal surface matching. Neuroimage 100, 414–426.
Rosvall, M., Bergstrom, C.T., 2008. Maps of random walks on complex networks reveal community structure. Proc. Natl. Acad. Sci. 105 (4), 1118–1123.
Salehi, M., Karbasi, A., Barron, D.S., Scheinost, D., Constable, R.T., 2020. Individualized functional networks reconfigure with cognitive state. NeuroImage 206, 116233.
Satterthwaite, T.D., Elliott, M.A., Gerraty, R.T., Ruparel, K., Loughead, J., Calkins, M.E., Eickhoff, S.B., Hakonarson, H., Gur, R.C., Gur, R.E., et al., 2013. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. Neuroimage 64, 240–256.
Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.-N., Holmes, A.J., Eickhoff, S.B., Yeo, B.T., 2017. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. Cerebr. Cortex 1–20.
Seitzman, B.A., Gratton, C., Laumann, T.O., Gordon, E.M., Adeyemo, B., Dworetsky, A., Kraus, B.T., Gilmore, A.W., Berg, J.J., Ortega, M., et al., 2019. Trait-like variants in human functional brain networks. Proc. Natl. Acad. Sci. 116 (45), 22851–22861.
Sepulcre, J., Liu, H., Talukdar, T., Martincorena, I., Yeo, B.T., Buckner, R.L., 2010. The organization of local and distant functional connectivity in the human brain. PLoS Comput. Biol. 6 (6), e1000808.
Shou, H., Eloyan, A., Lee, S., Zipunnikov, V., Crainiceanu, A., Nebel, M., Caffo, B., Lindquist, M., Crainiceanu, C.M., 2013. Quantifying the reliability of image replication studies: the image intraclass correlation coefficient (i2c2). Cognit., Affect. Behav. Neurosci. 13 (4), 714–724.
Smith, S.M., Nichols, T.E., Vidaurre, D., Winkler, A.M., Behrens, T.E., Glasser, M.F., Ugurbil, K., Barch, D.M., Van Essen, D.C., Miller, K.L., 2015. A positive-negative mode of population covariation links brain connectivity, demographics and behavior. Nat. Neurosci. 18 (11), 1565–1567.
Sporns, O., Betzel, R.F., 2016. Modular brain networks. Annu. Rev. Psychol. 67, 613–640.
Sporns, O., Faskowitz, J., Teixera, A.S., Betzel, R., 2020. Dynamic expression of brain functional systems disclosed by fine-scale analysis of edge time series. bioRxiv.
Sporns, O., Faskowitz, J., Teixera, S., Betzel, R., 2020. Dynamic expression of brain functional systems disclosed by fine-scale analysis of edge time series. bioRxiv.
Sripada, C., Taxali, A., Angstadt, M., Rutherford, S., 2020. Boost in test-retest reliability in resting state fmri with predictive modeling. bioRxiv doi:10.1101/796714.
Svaldi, D. O., Goñi, J., Abbas, K., Amico, E., Clark, D. G., Muralidharan, C., Dzemidzic, M., West, J. D., Risacher, S. L., Saykin, A. J., et al., 2019. Optimizing differential identifiability improves connectome predictive modeling of cognitive deficits in alzheimer\'s disease. arXiv preprint arXiv:1908.06197.
Svaldi, D.O., Goñi, J., Sanjay, A.B., Amico, E., Risacher, S.L., West, J.D., Dzemidzic, M., Saykin, A., Apostolova, L., 2018. Towards subject and diagnostic identifiability in the alzheimer's disease spectrum based on functional connectomes. In: Graphs in Biomedical Image Analysis and Integrating Medical Imaging and Non-Imaging Modalities. Springer, pp. 74–82.
Thomas Yeo, B., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, L., Polimeni, J.R., et al., 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. 106 (3), 1125–1165.
Thompson, B., 1984. Canonical correlation analysis: Uses and interpretation. Sage.
Tipnis, U., Abbas, K., Tran, E., Amico, E., Shen, L., Kaplan, A. D., Goñi, J., 2020. Processed functional connectomes for the hcp young adult: Data release and assessment on brain fingerprints. arXiv preprint arXiv:2011.05212.
Trinh, H.-C., Kwon, Y.-K., 2016. Edge-based sensitivity analysis of signaling networks by using boolean dynamics. Bioinformatics 32 (17), i763–i771.
Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C., 2010. N4itk: improved n3 bias correction. IEEE Trans. Med. Imaging 29 (6), 1310–1320.
Uddin, L.Q., 2020. An edgy'new look. Nat. Neurosci. 23 (12), 1471–1472.
Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., Consortium, W.-M.H., et al., 2013. The wu-minn human connectome project: an overview. Neuroimage 80, 62–79.
Van Essen, D.C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S.W., et al., 2012. The human connectome project: a data acquisition perspective. Neuroimage 62 (4), 2222–2231.
Venkatesh, M., Jaja, J., Pessoa, L., 2020. Comparing functional connectivity matrices: a geometry-aware approach applied to participant identification. NeuroImage 207, 116398.
Xia, C.H., Ma, Z., Ciric, R., Gu, S., Betzel, R.F., Kaczkurkin, A.N., Calkins, M.E., Cook, P.A., de la Garza, A.G., Vandekar, S.N., et al., 2018. Linked dimensions of psychopathology and connectivity in functional brain networks. Nat. Commun. 9 (1), 1–14.
Xu, T., Nenning, K.-H., Schwartz, E., Hong, S.-J., Vogelstein, J.T., Goulas, A., Fair, D.A., Schroeder, C.E., Margulies, D.S., Smallwood, J., et al., 2020. Cross-species functional

alignment reveals evolutionary hierarchy within the connectome. NeuroImage 223, 117346.
Yoo, K., Rosenberg, M.D., Noble, S., Scheinost, D., Constable, R.T., Chun, M.M., 2019. Multivariate approaches improve the reliability and validity of functional connectivity and prediction of individual behaviors. Neuroimage 197, 212–223.
Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain mr images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans. Med. Imaging 20 (1), 45–57.
Zilles, K., Armstrong, E., Schleicher, A., Kretschmann, H.-J., 1988. The human pattern of gyrification in the cerebral cortex. Anatomy Embryol. 179 (2), 173–179.