MGN-Net: A multi-view graph normalizer for integrating heterogeneous biological network populations ☆

Mustafa Burak Gürbüz[a], Islem Rekik[a,b,*]

[a] *BASIRA Lab, Faculty of Computer and Informatics Engineering, Istanbul Technical University, Istanbul, Turkey*
[b] *School of Science and Engineering, Computing, University of Dundee, UK*

A B S T R A C T

With the recent technological advances, biological datasets, often represented by networks (i.e., graphs) of interacting entities, proliferate with unprecedented complexity and heterogeneity. Although modern network science opens new frontiers of analyzing connectivity patterns in such datasets, we still lack data-driven methods for extracting an integral connectional fingerprint of a multi-view graph population, let alone disentangling the typical from the atypical variations across the population samples. We present the multi-view graph normalizer network (MGN-Net[2]), a graph neural network based method to normalize and integrate a set of multi-view biological networks into a single connectional template that is centered, representative, and topologically sound. We demonstrate the use of MGN-Net by discovering the connectional fingerprints of healthy and neurologically disordered brain network populations including Alzheimer's disease and Autism spectrum disorder patients. Additionally, by comparing the learned templates of healthy and disordered populations, we show that MGN-Net significantly outperforms conventional network integration methods across extensive experiments in terms of producing the most centered templates, recapitulating unique traits of populations, and preserving the complex topology of biological networks. Our evaluations showed that MGN-Net is powerfully generic and easily adaptable in design to different graph-based problems such as identification of relevant connections, normalization and integration.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

Modern network science has introduced exciting new opportunities for understanding the underpinning mechanisms of biological systems by examining interactions within their components (Ideker et al., 2001). In the face of the ongoing 'tsunami' of biological data collection spanning the range from genetic (Adams et al., 1991) and metabolic networks (Kanehisa and Goto, 2000) all the way up to social and economic systems (Salathé et al., 2010), data-driven network representations have allowed us to map complex interplay between components of biological systems such as genetic data by revealing gene co-expression and connectomic data by investigating correlations in neural signaling between different brain regions. Namely, graphs present a natural tool to study such interactions (or connections) in complex biological data including protein-protein interactions (Safari et al., 2014), metabolic networks (Lee et al., 2008), and brain connectivity networks which span the field of network neuroscience (Bullmore and Sporns, 2009). In the former context, network neuroscience proposes to encode the brain wiring in a graph by representing the brain regions as nodes and their interactions as edges linking those nodes, which has propelled the development of advanced network-based analysis techniques of the brain construct (Bassett and Sporns, 2017). Particularly given the recent proliferation in large and multi-modal connectomic datasets such as the Human Connectome Project (Essen et al., 2012) acquired using multiple neuroimaging modalities including structural T1-weighted, diffusion, and functional magnetic resonance imaging (MRI), it is not always obvious how to integrate multi-modal connectomic data together (Van Essen and Glasser, 2016), nor easy to do so in practice, in order to first understand how the brain's structural, morphological and functional levels interlink to form this integrated com-

Fig. 1. Connectional template comparison for identifying discriminative connections differentiating between typical and atypical populations of networks.

plex system, and then identify typical and atypical connectional trends fingerprinting the human brain. This is substantially due to the large variability in brain connectivity across individuals, which limits our ability to disentangle the 'healthy' brain connectional variability from the 'pathological' variability. For instance, two individuals who largely differ in particular brain connections might not indicate that one of them has a pathological connection, this difference can still fit healthy connectional brain patterns. To distinguish between healthy and disordered connectional variability, we need to define a 'normalization' or 'standardization' process of brain networks (Fig. 1). Eventually, we hypothesize that reducing inter-subject variability in both healthy and disordered populations through a *normalization process* will contribute towards helping better identify 'pathological' alterations in brain networks as deviations from the 'standard/normalized' brain network representation. Without any loss of generalizability, this line of reasoning extends to other biological networks with multiple views, meaning that each sample is represented by a set of networks, where each network view captures unique traits of the sample (for example, functional, morphological and structural).

Here we propose a novel multi-view graph integration (MGI) method which produces a unified normalized connectional representation of a population of multi-view networks; for example a connectional brain template (CBT) from a heterogeneous population of multi-view brain networks. Specifically, our MGI method taps into the power of deep-learning multi-disciplinary architectures, which can handle large-scale, highly non-linear and heterogeneous datasets. Although deep learning has been recently used to regress brain networks (Bronstein et al., 2017), there has not been any work on applying deep learning for fusing networks in general. Critically, fusing multi-view biological networks is an uncharted territory where network science and deep learning have not cross-fertilized, especially the emerging field of geometric deep learning on graphs and manifolds (i.e., graph neural networks). Here we set out to integrate a population of multi-view biological graphs with the aim to estimate a representative reference connectional template by normalizing connections across the population

samples, which is an essential step for group comparison studies as well as discovering the integral signature of an anomaly in a given population (e.g., disordered) by comparison with a typical connectional template (e.g., healthy). In this context, we hypothesize that a population-driven connectional template satisfies the following properties: (1) well-centeredness, (2) discriminativeness, and (3) topological soundness. A *well-centered* template occupies the 'center' of a population by achieving the minimum distance to all population samples. A *discriminative* template implies that the estimated template consistently captures the unique and distinctive traits of a population. Last but not least, convergent studies show that a large variety of biological networks has extraordinarily complex yet highly organized topological patterns such as the spatially economical layout of brain regions that are likely to be a consequence of the conservation of wiring costs being an important selection pressure on the evolution of brain networks (Bullmore and Sporns, 2009). Therefore, the estimated template should be topologically sound by preserving the population topological properties during the normalization process.

From a deep learning perspective, methods for multi-view network integration are currently lacking. The simplest way to integrate a set of biological networks is to linearly average them. However, such a normalization technique alone is very sensitive to outliers and cannot be generalized to blend information of multi-view network datasets with heterogeneous distributions. Currently, the prevailing technique for non-linear network integration is similarity network fusion (SNF) (Wang et al., 2014), which is based on message passing theory (Pearl, 1988). SNF aims to estimate a status matrix for each network that carries the whole information in the networks and a sparse local matrix that only takes up to top-k neighbors into consideration. Next, an iterative integration step is conducted to update each status network through diffusing mean global structure of remaining networks and along with the sparse local network. Even though SNF is a powerful tool since it is a generic unsupervised technique, it comes with strong assumptions such as emphasizing the top k local connections for each node and equally averaging the global topology of comple-

mentary networks for each iterative update to ultimately merge them. Another very recent approach, the netNorm (Dhifallah and Rekik, 2019), utilizes a graph-based feature selection along with SNF to integrate multi-view networks. netNorm first constructs a high-order graph using cross-view connectional features as nodes and their Euclidean distance as a dissimilarity measure to select the most centered ones across the population. Next, for each network view, it composes a mosaic of the selected edges across subjects and eventually integrates the mosaic network views into a single network using SNF. Despite the promising results on multi-view datasets, netNorm has recognized limitations. First, it uses the Euclidean distance as a predefined metric for selecting the most representative connections which might fail to capture complex non-linear patterns in the given population. Second, netNorm consists of independent feature extraction, feature selection, and fusion steps. These fully independent steps cannot provide feedback to each other in order to globally optimize the template estimation process. Therefore, the pipeline is agnostic to cumulative errors. More importantly, both SNF (Wang et al., 2014) and netNorm (Dhifallah and Rekik, 2019) do not have any mechanism to preserve complex topological patterns in biological networks during the integration process, which is undeniably substantial for outputting topologically sound connectional templates.

Here we propose the multi-view graph normalizer network (MGN-Net), a novel graph neural network (GNN) based method for integrating and normalizing a set of multi-view graphs to *learn* a representative connectional template for a given population. Our approach is inspired by cutting-edge, but so far neglected GNN frameworks (Kipf and Welling, 2017; Defferrard et al., 2016; Xu et al., 2019; Veličković et al., 2018) in the field of network integration. It is also distinct in that it circumvents the need for hand-crafted steps and general assumptions as it learns how to estimate the best template within an end-to-end optimization framework. GNNs are an emerging subfield of deep learning which extends the idea of convolutional neural networks (CNNs) (Krizhevsky et al., 2012) to non-Euclidean data such as graphs and surfaces. GNNs achieved remarkable results in several recent biomedical data analysis studies such as disease classification (Rhee et al., 2018; Parisot et al., 2018) and protein interaction prediction (Gainza et al., 2020; Fout et al., 2017). MGN-net capitalizes on graph neural network layers to explore implicit patterns that exist in the population of multi-view graphs and estimate the best template that is well-centered, discriminative, and topologically sound.

This work presents an extension to the recent conference MIC-CAI 2020 paper (Gurbuz and Rekik, 2020). The method presented in (Gurbuz and Rekik, 2020) introduces the Subject Normalization Loss (SNL) function for optimizing the proposed Deep Graph Normalizer (DGN) architecture. However, SNL does not constrain the integration process in terms of maintaining the complex topology of biological networks. Furthermore, it does not evaluate the topological soundness of generated connectional templates. To address these limitations, we further propose a novel loss function that penalizes the deviation from the ground-truth node strength *topological* distribution. Through additional experiments, we show that this novel loss function not only enforces the learning of more topologically sound templates but also increases the performance in terms of centeredness and preserving discriminative traits of the network populations during the population multigraph integration.

2. Method

MGN-Net: a general framework to learn an integral and holistic connectional template of a multi-view graph population. MGN-Net takes (Fig. 2–a) two or more isomorphic weighted (possibly unweighted) multi-view graphs and maps them onto an output population center graph (i.e., connectional template). This

learning task is fundamentally rooted in embedding connectivity patterns onto a high-dimensional vector representation for each node in each graph in the given population, namely a node embedding (Fig. 2–b). During the embedding process, we preserve the unique domain-specific topological properties of the population graphs thanks to our novel topology-constrained normalization loss function which penalizes the deviation from ground-truth population topology. Next, we derive the template edges from the pairwise relationship of node embeddings (Fig. 2–c). This relationship can be measured by any function that allows for deep network backpropagation depending on the application so that the MGN-Net can be trained in an end-to-end manner. Finally, we utilize our novel topology-constrained randomized loss function to minimize the distance between the population network views and the learned biological template.

Following the optimization of our MGN-Net architecture in the training stage, the learned connectional templates by the MGN-Net encapsulates both shared traits among samples and different yet complementary information offered by multi-view biological measurements. We used our method to integrate both small-scale and large-scale multi-view brain connectome datasets derived from T1-weighted MRI to estimate CBTs. We demonstrate that MGN-Net significantly outperforms other baseline network integration tools on both datasets in terms of centeredness (Fig. 4), preserving the highly organized topology of brain networks (Table 4), and identifying the most discriminative connections distinguishing between healthy and disordered brain states (Table 5). We tested the fingerprinting property of the learned CBTs by conducting a simple comparison between the learned healthy and disordered population CBTs and selecting the most discriminative connectional features that boost the classification accuracy of an independent machine learning-based diagnostic model. Our experiments highlighted the necessity of multi-view network integration to provide a normalized and expressive characterization of a population of brain multi-view graphs in both Alzheimer's diseases (AD) and the autism spectrum disorder (ASD) populations (Table 6).

MGN-Net is an end-to-end graph neural network based architecture for normalizing multi-view graphs. To fully exploit the topological information offered by complex graph structures, we tap into the nascent field of GNNs on non-Euclidean spaces such as graphs instead of conventional learning methods where the implicit patterns are overlooked since graphs are projected onto a Euclidean space and processed similarly to any other numerical data. Given a population of multi-view samples, where each sample is represented by two or more graphs derived from different measurements (Fig. 2–a), MGN-Net first feeds each multi-view sample through 3 graph convolutional layers. From layer to layer, deeper holistic embeddings are learned for each node recapitulating the complementary information offered by heterogeneous measurements (Fig. 2–b). Next, we produce the normalized connectional template graph using pairwise relationships of node embeddings outputted by the final layer (Fig. 2–c).

To evaluate the representativeness of the learned population templates, we propose the topology-constrained normalization loss (TCNL) function. The TCNL loss is composed of two parts. First, we compute the mean Frobenius distance between the learned template and a random subset of training samples to evaluate its centredness. Second, we compute the Kullback-Leibler divergence of node strength distributions of the connectional template and the random subset of training samples with the aim of measuring deviation from the real topology of the population multi-view graphs. We have particularly chosen the node strength distribution to constrain our loss optimization given that node strength presents a simple yet the most fundamental network measure and other advanced topological measures such as centrality measures depend on it (Bullmore and Sporns, 2009). Note that TCNL compares the
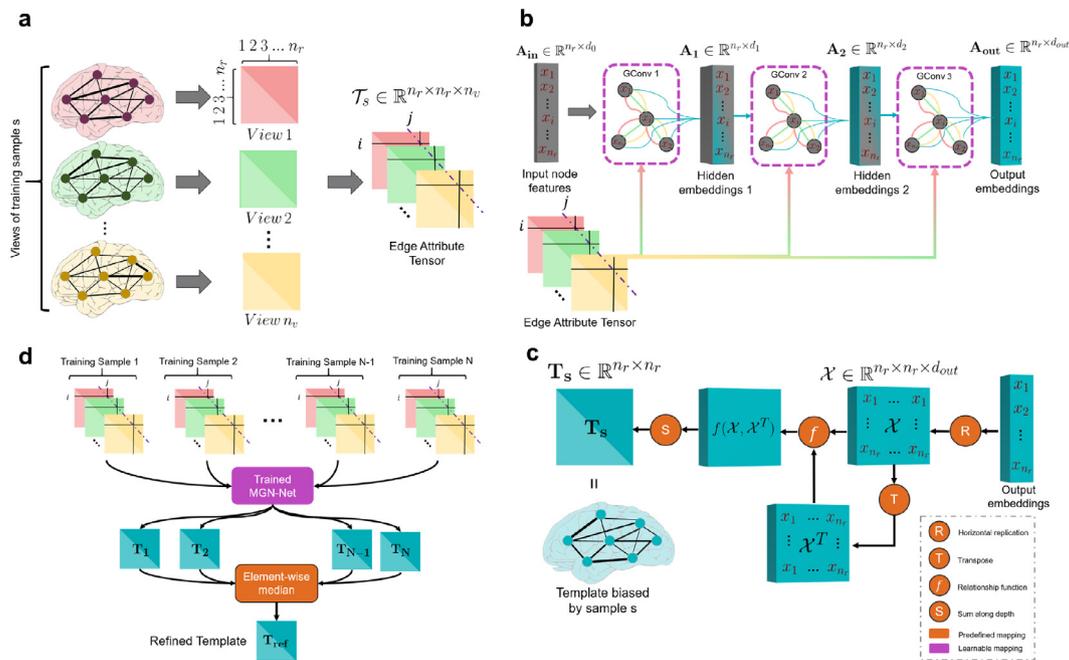
Fig. 2. Overview of the proposed multi-view graph normalizer network (MGN-Net) architecture for integrating a population of multi-view graphs. **a) Tensor representation of multi-view networks.** We represent each input sample by a tensor $\mathcal{T}_s \in \mathbb{R}^{n_r \times n_r \times n_v}$ where n_r is the number of nodes and n_v is the number of heterogeneous views. **b) Node embedding learning.** Abstract vector representation for each node of the input network is learned through 3 graph neural network layer. **c) Tensor representation of pair-wise relationship computation.** Calculation of pairwise relation of node embeddings is reformulated as a tensor operation for easy and efficient backpropagation. **d) Template refinement after training.** We execute a post-training refinement step to eliminate any bias towards the given input subject by computing the median of all possible templates to create the ultimate template $\mathbf{T}_{ref}$ for the given population.

learned graph template with other multi-view graphs in the population, therefore the MGN-Net, by learning the most essential connectivities and normalizing trivial variabilities specific to samples, maps each multi-view graph to a population-representative template. The randomization of the training samples used for network normalization induces a loss regularization effect since it is much easier for the model to overfit if the loss is calculated against the whole dataset in each iteration. Moreover, the random sampling size is predefined as a hyperparameter that is much smaller than the training dataset size, hence the magnitude and the computation time of loss are independent of the number of subjects in the dataset.

Following the completion of the training phase, MGN-Net can map *any* multi-view biological graph to a connectional template that fingerprints the input population. However, each fingerprint is biased towards the particular subject that is used to create it. In order to avoid such bias, we further embed a post-training regularization step (Fig. 2–d). First, we generate biased population templates by feeding each input to the trained MGN-Net. Then we pick the most centered connections across biased templates by calculating the element-wise median to obtain a finalized connectional template that represents the population most.

2.1. Multi-view graph normalizer network

Graphs (i.e. networks) are used extensively in various fields ranging from drug discovery to computational linguistics. They are also the backbone of many biomedical applications due to their great ability to represent knowledge of interacting entities. Despite their ubiquity, graphs cannot be easily used for machine learning applications since there is no straightforward way to encode their non-Euclidean structure into a feature vector representation. Classical approaches such as handcrafted heuristics (Liben-nowell and Kleinberg, 2003), graph statistics (Bhagat et al., 2011), and kernel functions (Vishwanathan et al., 2008) are widely used to compute

feature vectors for predictive tasks on graphs. However, such approaches treat the graph-driven feature vector estimation (i.e. embedding) process as a preprocessing step, therefore, they generally fail to learn task-optimized graph representations that preserve the graph structure and topology.

The search for data-driven representation learning on graphs ushered in a new deep learning-based approach called graph neural networks (GNNs) (Kipf and Welling, 2017; Defferrard et al., 2016; Xu et al., 2019; Veličković et al., 2018). As other conventional methods, GNNs aims to compute a vector embedding for each node in the graph. However, GNNs are unique in the sense that they learn graph embedding through optimizing a predefined function, therefore, they automatically learn the accurate representation that is specific to the problem without any hand engineering. GNNs have recently revolutionized the field of graph theory and network analysis by generalizing convolutional neural networks (Krizhevsky et al., 2012) (CNNs), which naturally operate on Euclidean data such as images, to graphs. In particular, CNNs slide small learnable filters through the image and at each position of the filters, small patch of pixels are multiplied by the learnable filters to extract some useful local features of the image such as lines and corners. In deeper layers, these local features are then combined to learn more detailed structural features. By analogy to CNNs working principle, GNNs exploit learnable filters to control how each node aggregates information from its local neighborhood and in each layer, hierarchically merges information passed by the previous layer and neighbors to learn a comprehensive vector representation for each node.

Many GNN based methods can be framed in terms of Message Passing Neural Networks (MPNN) paradigm (Gilmer et al., 2017). Let us represent a graph G with node features $\mathbf{v}_i$ and edge features $\mathbf{e}_{ij}$. The forward pass consists of message passing and the readout phase. The message passing phase runs for L times (number of graph convolution layers) and is formulated in terms of message functions M_l and node update functions U_l. In the message-passing

Table 1

Major mathematical notations used in the paper. We denote tensors by boldface Euler script letters, e.g., $\mathcal{X}$. Matrices are denoted by boldface capital letters, e.g., $\mathbf{X}$, vectors are denoted by boldface lowercase letters, e.g., $\mathbf{x}$, scalars and distributions are denoted by lowercase letters, e.g., x, and sets are denoted by uppercase letters e.g., X.

Notation	Definition
n_r	total number of nodes (regions of interests) in the networks
n_v	total number of measurements (network views) for a sample
$\mathcal{T}_s$	subject's tensor representation $\in \mathbb{R}^{n_r \times n_r \times n_v}$
T	Training dataset
$\mathbf{T}_s$	template generated for subject $\mathcal{T}_s$, $\mathbf{T}_s \in \mathbb{R}^{n_r \times n_r}$
$\mathbf{T}_{ref}$	refined template for training dataset
l	index of a graph convolution layer
$\mathbf{e}_{ij}$	cross connectional features between node ij, $\mathbf{e}_{ij} \in \mathbb{R}^{n_r \times n_r}$
F^l	filter-generating network at layer l that maps $\mathbf{e}_{ij}$ to dynamic weights $\Theta_{ij}^l \in \mathbb{R}^{d_l \times d_{l-1}}$
$\mathbf{v}_i^l$	embedding of ith node at layer l
$N(i)$	set of neighbors of node i
S	set of indices of randomly selected training samples for loss computation
$\mathbf{X}_i^v$	v^{th} view of ith random sample.
λ_v	normalization term for view v
t_s	normalized node strength distribution of $\mathbf{T}_s$
x_S^v	ground truth normalized node strength distribution for view v^{th} and randomly selected samples S
D_{KL}	Kullback-Liebler divergence
$\mathcal{L}_{c_s^v}$	centeredness loss for training sample s and view v
$\mathcal{L}_{kl_v^s}$	KL divergence loss for training sample s and view v
$\mathcal{L}_{tcnl_s}$	overall loss computed for the training sample s

phase, node embeddings $\mathbf{v}_i^{l-1}$ are updated based on messages $\mathbf{m}_i^l$ according to:

$$\mathbf{m}_i^l = \text{AGGR}\left\{ M_l\left(\mathbf{v}_i^{l-1}, \mathbf{v}_j^{l-1}, \mathbf{e}_{ij}\right) \right\}_{j \in N(i)} \tag{1}$$

$$\mathbf{v}_i^l = U_l\left(\mathbf{v}_i^{l-1}, \mathbf{m}_i^l\right) \tag{2}$$

where AGGR is a permutation invariant aggregation function such as mean or max, $N(i)$ is the set of neighbors of i in graph G. Next, the readout function R takes final node embeddings $\mathbf{v}_i^L$, and performs the given task. Finally, a loss computed for the output of R, and the network is trained in an end-to-end fashion.

$$R\left(\left\{\mathbf{v}_i^L \mid i \in G\right\}\right) \tag{3}$$

The functions M_l, U_l, and R are all differentiable functions (at least differentiable everywhere except few points in their domain) so that they can be learned via gradient-based optimization. In what follows, we first formalize our multi-view graph normalization problem then explain the components of our architecture in the frame of MPNN.

We propose the MGN-Net to solve the problem of integrating a population of multi-view networks. This problem can be defined as follows. Let sample s be represented by a set of n_v weighted undirected graphs with n_r nodes. We model this sample as a single tensor $\mathcal{T}_s \in \mathbb{R}^{n_r \times n_r \times n_v}$ that is composed of stacked n_v adjacency matrices $\{\mathbf{X}_s^v\}_{v=1}^{n_v}$ of $\mathbb{R}^{n_r \times n_r}$. Our objective is to integrate a set of multi-view graphs $T = \{\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_N\}$ in order to obtain a population-representative connectional template $\mathbf{T} \in \mathbb{R}^{n_r \times n_r}$ that is (1) well-centered, (2) discriminative, and (3) topologically sound. Since there is no general heuristics or conventional methods to meet these three broad constraints in a generic manner, we devised a novel architecture and loss function that learns the best node representation for mapping each sample multi-view graph to a population-representative template using a GNN. We have summarized the major mathematical notations presented in this paper in Table 1.

First we feed sample s represented by a tensor $\mathcal{T}_s \in \mathbb{R}^{n_r \times n_r \times n_v}$ to our MGN-Net architecture (Fig. 2–a). Then holistic embeddings (i.e. representations) are learned for each node through GNN layers (Fig. 2–b). These embeddings encapsulate all the complementary information supplied by the different views thanks to the graph convolution operation. There is an abundant variety of graph

convolution operations. Particularly for our MGN-Net, we chose edge-conditioned convolution (Simonovsky and Komodakis, 2017), a graph convolution tailored to simultaneously handle an arbitrary number of different types of weighted edges, which is essential for MGN-Net to blend connectivity patterns across the multiple graph views of each sample.

Given a multi-view graph, let $\mathbf{e}_{ij} \in \mathbb{R}^{n_v}$ denote cross-view features between node i and j acquired by stacking edge weights for all the available views such as cortical thickness and sulcal depth. In other words, for each sample s, we define the cross-view feature vector $\mathbf{e}_{ij}$ associated with nodes i and j as $[\mathcal{T}_s(i, j, 1), \mathcal{T}_s(i, j, 2), \ldots, \mathcal{T}_s(i, j, n_v)]$. Also, let $l \in \{l_0, \ldots, l_{out}\}$ be the index of a layer in the architecture and d_l the output dimension of the lth layer. Each layer includes a filter-generating network $F^l : \mathbb{R}^{n_v} \mapsto \mathbb{R}^{d_l \times d_{l-1}}$ that takes $\mathbf{e}_{ij}$ as input and outputs edge-specific weight matrix Θ_{ij}^l which dictates the information flow between node i and j. This network F^l allows the MGN-Net model to learn unique filters at each layer for each pair of nodes (i.e., connection) while exploiting the cross-view edge features.

In terms of the MPNN paradigm, the edge-conditioned convolution defines the message passing function M_l as:

$$M_l = F^l(\mathbf{e}_{ij}; \mathbf{W}^l)\mathbf{v}_j^{l-1} + \mathbf{b}^l; F^l(\mathbf{e}_{ij}; \mathbf{W}^l) = \Theta_{ij}^l \tag{4}$$

where $\mathbf{v}_i^l \in \mathbb{R}^{d_l \times 1}$ is the embedding of node i in layer l and $\mathbf{b}^l \in \mathbb{R}^{d_l}$ denotes a network bias and F^l is the defined filter-generating network that maps $\mathbb{R}^{n_v}$ to $\mathbb{R}^{d_l \times d_{l-1}}$ with learnable weights $\mathbf{W}^l$. Furthermore, node embedding update function U_l defined as:

$$U_l = \Theta^l \cdot \mathbf{v}_i^{l-1} + \mathbf{m}_i^l; \mathbf{m}_i^l = \text{AGGR}\left\{ M_l\left(\mathbf{v}_i^{l-1}, \mathbf{v}_j^{l-1}, \mathbf{e}_{ij}\right) \right\}_{j \in N(i)} \tag{5}$$

where $\Theta^l \in \mathbb{R}^{d_l \times d_{l-1}}$ is a learnable parameter, and $N(i)$ denotes the neighbors of node i. Noting that our aggregation schema is averaging, the following operation is performed at each layer l for every node i.

$$\mathbf{v}_i^l = \Theta^l \cdot \mathbf{v}_i^{l-1} + \frac{1}{|N(i)|}\left(\sum_{j \in N(i)} F^l(\mathbf{e}_{ij}; \mathbf{W}^l)\mathbf{v}_j^{l-1} + \mathbf{b}^l\right) \tag{6}$$

Note that F^l can be any type of neural network and vary in each layer depending on the characteristics and complexity of edge weights. Furthermore, $\mathbf{v}_i^0$ corresponds to the initial node features of i so this convolution operation can also utilize the node features.

We note that since brain graphs or connectomes conventionally have no node features, we set entries of the node-specific feature vector $\mathbf{v}_i^0$ to '1' (i.e., identity vector). Next, we generate an edge-specific weight matrix $\boldsymbol{\Theta}_{ij}^l$ by learning the filtering function F^l for each layer l through optimizing our loss function which will be detailed in the next section. Thanks to filtering functions, each edge from node i to node j is associated with a unique weight matrix (generated by F^l from cross-view edge features $\mathbf{e}_{ij}$) which controls the degree of node i's contribution to node j's next layer embedding. Therefore following the first convolution, each node will have a different embedding even though they were identical in the beginning. However, breaking this symmetry of nodes and simultaneously learning distinct node embeddings is a heavy burden to our model. The availability of node features circumvents the need for breaking symmetry, and instead of a noisy input of 1's, it supplies extra information regarding nodes and their roles in the system. Therefore we foresee that given both node and edge features MGN-Net will output more comprehensive connectional templates and converge to an optimum faster.

After applying three layers of edge conditioned convolution separated by rectified linear unit (ReLU), we learn abstract embeddings $\mathbf{V}^{l_{out}} = \left[\mathbf{v}_1^{l_{out}}, \mathbf{v}_2^{l_{out}}, ..., \mathbf{v}_{n_r-1}^{l_{out}}, \mathbf{v}_{n_r}^{l_{out}}\right]^T$ for the multi-view graph nodes (Fig. 2–b). Next, we compute the pairwise relationship of these embeddings to construct the connectional template network (Fig. 2–c). This step constitutes the readout function R of our model. For our case, we simply used absolute difference since our dataset is composed of morphological dissimilarity networks. The intuition behind this operation is our dataset preparation. We derived our brain networks by computing the pairwise absolute difference in cortical measurements between pairs of ROIs. Since we need to map learned node embeddings (abstract ROIs measurements) back to network representation, we simply repeat our dataset preprocessing step. However, such relationship can be computed by any function that allows for backpropagation so that MGN-Net can be trained in an end-to-end fashion. For instance, a simple function such as cosine similarity can be used to measure the similarity between output node embeddings.

2.2. Loss function

Our MGN-Net architecture takes only one sample at a time, however, we aim to output a population-representative connectional template. Since we do not have a ground truth template, to learn the mapping from one sample to the target population-template, we propose to evaluate the output template $\mathbf{T}_s$ that is based on a single sample s against a random subset of the training dataset in the optimization process. In the meantime, we have to preserve the complex topology of the given biological networks while ensuring that the generated template occupies the center of a population by achieving the minimum distance to all population samples (i.e., multi-view networks) to meet our (1) well-centeredness and (3) topological soundness criteria. To address this problem, we present the topology-constrained normalization loss (TCNL) function to evaluate the representativeness of generated templates. TCNL computed against a random subset of training subjects drawn independently for each subject in each epoch. Indices of drawn samples are denoted by the set S. The TCNL loss is composed of a weighted sum of two terms. The first term of the TCNL computes the centredness loss $\mathcal{L}_c$ of the output template and is formalized as follows:

$$\mathcal{L}_{c_s}^v = \sum_{i \in S} \|\mathbf{T}_s - \mathbf{X}_i^v\|_{Frob} \times \lambda_v; \quad \lambda_v = \frac{\frac{1}{\mu_v}}{\max\left\{\frac{1}{\mu_j}\right\}_{j=1}^{n_v}} \tag{7}$$

where $\mathbf{T}_s$ denotes the connectional template for the sample s and $\mathbf{X}_i^v$ is the vth view of ith random sample. μ_v is the mean of connectivity weights of view v and $\max\left\{\frac{1}{\mu_j}\right\}_{j=1}^{n_v}$ is the maximum of mean reciprocals $\frac{1}{\mu_1}$ to $\frac{1}{\mu_{n_v}}$. We include an additional view-specific normalization weight λ_v since the value range of connectional weights for the input graphs might vary largely across views. For example, in our Alzheimer's diseases left hemisphere population (see Section *3.1. Evaluation Datasets*), the mean connectional weight for maximum principal curvature is 0.084 with a min-max range in $[0, 0.586]$ while the mean of cortical thickness is 0.723 with a min-max range in $[0, 3.740]$. Therefore, without a normalization term, our MGN-Net model is most likely to overfit the view with the largest connectional weights as it targets to optimize the defined loss function. Similar problems in the literature are addressed by normalizing the adjacency matrices. For instance, SNF (Wang et al., 2014) divides connectivities in each row by the sum of the entries in that row for normalization; however, this breaks the symmetry in the adjacency matrices of the views, therefore, it is not applicable in our case. Moreover, a simple normalization approach such as min-max scaling can saturate connections at 0 and 1 while standard z-score scaling generates negative connectivities in the network that is counter-intuitive for many types of fully positive networks. Therefore, we introduce λ_v to ensure that the model gives equal attention to each brain view regardless of their value range.

In addition to the centredness loss, we further add a second loss term $\mathcal{L}_{kl}$ to penalize deviations from the topology of the training networks based on Kullback-Leibler divergence of node strength distribution of the generated connectional brain template from the node strength distribution of randomly selected training samples. Let $\mathbf{A}$ denote the adjacency matrix of a graph. We defined the topological strength k_i of node i and the normalized node strength distribution $n(i)$ as follows:

$$k_i = \sum_j \mathbf{A}_{ij}; \quad n(i) = \frac{k_i}{\sum_j k_j} \tag{8}$$

First, we compute the normalized node strength distribution of the generated template for sample s which is denoted by $t_s(i)$. Next, we calculate the ground truth $\left\{x_S^1(i), ..., x_S^{n_v}(i)\right\}$ for n_v views separately by averaging normalized node strength distribution of a random subset S of the training population. We calculate the topological loss $\mathcal{L}_{kl_v}^s$ for subject s and view v as:

$$\mathcal{L}_{kl_v}^s = D_{KL}(t_s||x_S^v) + D_{KL}(x_S^v||t_s)$$
$$= \sum_{i=1}^{n_r} t_s(i) \log_2\left(\frac{t_s(i)}{x_S^v(i)}\right) + \sum_{i=1}^{n_r} x_S^v(i) \log_2\left(\frac{x_S^v(i)}{t_s(i)}\right) \tag{9}$$

Note that the Kullback-Leibler divergence $D_{KL}(P||Q)$ is not a symmetrical function ($D_{KL}(P||Q) \neq D_{KL}(Q||P)$) and defines the information gained by changing beliefs from a prior probability distribution Q to the posterior probability distribution P. Intuitively, P is the true distribution and Q is the estimate. Therefore $D_{KL}(x_S^v||t_s)$ solely is sufficient to represent the topological loss. However, for our datasets using a symmetrical expression $D_{KL}(t_s||x_S^v) + D_{KL}(x_S^v||t_s)$ provides a smoother training.

Given the training population T and MGN-Net with parameters $\left\{\mathbf{W}_1, \mathbf{b}_1, \ldots, \mathbf{W}_{l_{out}}, \mathbf{b}_{l_{out}}\right\}$. We define the TCNL loss $\mathcal{L}_{tcnl_s}$ for subject s and the overall optimization task as follows:

$$\mathcal{L}_{tcnl_s} = \sum_{v=1}^{n_v} \mathcal{L}_{c_s}^v + \beta \times \mathcal{L}_{kl_s}^v; \quad \min_{\mathbf{W}_1, \mathbf{b}_1...\mathbf{W}_{l_{out}}, \mathbf{b}_{l_{out}}} \frac{1}{|T|} \sum_{s=1}^{|T|} \mathcal{L}_{tcnl_s} \tag{10}$$

2.3. Post-training refinement

MGN-Net learns to map heterogeneous views of each subject to a population-representative template. After MGN-Net training, all learned connectional templates represent the population regardless of the input sample used to generate them. However, each template is biased towards its associated training sample s. To eliminate this bias, we suggest an extra step which is executed after the training to obtain more refined and representative templates (Fig. 2–d). First, each subject is fed through the trained MGN-Net in order to get the corresponding template. Then, the most centered connections are selected from these templates by taking the element-wise median. The median operation could also be replaced with other measures of central tendency such as average or truncated mean, however, we used the centeredness score to empirically verify that the median is the most suitable for our case.

$$\mathbf{T}_{ref} = med\left\{\mathbf{T}_1, \mathbf{T}_2, \ldots, \mathbf{T}_{|T|}\right\} \tag{11}$$

$\mathbf{T}_{ref}$ denotes the final refined connectional brain template of the input multi-view graph population.

3. Experiments and material

3.1. Evaluation datasets

We showcase MGN-Net with four different evaluation tests: (1) centeredness, (2) topological soundness, (3) accurate identification of most discriminative connections between two biological groups, and (4) biomarker discovery for neurological diseases. We benchmarked our method against SNF (Wang et al., 2014), netNorm (Dhifallah and Rekik, 2019) and DGN (Gurbuz and Rekik, 2020) (the ablated version of MGN-Net without the $\mathcal{L}_{kl}$ in the loss function) on a small-scale dataset and a relatively large-scale dataset at the neuroscience scale given that brain disorder datasets are quite scarce. The first datasets (AD/LMCI dataset) consists of 77 subjects (41 subjects diagnosed with Alzheimer's diseases (AD) and 36 with Late Mild Cognitive Impairment (LMCI)) from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database GO public dataset (Mueller et al., 2005). Each subject is represented by 4 cortical morphological brain graphs derived from maximum principal curvature, the mean cortical thickness, the mean sulcal depth, and the average curvature measurements (Mahjoub et al., 2018). The second dataset (NC/ASD dataset) is collected from the Autism Brain Imaging Data Exchange ABIDE I public dataset dataset (Di Martino et al., 2014) and includes 310 subjects (155 normal control (NC) and 155 subjects with autism spectral disorder (ASD)) with 6 cortical morphological brain networks extracted from the 4 aforementioned cortical measures in addition to cortical surface area and minimum principle area (Soussia and Rekik, 2017; 2018). For each hemisphere, the cortical surface is reconstructed from T1-weighted MRI using the FreeSurfer (Fischl, 2012) pipeline and parcellated into 35 ROIs using Desikan-Killiany (Desikan et al., 2006) atlas and its corresponding brain network is derived by computing the pairwise absolute difference in cortical measurements between pairs of ROIs (Table 2).

3.2. Hyperparameter setting and training

We trained 8 separate models to generate connectional templates for the right and left hemispheres of four groups namely AD, LMCI, NC, and ASD. We set all the hyperparameters for MGN-Net using a grid search. Each model includes 3 edge-conditioned convolution layers while each layer l contains a shallow neural network F^l with ReLU activation to map 4 (for AD/LMCI dataset) or 6 (for NC/ASD dataset) connectional features obtained from heterogeneous views to $\mathbb{R}^{d_l \times d_{l-1}}$ in order to dynamically learn a

Table 2

Data distribution for AD-LMCI and NC-ASD datasets. Each view in the datasets contains 35 nodes, and views are fully connected. In other words, each includes 1190 connections (35×34, no self connections).

Dataset	AD/LMCI		NC/ASD	
	AD	LMCI	NC	ASD
Number of subjects	41	36	155	155
Male	23	20	124	140
Female	18	16	31	15
Mean Age	75.27	72.54	15.36	16.92
Number of Views	4	4	6	6

unique message-passing filter $\mathbf{\Theta}_{ij}^l \in \mathbb{R}^{d_l \times d_{l-1}}$ for each pair of nodes. Furthermore, edge-conditioned convolution layers output embeddings with 36, 24, and 5 (for AD/LMCI dataset) or 8 (for NC/ASD dataset) dimensions for each node of the input network, respectively. Models are trained using Adam optimizer with a learning rate of 0.0006 and an exponential decay rate of 0.9 for the first moment and 0.99 for the second moment. Since we did not have a GPU memory bottleneck, we executed an update once for the epoch utilizing the gradients computed for the whole dataset (batch size = dataset size). We set the size of the random subset of training samples in our TCNL function to 10 and β which balances the $\mathcal{L}_c$ and $\mathcal{L}_{kl}$ to 25 for AD-LMCI and 10 for NC-ASD. The algorithms are implemented using PyTorch and PyTorch-Geometric (Fey and Lenssen, 2019) frameworks.

We split each group into training and testing sets using 5-fold cross-validation which yields 80% to 20% split. We let each model train for a maximum of 1200 epoch and apply early stopping if there is no improvement in the testing loss for more than 50 epochs. We note that during the training there was neither a significant difference between training and testing losses nor a sign of overfitting due to excessive training. This can be explained by our randomization process during loss calculation which perturbs the training distribution by selecting a different subset for each optimization step. For instance, given our smallest group LMCI (37 subjects), there are more than $\binom{37}{10} \approx 3.4 \times 10^8$ different targets for MGN-Net to fit which makes overfitting much harder compared to conventional learning tasks where targets are fixed.

3.3. Evaluating the Topological soundness of the learned connectional templates

Based on convergent evidence from empirical studies that reveal topological patterns in complex networks (Bullmore and Sporns, 2009; Guimerá et al., 2005; Watts et al., 2006), we suggest that the topology of networks should be preserved during the integration process to obtain more holistic and representative templates. There are many studies investigating the topological features of structural and functional brain networks such as small-world topology, highly connected hubs, and modularity at both the whole-brain scale of human neuroimaging and at a cellular scale in non-human animals (Bullmore and Sporns, 2009). Here we assume that imposing the simple node strength distribution constraint on the MGN-Net training is sufficient to preserve the population topology to a large> extent since more complex topological measures such as PageRank and effective size are derived from node strength.

As for the evaluation of the topological soundness of the learned brain connectional templates, we used additional topological measures including PageRank Page et al., effective size (Burt, 1992), and weighted clustering coefficient (Onnela et al., 2005) measures as they capture different topological properties of graphs. For each measure, we computed the Kullback-Liebler divergence of generated template distribution from the ground truth

distribution ($D_{KL}(g||t)$ where t denotes the template measure distribution and g is the ground truth measure distribution). We applied 5-fold cross-validation and generated connectional templates using the *training samples* using MGN-Net, SNF, netNorm, and DGN, respectively. Next, we computed the aforementioned measures on the connectional templates generated by all three methods. We acquired ground truth by averaging measures that are independently calculated for each view of each *testing sample*. Finally, we reported the average Kullback-Liebler divergence across folds between the generated template distribution and the ground truth distribution. Note that we normalized each distribution using the sum of all node measures to get a valid discrete probability distribution (see *4.3.Topological soundness test* for benchmark results).

3.3.1. PageRank

Originally PageRank algorithm is proposed to measure the importance of website pages based on a graph that is derived from World Wide Web pages and hyperlinks between them. However, it is a general measure which can be applied to investigate graph topologies in various domains including biology and neuroscience (Gleich, 2014). We used the power iteration method for calculation and set the damping parameter and maximum iteration to 0.85 and 100, respectively.

3.3.2. Effective node size

A node's ego network consists of its direct neighbors plus the ties among these neighbors. The effective size of a node measures the non-redundancy of the node's ego network (Burt, 1992). This measure is formulated as:

$$e(i) = \sum_{j \in N(i) - \{i\}} \left(1 - \sum_{k \in N(j)} p_{ik} m_{jk} \right) \quad (12)$$

where $N(i)$ is set of neighbors of i, p_{ik} is the proportion of the weight of the edge connecting node i to node j to the sum of all connection weights of node i. m_{jk} denotes the j's connection with k's divided by the j's strongest connection.

3.3.3. Clustering coefficient

There are multiple ways to define the clustering coefficient for weighted graphs. The one we used for our experiments utilizes the geometric average of edge weights of the triangles through the subject node (Onnela et al., 2005):

$$c(i) = \frac{1}{deg(i)(deg(i) - 1)} \sum_{j,k \in Tri(i)} \left(\hat{w}_{ij} \hat{w}_{ik} \hat{w}_{jk} \right)^{1/3} \quad (13)$$

$deg(i)$ is the degree of node i, $Tri(i)$ is the set of node pairs that form a triangle with node i and $\hat{w}_{ij}$ is the weight of the connection between node i and j normalized by the maximum connectivity weight in the graph.

3.4. Discriminative feature selection and biomarker discovery

Our second criterion is that the generated templates are discriminative which means that the templates will encapsulate the most distinctive traits of the input population. Note that our loss function does not include any term to enforce this criterion and that our model training is performed on a single population. However, through our feature selection and biomarker discovery experiments, we showed that the MGN-Net integration process fulfills this criterion.

Specifically, we propose to compare templates ($\mathbf{T}_A$ and $\mathbf{T}_B$) generated from two groups A and B to identify the most discriminative connections disentangling both groups. To do so, we calculate a discriminativeness score for each graph connection by taking both residual matrix $\|\mathbf{T}_A - \mathbf{T}_B\|_{Frob}$ and absolute alteration ratio matrix

$\max \left(\frac{\mathbf{T}_A}{\mathbf{T}_B}, \frac{\mathbf{T}_B}{\mathbf{T}_A} \right)$ into consideration. Alteration ratio plays a key role in giving equal attention to connections with relatively small weights since the residual matrix alone focuses on the deviation of larger connections. Next, we select the edges (brain connections) with the highest discriminativeness scores:

$$Score_{ij} = \max \left(\frac{\mathbf{T}_{A_{ij}}}{\mathbf{T}_{B_{ij}}}, \frac{\mathbf{T}_{B_{ij}}}{\mathbf{T}_{A_{ij}}} \right) + \alpha \times \left\| \mathbf{T}_{A_{ij}} - \mathbf{T}_{B_{ij}} \right\|_{Frob} \quad (14)$$

Here α is a parameter to balance residual and alteration values for computing the discriminativeness score. We define it as $\frac{2}{\mu_{\mathbf{T}_A} + \mu_{\mathbf{T}_B}}$ where $\mu_{\mathbf{T}_A}$ is the average of all connection weights of connectional template $\mathbf{T}_A$. Note that division by zero may occur during the computation of $\max \left(\frac{\mathbf{T}_A}{\mathbf{T}_B}, \frac{\mathbf{T}_B}{\mathbf{T}_A} \right)$; we simply set alteration ratio to zero for such cases.

For classifying populations (Fig. 3), we first pick the top k connections with the highest discriminativeness scores calculated using two connectional templates generated from the training samples. Next, we concatenate the cross-view edge features $\mathbf{e}_{ij}$ of the selected connections, so we obtain $k \times 4$ features for AD-LMCI populations and $k \times 6$ features for NC-ASD populations. Finally, we train a support vector machine (SVM) classifier on these features to perform binary classification. We present the average accuracy across both 5-folds and k values where $k = \{5, 10, \ldots, 25\}$. Note that we perform 16 classifications in total (4 populations namely, AD-LMCI left hemisphere, AD-LMCI right hemisphere, NC-ASD left hemisphere, and NC-ASD right hemisphere repeated for MGN-Net and three baseline methods) and we independently select the optimal hyperparameters such as kernel type, kernel-specific hyperparameters, regularization parameter C, and tolerance for stopping criterion for each classification task using grid search (over 13600 combinations) by choosing the hyperparameters that yield best average accuracy on the 5-fold cross-validation. SVM classification, cross-validation, and grid search procedures were implemented using Scikit-learn (Pedregosa et al., 2011).

3.5. Remarks on the SNF, netNorm and DGN benchmarks

Originally SNF (Wang et al., 2014) is applied to fuse similarity networks of cancer patients for clustering applications for cancer subtyping. However, it is mathematically demonstrated that SNF can also be used to fuse dissimilarity networks where connections denote the difference between nodes as in our datasets (Dhifallah and Rekik, 2019). Furthermore, SNF is tailored to integrate a population of *single-view* networks whereas our multi-view network integration task operates at two levels: multi-view network integration of individuals and population integration across subjects. Therefore, we adapt SNF to our problem by combining it with averaging. We tried three different pipelines SNF-SNF (SS), Average-SNF (AS), and SNF-Average (SA) where the first step is the merging of the multi-view network for each individual, and the second one is the fusion across subjects. For clarity, we only present the results by SA as it largely outperformed the 3 other SNF-based alternatives across all evaluation datasets. As for netNorm (Dhifallah and Rekik, 2019) and the DGN (ablated version) (Gurbuz and Rekik, 2020), we rely on the publicly available implementations[3] and no adjustments are made since they are directly applicable to our datasets.
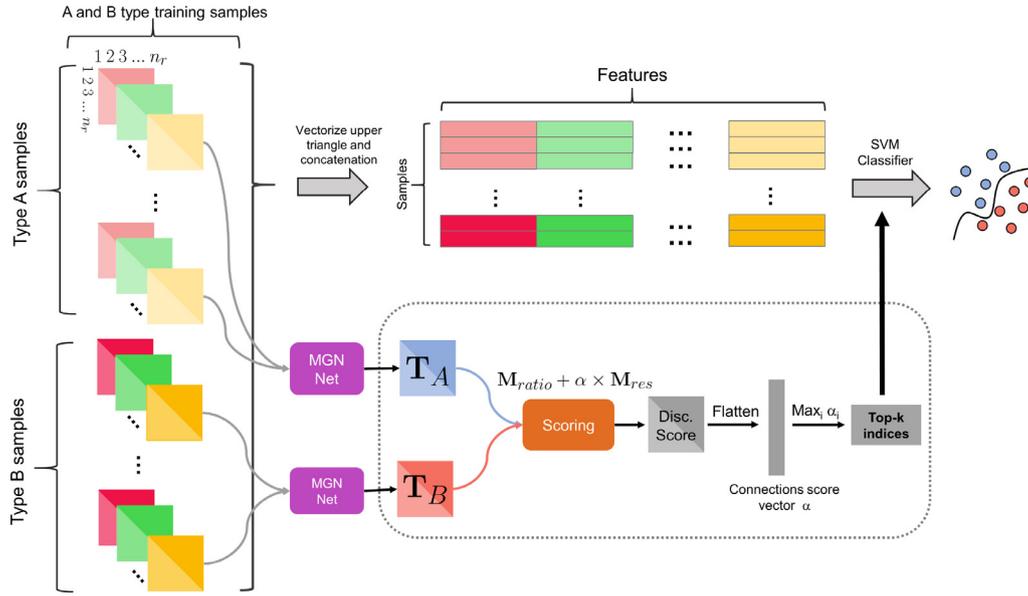
Fig. 3. Discriminative feature selection pipeline for classifying brain networks using templates. First, we generate two distinct connectional templates for type A samples and type B samples. Then we assign a discriminativeness score for each connection using both residual matrix $\|\mathbf{T}_A - \mathbf{T}_B\|_{Frob}$ and absolute alteration ratio matrix $\max\left(\frac{\mathbf{T}_A}{\mathbf{T}_B}, \frac{\mathbf{T}_B}{\mathbf{T}_A}\right)$. Next, we pick the top k connections to train a support vector machine (SVM) model for classifying new testing samples as A or B.
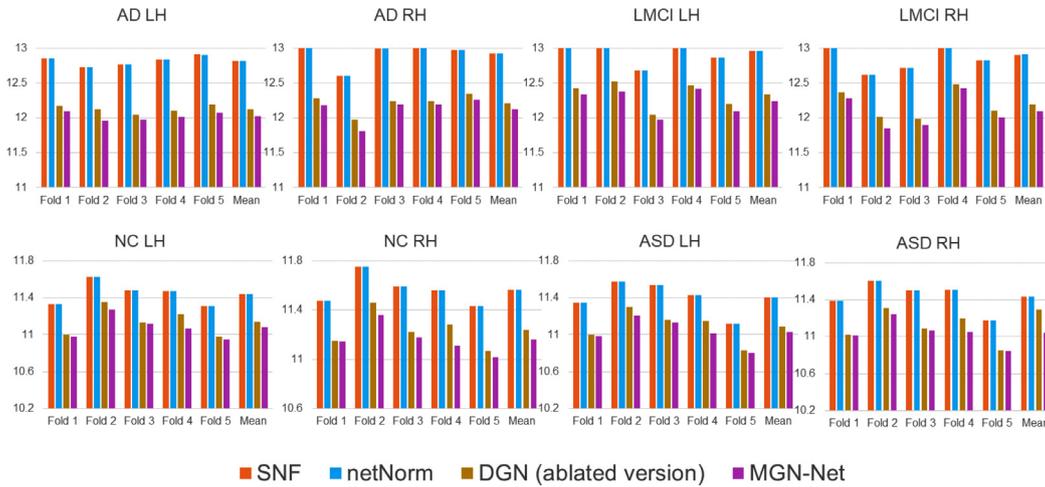
Fig. 4. Centeredness comparison of CBTs generated by SNF (Wang et al., 2014), netNorm (Dhifallah and Rekik, 2019), DGN (ablated version) (Gurbuz and Rekik, 2020) and MGN-Net. Charts illustrate the mean Frobenius distance between the connectional templates created using the training sets and networks of the samples in the testing set using a 5-fold cross-validation strategy. We report the average distance for each cross-validation fold as well as the average across folds ("Mean" bars on the right). MGN-Net achieved the lowest mean Frobenius distance to the population multi-view networks with a high statistical significance demonstrated by a two-tailed paired t-test (all $p > 0.0001$) for MGN-Net-SNF and MGN-Net-netNorm pairs. As for the MGN-Net-DGN pair, all $p > 0.05$ except NC LH and ASD RH. LH: left hemisphere. RH: right hemisphere.

4. Results

4.1. Centeredness test

For reproducibility and generalizability, we split the datasets into training and testing sets using 5-fold cross-validation. We used the training set to generate the CBTs for the population and compute the centeredness on the left out testing fold. To evaluate the centeredness of the integrated networks (CBTs), we measure the mean Frobenius distance between each network view of each sample in the testing left-out fold and the CBT of the four training folds. We note that MGN-Net significantly outperforms other methods across all folds and evaluation datasets (Fig. 4, two-tailed paired t-test, $p > 0.05$, see Table 3 for standard deviation across folds).

Table 3

Standard deviations of Frobenius distances between the connectional brain templates created using the training sets and networks of the samples in the testing set across 5 cross-validation folds.

Dataset	Centeredness standard dev. across folds			
	SNF	netNorm	DGN	MGN-Net
AD LH	0.07	0.07	0.05	0.06
AD RH	0.19	0.18	0.14	0.18
LMCI LH	0.19	0.19	0.2	0.19
LMCI RH	0.28	0.27	0.22	0.25
NC LH	0.13	0.13	0.15	0.13
NC RH	0.12	0.12	0.15	0.12
ASD LH	0.18	0.18	0.18	0.15
ASD RH	0.16	0.16	0.18	0.14

Fig. 5. This chart displays the average centeredness distance across folds of the learned AD and LMCI templates that are biased towards randomly selected 25 samples along with the finalized median-based templates generated by MGN-Net, DGN (ablated version) (Gurbuz and Rekik, 2020), SNF (Wang et al., 2014), and netNorm (Dhifallah and Rekik, 2019). AD: Alzheimer's disease. LMCI: late mild cognitive impairment.

4.2. Centeredness of subject biased templates

We run validation experiments only on the refined connectional templates produced by taking the element-wise median of subject-biased templates to further refine the learned templates by pruning biased connections. Here, we take a look at the centeredness of randomly selected 25 subject-biased CBTs. We note that even the subject-biased templates (i.e. preliminary outputs) learned by the MGN-Net significantly outperform templates that are produced by SNF and netNorm in terms of centeredness (Fig. 5, Fig. 6).

Although it was not the motivation behind the topology-constrained loss function, penalizing deviations from the node strength distributions of the population provided much more consistent subject biased templates. For example, the standard deviation of the centeredness score for randomly selected subject-biased templates (AD LH. population) estimated by DGN is 0.077 whereas

for the MGN-Net it reaches only 0.029. This phenomenon can be explained by the fact that preserving the population topological properties regularizes the subject-biased templates by avoiding motifs and connections that are not repeated across the population.

4.3. Topological soundness test

Network science provides us with tools to quantify the topological properties of networks such as clustering, structural holes, and centrality. It is shown that distinctive topological patterns occur in a wide range of complex networks from biological to social networks (Bullmore and Sporns, 2009; Watts et al., 2006; Guimerá et al., 2005). Therefore, it is essential to preserve these complex but systematic patterns during the biological network data integration process to generate more realistic and integral templates. Thanks to the proposed topology-constrained loss function (TCNL),
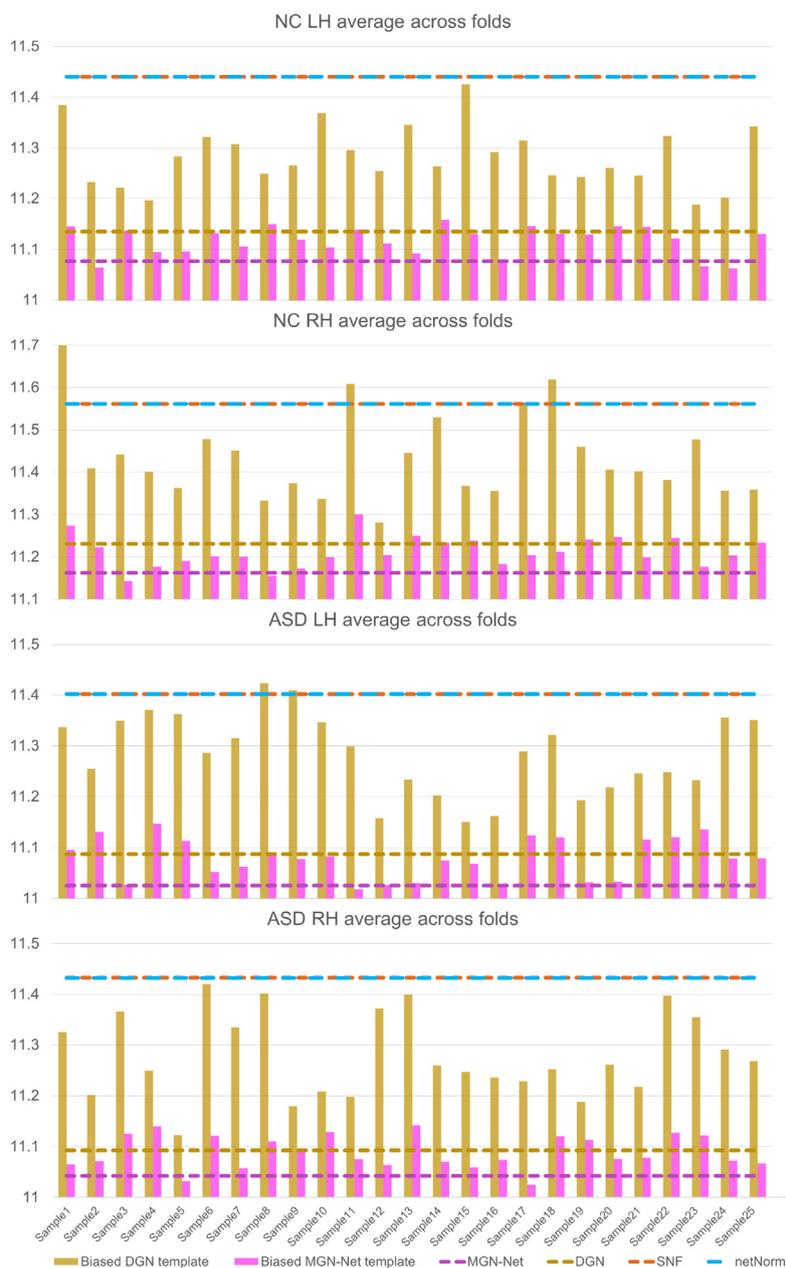
Fig. 6. This chart displays the average centeredness distance across folds of the learned NC and ASD templates that are biased towards randomly selected 25 samples along with the finalized templates generated by MGN-Net, DGN (ablated version) (Gurbuz and Rekik, 2020), SNF (Wang et al., 2014), and netNorm (Dhifallah and Rekik, 2019). NC: normal controls. ASD: autistic spectrum disorder.

we can preserve any topological pattern in a population of multi-view networks when transforming them to a unique holistic connectional template –in a fully generic manner (Fig. 7).

In the context of network neuroscience, we evaluated the topological soundness of the learned CBTs by comparing the discrepancy of the distribution of various topological measures including PageRank Page et al., effective node size (Burt, 1992), and clustering coefficient (Onnela et al., 2005) between the population multi-view brain networks and the estimated CBT. First, we calculated the ground truth by averaging the distribution of topological measures (e.g., clustering coefficient) of each network view of each testing subject. Next, we computed the distribution of topological measures for the CBT estimated using the training dataset. Specifically, each distribution is a discrete probability distribution that is composed of topological measures calculated for each node normalized by the total sum of measures across all nodes. Lastly, we

computed the Kullback-Leibler divergence of the ground truth distributions and distributions derived from the connectional brain templates. These steps were completed with a comprehensive battery of graph topology analysis, to assess the consistency and generalizability of the new TCNL function where we demonstrated that a simple node strength constraint is sufficient for endowing MGN-Net with the ability to capture much more complex topological measures such as PageRank and effective node size. We display the PageRank, effective size and clustering coefficient results in the form of distribution graphs for AD LH population (Fig. 8) and also in a tabular form for all populations (Table 4). Remarkably, the connectional brain template generated by our MGN-Net architecture captures the topology of the connectomic datasets by showing striking similarity with the ground truth while netNorm and SNF fail to preserve the multi-view connectomic data topology. Our statistical analysis using two-tailed paired t-test also demon-
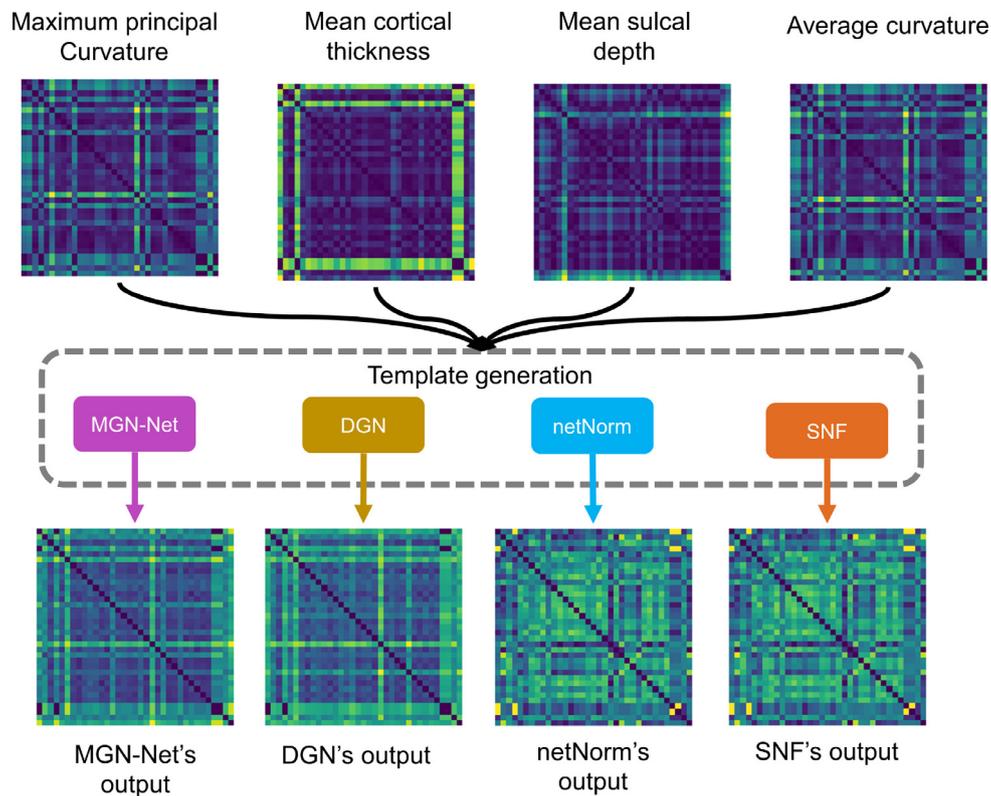
Fig. 7. The figure demonstrates templates estimated by MGN-Net, DGN (ablated version) (Gurbuz and Rekik, 2020), netNorm (Dhifallah and Rekik, 2019), and SNF (Wang et al., 2014) for the left hemisphere of the AD group along with view averages. It is apparent that the templates generated by MGN-Net and DGN encapsulates topological patterns which commonly exist in all views. As for netNorm and SNF case, they capture only a few local motifs across views.

Table 4

Evaluation of deviation from ground truth topology. We report the Kullback-Liebler divergence of the ground truth and learned connectional templates for the PageRank, effective size, and clustering coefficient distributions. * $p < 0.005$ for MGN-Net *vs* SNF, MGN-Net *vs* netNorm and MGN-Net *vs* DGN (ablated version) for all metrics using two-tailed paired t-test.

Dataset	PageRank Dis. *				Effective Size Dis. *				Clustering Coef. *			
	SNF	netNorm	DGN	MGN-Net	SNF	netNorm	DGN	MGN-Net	SNF	netNorm	DGN	MGN-Net
AD LH	0.0552	0.0552	0.0123	**0.0046**	0.0042	0.0041	0.0010	**0.0009**	0.0333	0.0333	0.0099	**0.0049**
LMCI LH	0.0560	0.0560	0.0147	**0.0071**	0.0041	0.0040	0.0011	**0.0010**	0.0338	0.0338	0.0109	**0.0060**
AD RH	0.0578	0.0578	0.0121	**0.0007**	0.0038	0.0037	0.0008	**0.0005**	0.0348	0.0348	0.0099	**0.0029**
LMCI RH	0.0588	0.0588	0.0117	**0.0007**	0.0038	0.0037	0.0008	**0.0005**	0.0354	0.0354	0.0097	**0.0029**
NC LH	0.0599	0.0600	0.0104	**0.0039**	0.0104	0.0104	0.0015	**0.0012**	0.0341	0.0341	0.0114	**0.0055**
ASD LH	0.0573	0.0573	0.0100	**0.0040**	0.0100	0.0100	0.0012	**0.0012**	0.0326	0.0327	0.107	**0.0056**
NC RH	0.0629	0.0630	0.0098	**0.0011**	0.0102	0.0102	0.0013	**0.0008**	0.0361	0.0362	0.0110	**0.0039**
ASD RH	0.0625	0.0625	0.0087	**0.0013**	0.0101	0.0101	0.0010	**0.0007**	0.0359	0.0359	0.0102	**0.0043**

strates that MGN-Net significantly outperformed other methods on both small-scale and large-scale evaluation datasets and across all topological measures (two-tailed paired t-test $p < 0.005$).

4.4. Discriminative feature selection test

Having demonstrated that MGN-Net generates both well-centered and topologically sound connectional templates, we next evaluated its capacity to preserve distinctive features (i.e., edge/connection weights) through the normalization process. We reason that a well-representative connectional template can encapsulate the most unique traits of a population of multi-view networks, which makes it easily distinguishable from other population templates. Those uniquely distinctive features can be used for biological network classification tasks such as differentiating the healthy from the disordered brain network. We designed a simple strategy based on the learned templates that automatically identify the most discriminative network connections distinguishing

between two populations. First, we used integration methods to generate two CBTs for populations *A* and *B*, respectively. Then we assigned a discriminativeness score to each brain connectivity between a pair of ROIs based on the high peaks in the difference (weighted sum of the alteration ratio and magnitude) between the adjacency matrices of connectional templates *A* and *B*. The intuition behind this comparison is driven by our hypothesis that samples belonging to a population *A* lie further away from the CBT of population *B* in comparison with their induced CBT (i.e., from population *A*) and vice versa. Hence, a simple comparison between the connectivity matrices of the CBTs of populations *A* and *B* can easily reveal the most distinctive connectivities to be used in boosting the accuracy of an independent classification method (Fig. 3).

As a proof of concept for the validity of our hypothesis, we further explored the discriminative power of the learned CBTs by our MGN-Net architecture in comparison with benchmark methods. We demonstrated the reproducibility of our classification results against different perturbation of the training and testing sets
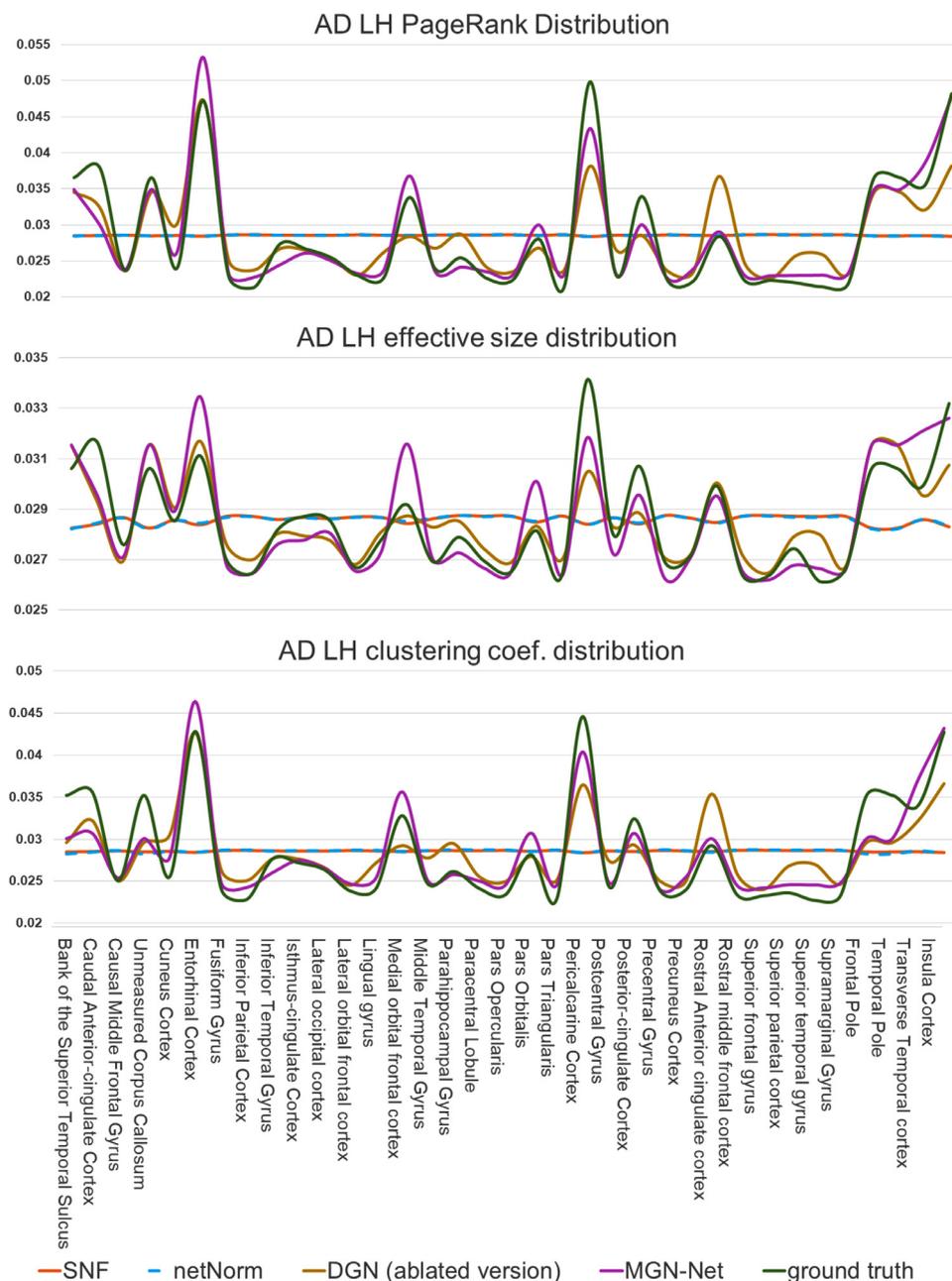
Fig. 8. Comparison of topological distributions of templates generated by SNF (Wang et al., 2014), netNorm (Dhifallah and Rekik, 2019), DGN (ablated version) (Gurbuz and Rekik, 2020), and MGN-Net against the ground truth network population distribution for the AD left hemisphere population. AD: Alzheimer's disease.

using 5-fold cross-validation and across both AD and ASD neurological disorders. We set up 4 different classification tasks namely; AD-LMCI left hemisphere, AD-LMCI right hemisphere, NC-ASD left hemisphere, and NC-ASD right hemisphere. Next, we selected the top k ($k = \{5, 10, \ldots, 25\}$) most discriminative connections revealed by the learned connectional templates of each method and fed them to a support vector machines (SVM) classifier for training. Note that by setting small k values, we are only using less than one percent of the available features. Also, note that brain network classification is a very challenging task that requires deliberately designed preprocessing steps and architectures to achieve satisfactory accuracy. Here, our very modest pipeline that consists of template-based feature selection and SVM is designed to evaluate the templates' capability of identifying important connections. We empirically set the best SVM parameters including the kernels

for each method independently using grid search. Table 5 displays the average accuracy across folds and selected k values. Remarkably, the SVM with training features selected by our MGN-Net outperformed baseline methods on 3 classification tasks –except the AD-LMCI (RH). These results imply that our proposed model not only generates more centered and topologically sound connectional templates but is significantly better at capturing unique traits of multi-view graph populations. MGN-Net outperformance was also replicable in different multi-view connectomic datasets with subtle connectional changes between comparison groups, suggesting that our model could successfully spot integral and holistic brain connections that largely vary across both populations –which leads us to the next series of experiments on population template fingerprinting.

Table 5

SVM binary classification results using features selected by SNF (Wang et al., 2014), net-Norm (Dhifallah and Rekik, 2019), DGN (ablated version) (Gurbuz and Rekik, 2020) and MGN-Net. Average accuracy across 5 folds and *k* values are presented for classifying AD-LMCI and NC-ASD using the left and right hemispheres features separately. The standard deviations σ of average accuracies across k values are indicated in parenthesis.

Dataset	Ave. Accuracy			
	SNF	netNorm	DGN	**MGN-Net**
AD-LMCI LH	70.92 (σ 1.63)	71.37 (σ 0.93)	64.63 (σ 3.30)	**74.23** (σ 2.91)
AD-LMCI RH	**56.72** (σ 0.73)	55.92 (σ 0.93)	53.25 (σ 0.00)	54.05 (σ 1.19)
NC-ASD LH	54.58 (σ 1.06)	54.97 (σ 1.87)	55.94 (σ 0.54)	**56.26** (σ 0.58)
NC-ASD RH	55.45 (σ 0.77)	56 (σ 1.56)	57.16 (σ 1.52)	**58.13** (σ 1.05)
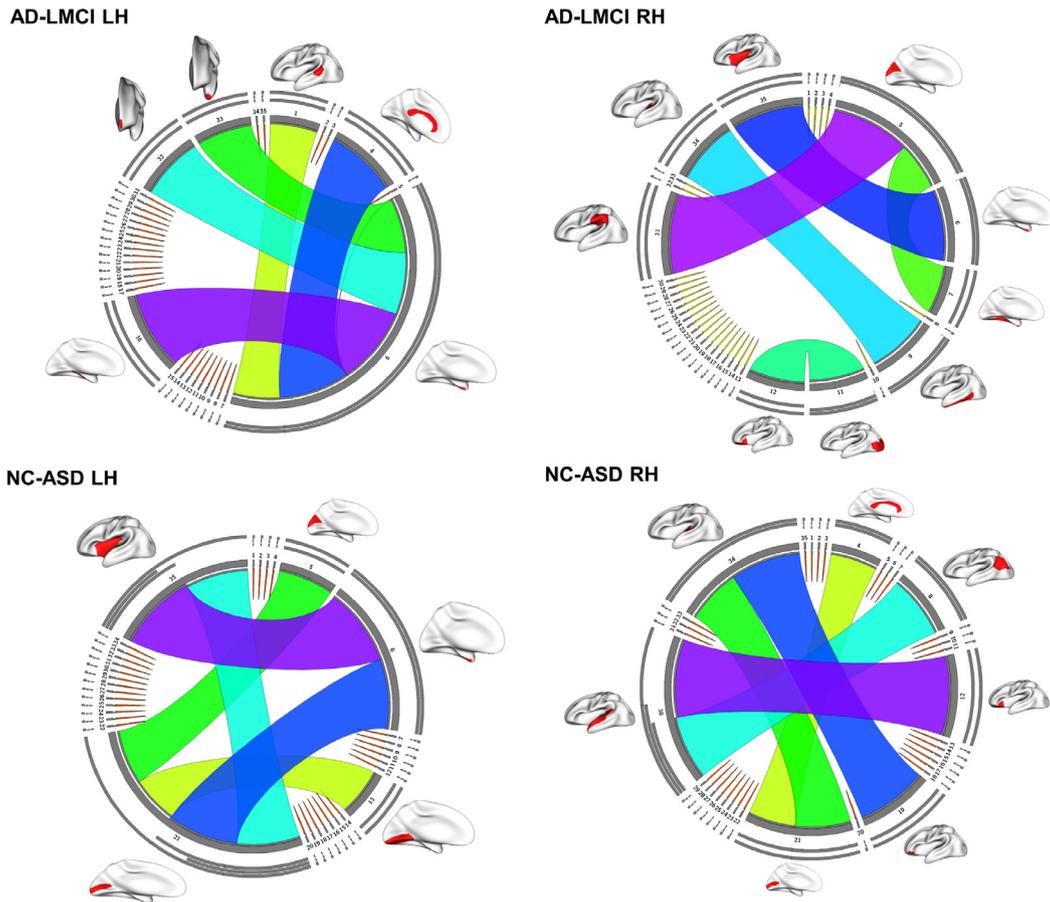
Fig. 9. Circular graphs illustrating the top 5 most discriminative connections that differentiate between two brain states in each pair of groups AD-LMCI and NC-ASD, respectively, for both right and left hemispheres.

4.5. Biomarker discovery for Alzheimer's disease and autism spectrum disorder

Given a particular brain disorder, we further investigated whether the connectional features of the learned disordered CBT, with the largest deviations from the learned healthy CBT, present a connectional fingerprint of the disorder of interest. Specifically, we tested if a naive comparison between disordered and healthy population templates provides a meaningful description of how the multi-view connectional aspect of a brain is altered by a particular disorder. Particularly in this study, we leveraged the CBTs generated by MGN-Net to recognize and decipher the connectional morphological alterations of brain regions fingerprinting Alzheimer's disease and the autism spectrum disorder by selecting the top 5 connections with the highest discriminativeness scores for AD-LMCI and NC-ASD populations, respectively (Fig. 9).

Discovered Alzheimer's disease connectional fingerprint. We discovered that the most pathologically altered connection between brain regions that differentiates AD patients from LMCI patients are mostly clustered around the entorhinal cortex (EC) and cuneus cortex (CC) (Table 6). EC is part of the hippocampal memory system and plays an essential role in memory functions such as memory formation, memory optimization in sleep, and memory consolidation. Therefore, atrophy in EC is likely to be the cause of a significant decline in memory for AD patients when compared to mild symptoms of LMCI (Yassa, 2014; Van Hoesen et al., 1991). Moreover, the deviation of CC which is involved in response inhibition (Crockford et al., 2005) and generating finger movements based on gaze position (Bédard and Sanes, 2008) might explain the AD strong effect on tasks requiring controlled inhibition processes and motor skills. Insights derived from a simple template comparison also align with existing morphological clinical findings about the demented brain. For example, several studies show that morpho-

Table 6
Top 5 discriminative connections discovered for each population pairs.

Rank	AD-LMCI LH	AD-LMCI RH
1	**Entorhinal Cortex** ⟷ Superior Temporal Sulcus	**Cuneus Cortex** ⟷ Fusiform Gyrus
2	**Entorhinal Cortex** ⟷ Unmeasured Corpus Callosum	**Cuneus Cortex** ⟷ Supramarginal Gyrus
3	**Entorhinal Cortex** ⟷ Parahippocampal Gyrus	**Entorhinal Cortex** ⟷ Insula Cortex
4	**Entorhinal Cortex** ⟷ Frontal Pole	Inferior Temporal Gyrus ⟷ Transverse Temporal Cortex
5	**Entorhinal Cortex** ⟷ Temporal Pole	Lateral Occipital Cortex ⟷ Lateral Orbital Frontal Cortex
Rank	NC-ASD LH	NC-ASD RH
1	**Pericalcarine Cortex** ⟷ Cuneus Cortex	**Pericalcarine Cortex** ⟷ Unmeasured Corpus Callosum
2	**Pericalcarine Cortex** ⟷ Entorhinal Cortex	**Superior Temporal Gyrus** ⟷ Inferior Parietal Cortex
3	**Insula Cortex** ⟷ Entorhinal Cortex	**Superior Temporal Gyrus** ⟷ Lateral Orbital Frontal Cortex
4	**Pericalcarine Cortex** ⟷ Lingual Gyrus	**Transverse Temporal Cortex** ⟷ Pars Orbitalis
5	**Pericalcarine Cortex** ⟷ **Insula Cortex**	**Pericalcarine Cortex** ⟷ **Transverse Temporal cortex**

logical atrophy in the entorhinal cortex is the primary biomarker for the conversion of MCI to AD (Lóez et al., 2014; Devanand et al., 2007; Whitwell et al., 2007). Besides, changes in the volume and the cortical thickness of the CC is also extensively reported as an accurate indicator for the conversion of MCI to AD (Wee et al., 2013; Niskanen et al., 2011).

Discovered autism spectrum disorder connectional fingerprint. For the large-scale NC-ASD population, MGN-Net identifies the superior temporal gyrus (STG), transverse temporal gyrus (TTG, or Heschl's gyrus), the insular cortex (IC), and the pericalcarine cortex (PC) as top connectional morphological biomarkers of ASD (Table 6). STG plays an important role (Chang et al., 2010) in auditory, phonetic processing, and social cognition thus, it can explain the receptive language ability deficits which is one of the core features of autism (Bigler et al., 2007). Alteration of TTG which is the first cortical structure that processes the auditory information can be the cause of impaired or delayed language abilities of children with autism (Landa and Garrett-Mayer, 2006; Luyster et al., 2008). Furthermore, it is extensively reported that the atrophy in the IC which governs the processing of empathy (Singer, 2006), norm violations (Sanfey et al., 2003), and emotional (Phan et al., 2002) in the ASD population might be the cause of abnormalities in emotional and affective functions (Yamada et al., 2016a). The alteration of PC which is part of the human visual cortex is likely to be linked to visual symptoms of ASD such as gaze aversion (Mirenda et al., 1983), intense light sensitivity (Jones et al., 2003), and disorganized processing of face stimuli (Pelphrey et al., 2002). Again, from a brain morphology perspective, our connectional ASD blueprint resonates with the existing findings on morphological abnormalities of the autistic brain. For instance, multiple studies showed that a decrease of both white and gray matter volume particularly in the STG (Bonilha et al., 2008) elevated white matter volumes in the TTG (Xiao et al., 2014), significant volumetric increase in IC (Yamada et al., 2016b), and thicker PC (Zielinski et al., 2014) as common morphological traits that disentangle subjects with autism from healthy subjects.

5. Discussion and conclusion

In this work, we proposed the MGN-Net to generate a normalized connectional template that fingerprints a population of multiview networks in an end-to-end manner. MGN-Net is topology-aware thanks to its graph neural network layers and novel loss function that preserves domain-specific topological patterns during the integration process. Moreover, MGN-Net is extremely flexible as it is composed of customizable modules that can be useful to the wide community focused on multi-view network integration or normalization stretching from systems biology to social networks. In general, these features of MGN-Net contrast with other network integration methods since they require a set of strong assumptions such as assuming linear relationships among samples, and neglecting distinctive topological characteristics of complex networks.

From a network neuroscience perspective, we demonstrated that our MGN-Net consistently and significantly outperforms traditional integration methods by generating well-centered, discriminative, and topologically sound connectional templates. Together, our work shows that normalization and integration of multi-view biological graphs can lead to valuable insights by discovering connectional biomarkers that disentangle the typical from the atypical connectivity variability. For example, our connectional brain templates that fingerprint the population of multi-view brain networks derived from T1-weighted MRI scans have revealed a set of biomarkers for both Alzheimer's diseases and the autism spectrum disorder.

Furthermore, conventional methods cannot easily be adapted to more sophisticated network structures such as networks with dynamic connectivity while our flexible architecture needs only small tweaks to operate on any type of network. In fact, our MGN-Net is powerfully generic and easily adaptable in design to different graph-based problems. For instance, geometric recurrent neural networks (RNNs) based on graph convolutional operation can be used to fuse dynamic brain networks derived from MRI measurements acquired at different time points to reveal the trajectory of neurological diseases (Ezzine and Rekik, 2019; Ghribi et al., 2021; Gürler et al., 2020; Nebli et al., 2020). Similar to other methods, MGN-Net assumes that all network views contain the same number of nodes. As MGN-Net is rooted in the powerful and adaptable deep learning mindset, it can be naturally extended to handling non-isomorphic graphs with varying numbers of nodes and local topologies by designing graph up-sampling or down-sampling layers. However, the design and physical interpretation of such up-sampling or down-sampling extensions require domain knowledge and may vary across applications. Therefore, in our future work, we intend to devise a build-in extension to MGN-Net that is agnostic to the number of nodes to overcome this limitation. Furthermore, generated connectional templates strongly rely on the selection of the node embedding relationship function. For our case we used the absolute difference, however, this can be also learned to better model the complex and heterogeneous interactions between graph nodes.

We generated representative connectional templates based on a single population for comparative studies. However, MGN-Net can be further enhanced by adding an *auxiliary classification component* right after the graph convolutional layers to create *targeted* connectional templates for tasks that primarily aim to disentangle two specific groups. For example, gender differences in cortical morphological networks can be studied using tailored *targeted* connectional templates method. However, we anticipate that the classification component can force the model to enhance differences between group connectional templates to boost classification results,

thus the learned templates can be useful to study specific population pairs.

Like other deep learning models, interpretation of GNNs remains a formidable challenge due to their black-box design. A new line of research studies this notorious problem. For instance, (Huang et al., 2020) proposed the GraphLIME that is a generic GNN-model explanation framework that interprets a node by generating a model from its N-hop neighbors and computes the k most representative features as the explanation using a kernel-based feature selection algorithm. Another work (Ying et al., 2019) proposed a model-agnostic approach called GNNExplainer for explaining any graph-based machine learning task. GNNExplainer identifies a compact subgraph (with a constraint size) G_S that minimizes the uncertainty of a trained model on a given task when its GNN computation is limited to G_S. In our future work, we will explore the potential of these methods to improve the interpretability of our MGN-Net architecture. Their explanations may allow us to identify and fix some of the systematic errors made by MGN-Net. More importantly, their feature attribution capabilities can be leveraged to guide our integration process to further ensure that the important connectional traits are preserved during the multi-graph integration task. For example, we can first train an auxiliary GNN classifier to differentiate several brain network populations, then GNNExplainer can be used to generate a subgraph pattern shared between instances that belong to the same populations. Finally, this graph pattern can be used to assign different importance weights to nodes in the MGN-Net's loss function.

In contrast to other baseline network integration tools, one of the major advantages of our method is that it can be trained collaboratively by different parties without sharing private data based on innovative deep learning strategies such as federated learning (Koneçný et al., 2015) and split neural networks (Vepakomma et al., 2018). Furthermore, typically, biomedical data analysis tasks are aggravated by the limited availability of the training data. However, MGN-Net can be trained using transfer learning to achieve good accuracy with limited data. We anticipate that MGN-Net will accelerate the modeling of biological networks and help to ultimately understand the complex dynamics of biological phenomena.

Precomputation and MGN-Net infrastructure

Our publicly available implementation takes a set of stacked adjacency matrices (tensors) as input, however, for efficient graph convolutional layer computation, a couple of preprocessing steps are executed such as casting tensors into different data structures such as edge, feature, and node embedding lists. These precomputation steps run on CPUs and the time and space complexity of these steps depend on the number of subjects, graph nodes as well as the number of views. The main MGN-Net computations including graph convolution operations, pair-wise node embedding relationship computation, and backpropagation are performed on GPUs. MGN-Net training and testing were conducted on a machine with Tesla V100 32GB GPU and Intel Xeon e5-2698 v4 2.2 GHz (20-core) CPU. As for every deep learning architecture, backpropagation is the most computationally demanding step for MGN-Net as well. Particularly, our random subset size hyperparameter is the most important factor affecting this computational cost followed by the number of samples, nodes, and views.

Data availability

The data that support the findings of this study are publicly available from ADNI data (http://adni.loni.usc.edu/). For reproducibility and comparability, the authors will make available upon request all morphological networks generated based on the four cortical attributes (maximum principal curvature, cortical thickness, sulcal depth, and average curvature) for the 77 subjects (41 AD and 36 LMCI) following the approval by ADNI Consortium. Our large-scale dataset is also available from the public ABIDE initiative (http://fcon_1000.projects.nitrc.org/indi/abide/). Following the approval by the ABIDE initiative, all morphological networks generated from the six cortical attributes (cortical surface area and minimum principle area in addition to 4 aforementioned measures) for the 310 subjects (155 NC and 155 ASD) are also accessible from the authors upon request.

Code availability

An open-source Python implementation of MGN-Net is available on GitHub at https://github.com/basiralab/MGN-Net. The release includes a tutorial, notes regarding Python packages, which need to be installed and, connectional brain templates (CBTs) learned from our four datasets by MGN-Net. Users can directly run MGN-Net either on simulated or externally supplied datasets. Information regarding dataset format can be also found in the same repository. Hyperparameters for the model and training routine can be easily tuned by editing a configuration file.

Declaration of Competing Interest

The authors declare no competing interests.

CRediT authorship contribution statement

Mustafa Burak Gürbüz: Methodology, Software, Formal analysis, Validation, Visualization, Writing - original draft, Writing - review & editing. **Islem Rekik:** Conceptualization, Supervision, Methodology, Resources, Writing - review & editing, Funding acquisition.

References

Adams, M., Kelley, J., Gocayne, J., Dubnick, M., Polymeropoulos, M., Xiao, H., Merril, C., Wu, A., Olde, B., Moreno, R., et al., 1991. Complementary dna sequencing: expressed sequence tags and human genome project. Science 252 (5013), 1651–1656. doi:10.1126/science.2047873.

Bassett, D., Sporns, O., 2017. Network neuroscience. Nat. Neurosci. 20, 353–364. doi:10.1038/nn.4502.

Bhagat, S., Cormode, G., Muthukrishnan, S., 2011. Node classification in social networks. Comput. Res. Repos. - CORR doi:10.1007/978-1-4419-8462-3_5.

Bigler, E., Mortensen, S., Neeley, E., Ozonoff, S., Krasny, L., Johnson, M., Lu, J., Provencal, S., McMahon, W., Lainhart, J., 2007. Superior temporal gyrus, language function, and autism. Dev. Neuropsychol. 31, 217–238. doi:10.1080/87565640701190841.

Bonilha, L., Cendes, F., Rorden, C., Eckert, M., Dalgalarrondo, P., Min, L., Steiner, C., 2008. Gray and white matter imbalance - typical structural abnormality underlying classic autism? Brain Dev. 30, 396–401. doi:10.1016/j.braindev.2007.11.006.

Bronstein, M.M., Bruna, J., LeCun, Y., Szlam, A., Vandergheynst, P., 2017. Geometric deep learning: going beyond euclidean data. IEEE Signal Process. Mag. 34 (4), 18–42.

Bullmore, E., Sporns, O., 2009. Complex brain networks: Graph theoretical analysis of structural and functional systems. Nat. Rev. Neurosci. 10, 186–198. doi:10.1038/nrn2575.

Burt, R.S., 1992. Structural Holes: The Social Structure of Competition. Harvard University Press, Cambridge, MA.

Bédard, P., Sanes, J., 2008. Gaze and hand position effects on finger-movement-related human brain activation. J. Neurophysiol. 101, 834–842. doi:10.1152/jn.90683.2008.

Chang, E., Rieger, J., Johnson, K., Berger, M., Barbaro, N., Knight, R., 2010. Categorical speech representation in human superior temporal gyrus. Nat. Neurosci. 13, 1428–1432. doi:10.1038/nn.2641.

Crockford, D., Goodyear, B., Edwards, J., Quickfall, J., el Guebaly, N., 2005. Cue-induced brain activity in pathological gamblers. Biol. Psychiatry 58, 787–795. doi:10.1016/j.biopsych.2005.04.037.

Defferrard, M., Bresson, X., Vandergheynst, P., 2016. Convolutional neural networks on graphs with fast localized spectral filtering. In: Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., Garnett, R. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc., pp. 3844–3852. https://proceedings.neurips.cc/paper/2016/file/04df4d434d481c5bb723be1b6df1ee65-Paper.pdf

Desikan, R., Ségonne, F., Fischl, B., Quinn, B., Dickerson, B., Blacker, D., Buckner, R., Dale, A., Maguire, R., Hyman, B., Albert, M., Killiany, R., 2006. An automated labeling system for subdiving the human cerebral cortex on mri scans into gyral based regions of interest. NeuroImage 31, 968–980. doi:10.1016/j.neuroimage.2006.01.021.

Devanand, D., Pradhaban, G., Liu, X., Khandji, A., De Santi, S., Segal, S., Rusinek, H., Pelton, G., Honig, L., Mayeux, R., Stern, Y., Tabert, M., Leon, M., 2007. Hippocampal and entorhinal atrophy in mild cognitive impairment - prediction of alzheimer disease. Neurology 68, 828–836. doi:10.1212/01.wnl.0000256697.20968.d7.

Dhifallah, S., Rekik, I., 2019. Estimation of connectional brain templates using selective multi-view network normalization. Med. Image Anal. 59, 101567. doi:10.1016/j.media.2019.101567.

Di Martino, A., Yan, C., Li, Q., Denio, E., Castellanos, F., Alaerts, K., Anderson, J., Assaf, M., Bookheimer, S., Dapretto, M., Deen, B., Delmonte, S., Dinstein, I., Ertl-Wagner, B., Fair, D., Gallagher, L., Kennedy, D., Keown, C., Keysers, C., Lainhart, J., Lord, C., Luna, B., Menon, V., Minshew, N., Monk, C., Mueller, S., Müller, R., Nebel, M., Nigg, J., O'Hearn, K., Pelphrey, K., Peltier, S., Rudie, J., Sunaert, S., Thioux, M., Tyszka, J., Uddin, L., Verhoeven, J., Wenderoth, N., Wiggins, J., Mostofsky, S., Milham, M., 2014. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Mol. Psychiatry 19 (6), 659–667. doi:10.1038/mp.2013.78.

Essen, D., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S., Della Penna, S., Feinberg, D., Glasser, M., Harel, N., Heath, A., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., Yacoub, E., 2012. The human connectome project: a data acquisition perspective. NeuroImage 62, 2222–2231. doi:10.1016/j.neuroimage.2012.02.018.

Ezzine, B.E., Rekik, I., 2019. Learning-guided infinite network atlas selection for predicting longitudinal brain network evolution from a single observation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 796–805.

Fey, M., Lenssen, J.E., 2019. Fast graph representation learning with PyTorch geometric. ICLR Workshop on Representation Learning on Graphs and Manifolds.

Fischl, B., 2012. Freesurfer. NeuroImage 62, 774–781. doi:10.1016/j.neuroimage.2012.01.021.

Fout, A., Byrd, J., Shariat, B., Ben-Hur, A., 2017. Protein interface prediction using graph convolutional networks. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 30. Curran Associates, Inc., pp. 6530–6539.

Gainza, P., Sverrisson, F., Monti, F., Rodolá, E., Boscaini, D., Bronstein, M., Correia, B., 2020. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. Nat. Methods 17, 1–9. doi:10.1038/s41592-019-0666-6.

Ghribi, O., Li, G., Lin, W., Shen, D., Rekik, I., 2021. Multi-regression based supervised sample selection for predicting baby connectome evolution trajectory from neonatal timepoint. Med. Image Anal. 68, 101853.

Gilmer, J., Schoenholz, S.S., Riley, P.F., Vinyals, O., Dahl, G.E., 2017. Neural message passing for quantum chemistry. In: Proceedings of the 34th International Conference on Machine Learning, pp. 1263–1272.

Gleich, D., 2014. Pagerank beyond the web. SIAM Rev. 57. doi:10.1137/140976649.

Guimerá, R., Mossa, S., Turtschi, A., Amaral, L., 2005. The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. Proc. Natl. Acad. Sci. U. S. A. 102, 7794–7799. doi:10.1073/pnas.0407994102.

Gurbuz, M.B., Rekik, I., 2020. Deep graph normalizer: A geometric deep learning approach for estimating connectional brain templates. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 155–165.

Gürler, Z., Nebli, A., Rekik, I., 2020. Foreseeing brain graph evolution over time using deep adversarial network normalizer. In: International Workshop on PRedictive Intelligence in MEdicine, pp. 111–122.

Huang, Q., Yamada, M., Tian, Y., Singh, D., Yin, D., Chang, Y., 2020. Graphlime: local interpretable model explanations for graph neural networks.

Ideker, T., Galitski, T., Hood, L., 2001. A new approach to decoding life: Systems biology. Annu. Rev. Genom. Hum. Genet. 2, 343–372. doi:10.1146/annurev.genom.2.1.343.

Jones, R., Quigney, C., Huws, J., 2003. First-hand accounts of sensory perceptual experiences in autism: a qualitative analysis. J. Intel. Dev.Disabil. 28(2), 112–121. J. Intel. Dev.Disabil. 28, 112–121. doi:10.1080/1366825031000147058.

Kanehisa, M., Goto, S., 2000. Kegg: kyoto encyclopedia of genes and genomes. Nucl. Acids Res. 28, 27–30. doi:10.1093/nar/28.1.27.

Kipf, T.N., Welling, M., 2017. Semi-supervised classification with graph convolutional networks. In: Proceedings of the 5th International Conference on Learning Representations. In:

Konečný, J., McMahan, B., Ramage, D., 2015. Federated optimization: distributed optimization beyond the datacenter. arXiv: 1511.03575.

Krizhevsky, A., Sutskever, I., Hinton, G., 2012. Imagenet classification with deep convolutional neural networks. Neural Inf. Process. Syst. 25. doi:10.1145/3065386.

Landa, R., Garrett-Mayer, E., 2006. Development in infants with autism spectrum disorders: a prospective study. J. Child Psychol. PsychiatryAllied Discip. 47, 629–638. doi:10.1111/j.1469-7610.2006.01531.x.

Lee, D.-S., Park, J., Kay, K., Christakis, N., Oltvai, Z., Barabási, A.-L., 2008. The implications of human metabolic network topology for disease comorbidity. Proc. Natl.. Acad. Sci. U. S. A. 105, 9880–9885. doi:10.1073/pnas.0802208105.

Liben-nowell, D., Kleinberg, J., 2003. The link prediction problem for social networks. J. Am. Soc. Inf.Sci. Technol. 58. doi:10.1002/asi.20591.

Luyster, R., Kadlec, M., Carter, A., Tager-Flusberg, H., 2008. Language assessment and development in toddlers with autism spectrum disorders. J. AutismDev. Disord. 38, 1426–1438. doi:10.1007/s10803-007-0510-1.

López, M., Bruña, R., Aurtenetxe, S., Pineda-Pardo, J., Marcos, A., Arrazola, J., Reinoso, A., Pedro, M., Bajo, R., Maestú, F., 2014. Alpha-band hypersynchronization in progressive mild cognitive impairment: a magnetoencephalography study. J. Neurosci. 34, 14551–14559. doi:10.1523/JNEUROSCI.0964-14.2014.

Mahjoub, I., Mahjoub, M., Rekik, I., 2018. Brain multiplexes reveal morphological connectional biomarkers fingerprinting late brain dementia states. Sci. Rep. 8. doi:10.1038/s41598-018-21568-7.

Mirenda, P., Donnellan, A.M., Yoder, D.E., 1983. Gaze behavior: a new look at an old problem. J. Autism Dev. Disord. 13, 397–409.

Mueller, S., Weiner, M., Thal, L., Petersen, R., Jack, C., Jagust, W., Trojanowski, J., Toga, A., Beckett, L., 2005. The Alzheimer's disease neuroimaging initiative. Neuroimaging Clin. N. Am. 15, 869–877. doi:10.1016/j.nic.2005.09.008. xi

Nebli, A., Kaplan, U.A., Rekik, I., 2020. Deep evographnet architecture for time-dependent brain graph data synthesis from a single timepoint. In: International Workshop on PRedictive Intelligence in MEdicine, pp. 144–155.

Niskanen, E., Könönen, M., Määttä, S., Hallikainen, M., Kivipelto, M., Casarotto, S., Massimini, M., Vanninen, R., Mervaala, E., Karhu, J., Soininen, H., 2011. New insights into Alzheimer's disease progression: acombined tms and structural mri study. PloS One 6, e26113. doi:10.1371/journal.pone.0026113.

Onnela, J.-P., Saramäki, J., Kertész, J., Kaski, K., 2005. Intensity and coherence of motifs in weighted complex networks. Phys. Rev. E Stat. NonlinearSoft Matter Phys. 71, 065103. doi:10.1103/PhysRevE.71.065103.

Page, L., Brin, S., Motwani, R., Winograd, T.,. The pagerank citation ranking: bringing order to the web. In: Proceedings of the 7th International World Wide Web Conference. Brisbane, Australia, pp. 161–172.

Parisot, S., Ktena, S.I., Ferrante, E., Lee, M., Guerrero, R., Glocker, B., Rueckert, D., 2018. Disease prediction using graph convolutional networks: application to autism spectrum disorder and Alzheimer's disease. Med. Image Anal. 48. doi:10.1016/j.media.2018.06.001.

Pearl, J., 1988. In: Pearl, J. (Ed.), Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann, San Francisco (CA) doi:10.1016/B978-0-08-051489-5.50007-2. http://www.sciencedirect.com/science/article/pii/B9780080514895500072

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Pelphrey, K., Sasson, N., Reznick, J., Paul, G., Goldman, B., Piven, J., 2002. Visual scanning of faces in autism. J. AutismDev. Disord. 32, 249–261. doi:10.1023/A:1016374617369.

Phan, K.L., Wager, T.D., Taylor, S.F., Liberzon, I., 2002. Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in pet and fmri. NeuroImage 16, 331–348.

Rhee, S., Seo, S., Kim, S., 2018. Hybrid approach of relation network and localized graph convolutional filtering for breast cancer subtype classification. arXiv: 1711.05859

Safari, N., Taghizadeh, M., Rezaei tavirani, M., Goliaei, B., Peyvandi, A., 2014. Protein-protein interaction networks (PPI) and complex diseases. Gastroenterol. Hepatol. BedBench 7, 17–31.

Salathé, M., Kazandjieva, M., Lee, J.W., Levis, P., Feldman, M., Jones, J., 2010. A high-resolution human contact network for infectious disease transmission. Proc. Nat Acad. Sci. U. S. A. 107, 22020–22025. doi:10.1073/pnas.1009094108.

Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L., Cohen, J.D., 2003. The neural basis of economic decision-making in the ultimatum game.. Science 300 5626, 1755–1758.

Simonovsky, M., Komodakis, N., 2017. Dynamic edge-conditioned filters in convolutional neural networks on graphs. CVPR 1704.02901.

Singer, T., 2006. The neuronal basis and ontogeny of empathy and mind reading: Review of literature and implications for future research. Neurosci. Biobehav. Rev. 30, 855–863. doi:10.1016/j.neubiorev.2006.06.011.

Soussia, M., Rekik, I., 2017. High-order connectomic manifold learning for autistic brain state identification. In: International Workshop on Connectomics in Neuroimaging, pp. 51–59.

Soussia, M., Rekik, I., 2018. Unsupervised manifold learning using high-order morphological brain networks derived from T1-w MRI for autism diagnosis. Front. Neuroinformatics 12.

Van Essen, D.C., Glasser, M.F., 2016. The human connectome project: progress and prospects. Cerebrum 2016.

Van Hoesen, G., Hyman, B., Damasio, A., 1991. Entorhinal cortex pathology in Alzheimer's disease. Hippocampus 1 (1), 1–8. doi:10.1002/hipo.450010102.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lió, P., Bengio, Y., 2018. Graph attention networks. In: International Conference on Learning Representations.

Vepakomma, P., Gupta, O., Swedish, T., Raskar, R., 2018. Split learning for health: distributed deep learning without sharing raw patient data.

Vishwanathan, S., Borgwardt, K., Kondor, R., Schraudolph, N., 2008. Graph kernels. J. Mach. Learn. Res. 11.

Wang, B., Mezlini, A., Demir, F., Fiume, M., Tu, Z., Brudno, M., Haibe-Kains, B., Goldenberg, A., 2014. Similarity network fusion for aggregating data types on a genomic scale. Nat. Methods 11. doi:10.1038/nmeth.2810.

Watts, D., J, D., Strogatz, H, S., 2006. Collective Dynamics of 'Small World' Networks, pp. 301–303.

Wee, C.-Y., Yap, P.-T., Shen, D., 2013. Prediction of alzheimer's disease and mild cognitive impairment using cortical morphological patterns. Hum. Brain Mapp. 34. doi:10.1002/hbm.22156.

Whitwell, J., Przybelski, S., Weigand, S., Knopman, D., Boeve, B., Petersen, R., Jack, C., 2007. 3d maps from multiple MRI illustrate changing atrophy patterns as subjects progress from mild cognitive impairment to alzheimer's disease. Brain 130, 1777–1786. doi:10.1093/brain/awm112.

Xiao, Z., Qiu, T., Ke, X., Xiao, X., Xiao, T., Liang, F., Zou, B., Huang, H., Fang, H., Chu, K., Zhang, J., Liu, Y., 2014. Autism spectrum disorder as early neurodevelopmental disorder: evidence from the brain imaging abnormalities in 2–3 years old toddlers. J. Autism Dev. Disord. 44. doi:10.1007/s10803-014-2033-x.

Xu, K., Hu, W., Leskovec, J., Jegelka, S., 2019. How powerful are graph neural networks?arXiv: 1810.00826.

Yamada, T., Itahashi, T., Nakamura, M., Watanabe, H., Kuroda, M., Ohta, H., Kanai, C., Kato, N., Hashimoto, R.-i., 2016. Altered functional organization within the insular cortex in adult males with high-functioning autism spectrum disorder: evidence from connectivity-based parcellation. Mol. Autism 7. doi:10.1186/s13229-016-0106-8.

Yamada, T., Itahashi, T., Nakamura, M., Watanabe, H., Kuroda, M., Ohta, H., Kanai, C., Kato, N., Hashimoto, R.-i., 2016. Altered functional organization within the insular cortex in adult males with high- functioning autism spectrum disorder: evidence from connectivity-based parcellation. Mol. Autism 7. doi:10.1186/s13229-016-0106-8.

Yassa, M., 2014. Ground zero in alzheimer's disease. Nat. Neurosci. 17, 146–147. doi:10.1038/nn.3631.

Ying, R., Bourgeois, D., You, J., Zitnik, M., Leskovec, J., 2019. Gnnexplainer: generating explanations for graph neural networks. Adv. Neural Inf. Process. Syst. 32, 9240–9251.

Zielinski, B., Prigge, M., Nielsen, J., Froehlich, A., Abildskov, T., Anderson, J., Fletcher, P., Zygmunt, K., Travers, B., Lange, N., Alexander, A., Bigler, E., Lainhart, J., 2014. Longitudinal changes in cortical thickness in autism and typical development. Brain 137. doi:10.1093/brain/awu083.