Metric Learning with Spectral Graph Convolutions on Brain Connectivity Networks

Sofia Ira Ktena[a,*], Sarah Parisot[a], Enzo Ferrante[a,b], Martin Rajchl[a], Matthew Lee[a], Ben Glocker[a], Daniel Rueckert[a]

[a] *Biomedical Image Analysis Group, Imperial College London, UK*
[b] *Universidad Nacional del Litoral / CONICET, Santa Fe, Argentina*

Abstract

Graph representations are often used to model structured data at an individual or population level and have numerous applications in pattern recognition problems. In the field of neuroscience, where such representations are commonly used to model structural or functional connectivity between a set of brain regions, graphs have proven to be of great importance. This is mainly due to the capability of revealing patterns related to brain development and disease, which were previously unknown. Evaluating similarity between these brain connectivity networks in a manner that accounts for the graph structure and is tailored for a particular application is, however, non-trivial. Most existing methods fail to accommodate the graph structure, discarding information that could be beneficial for further classification or regression analyses based on these similarities. We propose to learn a graph similarity metric using a siamese graph convolutional neural network (s-GCN) in a supervised setting. The proposed framework takes into consideration the graph structure for the evaluation of similarity between a pair of graphs, by employing spectral graph convolutions that allow the generalisation of traditional convolutions to irregular graphs and operates in the graph spectral domain. We apply the proposed model on two datasets: the challenging ABIDE database, which comprises functional MRI data of 403 patients with autism spectrum disorder (ASD) and 468 healthy

*Corresponding author. *Email address*: ira.ktena@imperial.ac.uk

controls aggregated from multiple acquisition sites, and a set of 2,500 subjects from UK Biobank. We demonstrate the performance of the method for the tasks of classification between matching and non-matching graphs, as well as individual subject classification and manifold learning, showing that it leads to significantly improved results compared to traditional methods.

Keywords: functional brain connectivity, spectral graph convolutions, convolutional neural networks, autism spectrum disorder, UK Biobank

1. Introduction

During the last decade, there has been increasing interest in the study of the human connectome (Sporns, 2011), which involves representing a set of brain regions along with their structural and/or functional interactions as networks. These brain connectivity networks result from the subdivision of the brain into regions using anatomical landmarks, cytoarchitecture or function (Arslan et al., 2017). Their topological properties have been thoroughly explored in recent neuroscience studies (Achard et al., 2006; Rubinov and Sporns, 2010). This is primarily due to the fact that associations between the topological organisation of these networks and brain development (Hagmann et al., 2010), function (Smith et al., 2015) as well as disease (Catani and ffytche, 2005; Fornito et al., 2015) have been established. Recent advances in neuroimaging have led to significant improvements in the spatial resolution of functional Magnetic Resonance Imaging (fMRI). Therefore, resting-state fMRI (rs-fMRI) is currently one of the most widespread approaches to map the putative connections between spatially remote brain regions by means of correlations between their corresponding time series. The obtained functional connectivity networks incorporate the strength of these connections in their edge labels (Sporns, 2013), yielding a so-called labelled graph representation.

Apart from the study of statistically significant group differences with network theoretical approaches (Rudie et al., 2013), brain networks, or graphs in mathematical terms, can be studied at a subject level in order to identify

distinctive patterns related to brain disease or development (Kawahara et al., 2017). In this context, disruptions to the functional network organisation of the human brain have been associated with neurological disorders, such as attention deficit hyperactivity disorder (ADHD) (Konrad and Eickhoff, 2010) and autism spectrum disorder (ASD) (Abraham et al., 2017). These findings suggest that the study of functional brain organisation has the potential to identify predictive biomarkers for neurodevelopmental and neuropsychiatric disorders and shed light on the disorder's underlying mechanisms. At the same time, sex- (Satterthwaite et al., 2014) and age- (Geerligs et al., 2014) related differences in functional connectivity networks have also been reported. The above are common examples of classification and regression problems, which can benefit from an accurate measure of similarity between the network representations to allow for the application of statistical and machine learning methods. Automatically learning meaningful pairwise similarities from graph-structured data is, additionally, very important in applications like graph-based label propagation (Wauquier and Keller, 2015; Parisot et al., 2017).

1.1. Inexact graph matching

The problem of evaluating how similar or different two graphs are can be addressed through inexact graph matching (Livi and Rizzi, 2013). These approaches estimate (dis)similarity between two graphs at a global scale and yield a meaningful pairwise metric that can further facilitate classification, regression and clustering applications. In these settings, the estimation of (dis)similarity between a pair of graphs has, most commonly, been dealt with using one of the following approaches (Livi and Rizzi, 2013): graph embedding, graph kernels, motif counting and graph edit distance. In graph embedding techniques, a feature vector representation is used to summarise either the complete set of network edges or its topology in terms of well-known network features, e.g. nodal strength and degree, network efficiency and modularity. The brain network representations obtained with these techniques can, then, be directly fed into traditional classification and regression algorithms in a straightforward manner.

Hence, graph embedding has been widely used to estimate brain network similarity (Dosenbach et al., 2010; Richiardi et al., 2011; Zeng et al., 2012; Abraham et al., 2017), although it often discards valuable information about the graph structure. Graph kernels, e.g. random walk kernels or the Weisfeiler-Lehman graph kernel (Shervashidze et al., 2011), have been employed to compare functional connectivity networks (Mokhtari and Hossein-Zadeh, 2013; Jie et al., 2014; Takerkart et al., 2014), but often fail to capture global properties as they compare features of smaller subgraphs. Motif counting, in turn, is a computationally expensive process, since it involves counting the occurrences of important recurring subgraph patterns (Shervashidze et al., 2009). Methods based on graph edit distance neatly model both structural and label variation within the graphs and are particularly useful to identify unknown node correspondences in brain connectivity networks (Raj et al., 2010; Ktena et al., 2017a), but are limited by the a priori definition of the edit costs. Automated methods have been proposed to address this problem of manually defining the cost functions, e.g. by utilising a Gaussian mixture model (GMM) (Neuhaus and Bunke, 2007) or self-organising maps (SOMs) (Neuhaus and Bunke, 2005), to learn the edit operation costs from the data samples. These early works paved the way for performing metric learning directly in the graph domain.

1.2. Non-Euclidean Convolutional Neural Networks

Previous works by Zagoruyko and Komodakis (2015) and Kumar et al. (2016) on the comparison of image patches explored different neural network models, including siamese and triplet convolutional networks, to learn a similarity metric. These network architectures employed standard 2D convolutions to yield hierarchies of image features and model the different factors that affect the final appearance of 2D images. However, the application and generalisation of convolutions to graph-structured data and irregular domains, such as brain connectivity networks, is not straightforward. This research topic has recently attracted a lot of attention and relevant work is focusing on the challenging problem of defining a local neighbourhood structure, which is required for convolution op-

erations (Henaff et al., 2015; Niepert et al., 2016; Masci et al., 2016). Niepert et al. (2016) attempted to address this challenge by employing a graph labelling procedure for the selection of a node sequence that served as a receptive field, but the labelling function limited node features to categorical, non-continuous values. An important step forward has been the introduction of the concept of signal processing on graphs by Shuman et al. (2013), which allowed to perform data processing tasks, like filtering, through the use of computational harmonic analysis. The extension of CNNs to irregular graphs was, then, rendered feasible by formulating convolutions in the graph spatial domain as multiplications in the graph spectral domain. Defferrard et al. (2016) relied on this property to define strictly localised filters by means of Chebyshev polynomials and employed a recursive formulation that allows fast filtering operations. Kipf and Welling (2016) later introduced a renormalisation trick to speedup computations, while Levie et al. (2017) proposed Cayley polynomial filters, a new class of parametric rational complex functions, to compute localised regular filters.

Alternative methods apply convolutions directly in the graph spatial domain, rather than the graph spectral domain. These approaches include using anisotropic heat kernels to extract intrinsic patches on manifolds (Boscaini et al., 2016) or, alternatively, representing local patches in polar coordinates (Masci et al., 2015). A recent work by Simonovsky and Komodakis (2017) proposed to use filter weights conditioned on the edge labels, instead, and dynamically generate those for each input sample. This group of methods addresses the problem of generalisation across different domains and have clear geometric interpretations on manifolds, where they have so far been applied. However, their interpretation on generic graphs is not straightforward neither as principled as the spectral approaches.

1.3. Contributions

In Ktena et al. (2017b) we proposed a novel method for learning a similarity metric between irregular graphs and demonstrated its potential in a classification task on functional connectivity networks with known node correspondences.

This method employs spectral graph convolutions to identify patterns that are significant for the estimation of similarity between a pair of graphs, allowing the learned metric to properly accommodate the graph structure. Our hypothesis is that a more accurate similarity metric of connectomes can be obtained when taking into account their graph structure, instead of operating on their vectorised equivalent. Additionally, learning the similarity function is considered beneficial when working with brain connectivity networks, since it allows fine-tuning the metric to a particular application, in contrast to more traditional approaches like euclidean distances or graph kernels that are commonly used in connectomics research. We used a siamese graph convolutional neural network that employed the polynomial filters formulated in Defferrard et al. (2016) and a global loss function that, according to Kumar et al. (2016), is robust to outliers and provides better regularisation. To the best of our knowledge, this has been the first application of metric learning with spectral graph convolutions on brain connectivity networks. In this work, we extend our preliminary study, exploring the impact of the different framework components and objective functions in a cross-validation setting. More specifically:

- We propose a modification to the global loss function that leads to better generalisation in applications on heterogeneous data, like the ABIDE database used in this work.

- We provide an extended validation on two large databases by 1) evaluating the influence of different loss functions and 2) comparing our approach to alternative methods that can be used for similarity estimation between pairs of graphs. We show that the learned similarities are more accurate for classification and manifold learning.

- We demonstrate the model performance on **two** different databases; the functional connectivity networks of 871 subjects from the challenging Autism Brain Imaging Data Exchange (ABIDE) database (Di Martino et al., 2014), which contains heterogeneous rs-fMRI data acquired at multiple international sites with different protocols, as well as a larger dataset

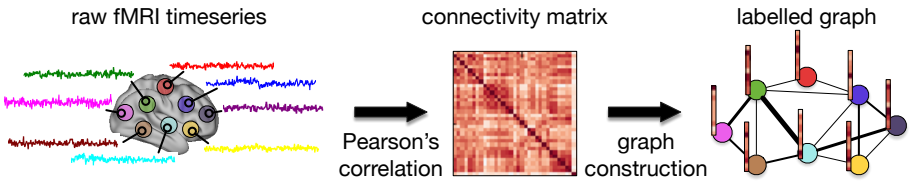
of 2,500 subjects from the UK Biobank (Sudlow et al., 2015). Our goal is to distinguish between patients with autism spectrum disorder (ASD) and healthy controls (HC) in the ABIDE case, while for UK Biobank the task is to distinguish between male and female subjects.

The learned metric leads to 12.9% better accuracy compared to baseline methods for disease classification with the ABIDE database and a 10.3% improvement for gender classification with UK Biobank compared to a metric learning approach operating on the vectorised connectivity matrices. Our implementations can be found online at `https://github.com/sk1712/gcn_metric_learning`.

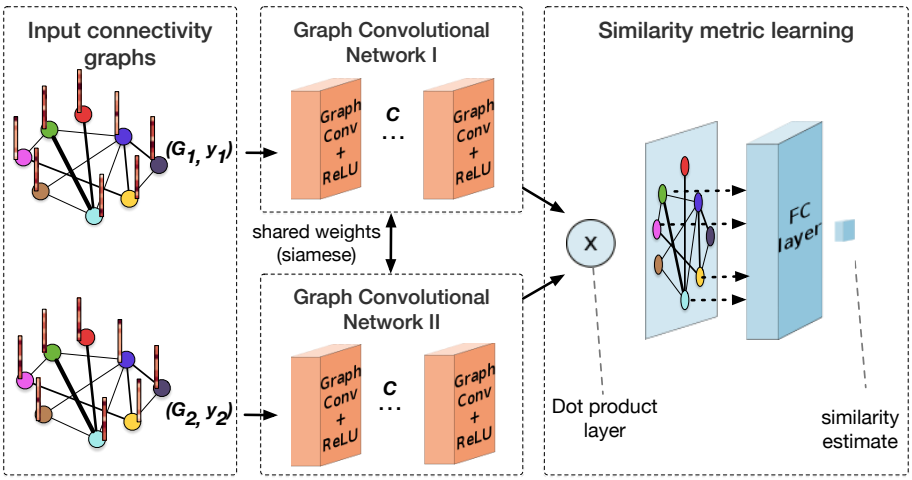
2. Materials and Methods

Figure 1 gives an overview of the proposed model for similarity metric learning on functional brain networks. In this section, we first present the datasets used and the process through which functional brain networks are derived from the original fMRI image data in 2.1. Additionally, we introduce the concept of graph convolutions and describe how the filtering operation can be performed in the graph spectral domain in 2.3. Finally, we describe the proposed network architecture in 2.4 and the loss functions explored as alternative objective functions in 2.5.

We introduce below a set of notations that will be used throughout this paper. We perform metric learning on N subjects using a siamese neural network with C graph convolutional layers. Each subject s is represented by a labelled graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where each node $v_i \in \mathcal{V}$ corresponds to a brain ROI and is associated with a signal $c_{si} : v_i \rightarrow \mathbb{R}^R$ containing the node's functional connectivity profile for an atlas with R regions. We denote the normalised graph Laplacian as L, and the i^{th} filter of the j^{th} layer parametrised on the Laplacian as $g_{\theta_{i,j}}(L)$.



(a) Estimation of single subject connectivity matrix and labelled graph representation. Pearson's correlation is used to obtain a functional connectivity matrix from the raw fMRI timeseries. After specifying the graph structure for all subjects, based on spatial or functional information, each row/column of the connectivity matrix serves as a signal for the corresponding node (node label).



(b) Siamese graph convolutional neural network for metric learning. A pair of graphs with the same structure but different signals is fed to this network, which outputs a similarity estimate between the two graphs. A same class (matching) / different class (non-matching) binary label is used for each pair during training.

Figure 1: Overview of the pipeline used for similarity metric learning on functional connectivity networks.

2.1. Datasets

2.1.1. ABIDE database

This dataset is provided by the Autism Brain Imaging Data Exchange (ABIDE) initiative (Di Martino et al., 2014) and has been preprocessed with the Config-

urable Pipeline for the Analysis of Connectomes (C-PAC)[1] (Craddock et al., 2013). This pipeline involves skull striping, slice timing correction, motion correction, global mean intensity normalisation, nuisance signal regression, band-pass filtering (0.01-0.1Hz) and registration of fMRI images to standard anatomical space (MNI152). The ABIDE database includes $N = 871$ subjects, 403 individuals suffering from ASD and 468 healthy controls, that met the imaging quality and phenotypic information criteria and were acquired with different imaging protocols at 20 acquisition sites. At this point it should be noted that the data is very different from one international site to the next, resulting from the different acquisition protocols at each site. We, subsequently, extract the mean time series for a set of brain regions based on the Harvard Oxford (HO) atlas, which comprises $R = 110$ cortical and subcortical ROIs (Desikan et al., 2006), and normalise them to zero mean and unit variance. The Fisher's transformed correlation matrices of the normalised timeseries are, then, used to specify the signal on the graph nodes $\mathbf{c} \in \mathbb{R}^{R \times R}$, i.e. each ROI is represented by its functional connectivity profile with the rest of the regions corresponding to one row of the correlation matrix. The process used to obtain the labelled graph representation is illustrated in Figure 1a.

2.1.2. UK Biobank

This is a large prospective study, planning to consistently acquire multimodal imaging data for 100,000 predominantly healthy subjects, in order to assist early disease prediction in the ageing population. As part of the acquired modalities, rs-fMRI data is available for the initial 5,000 participants' data release. These are accompanied by numerous non-imaging data, like age, sex, alcohol consumption, cognitive test scores and several others. We randomly select a subset of 2,500 subjects from this first release, including 1,181 male and 1,319 female subjects. The already preprocessed rs-fMRI images have been corrected for motion (Bannister et al., 2007) and distortion (Andersson et al., 2003), as well as

[1] `http://preprocessed-connectomes-project.org/abide/`

high-pass filtered to remove temporal drift. Miller et al. (2016) further apply an independent component analysis (ICA) based algorithm to automatically identify and remove artefacts. The 'cleaned' data is then fed to a group-level dimensionality reduction (Smith et al., 2014) and ICA to parcellate the brain into 100 spatially independent components that are not contiguous. Functional connectivity networks are, subsequently, estimated with L2-regularised partial correlation (Smith et al., 2011) and, after removing artefactual group-ICA components, comprise $R = 55$ nodes for each subject. Similarly to the ABIDE database, each row of the connectivity matrix, $\mathbf{c} \in \mathbb{R}^{R \times R}$, is used as the corresponding node's signal.

2.2. From fMRI Data to Graphs

Spectral graph convolutional networks filter signals defined on a common graph structure for all samples, meaning that they are not transferable from one domain to another, since these operations are parametrised on the graph's Laplacian. This is a constraint of the spectral approaches, which, however, provide a neat and principled way of performing convolutions in the irregular graph domain, where the notion of translation is not straightforward (Shuman et al., 2013). This property of the spectral filters is discussed in more detail in section 2.3. Although brain networks are very often treated as complete graphs, modelling brain connectivity as an irregular graph is more representative of their inherent complex architecture (Sporns et al., 2004). As a result, we model a common graph structure with two different approaches and explore their impact on the estimated similarities.

2.2.1. Spatial graph

The first approach we explore for defining the graph structure is based on anatomical information. We use the k-NN graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ of the regions spatial coordinates, where each ROI is represented by a node $v_i \in \mathcal{V}$ (located at the centre of the ROI) and the edges $\mathcal{E} = \{e_{ij}\}$ of the graph represent the spatial distances between connected nodes using $e_{ij} = d(v_i, v_j) = \sqrt{||v_i - v_j||^2}$ in terms

of 3D coordinates. Only the k spatially nearest nodes are then connected to the node v_i. For each subject, node v_i is associated with a signal $c_{si} : v_i \rightarrow \mathbb{R}^R,\ s = 1, ..., N$ which contains the node's connectivity profile in terms of Fisher's transformed Pearson's correlation between the representative rs-fMRI time series of each ROI.

2.2.2. Functional graph

As an alternative, we estimate the mean functional connectivity matrix among the training samples $\bar{A} = \frac{1}{N_{train}} \sum_{i=1}^{N_{train}} A_i$ and obtain the k-nn graph using the correlation distance between all region pairs. This kind of structure is more meaningful from a neuroscientific point of view, because it reflects the average functional connection strength between pairs of brain regions within a population. It is also data-driven, unlike the spatial structure which is purely based on anatomical information. Due to this fact it is more prone to introducing a bias towards the training set, if the latter does not contain enough samples to capture population variability. It should also be noted that we need to obtain an average connectivity matrix, because the filters learned with the spectral convolutions are tied to a specific domain. Hence, the individual connectivity matrices cannot be used, since the eigenbases of their Laplacian matrices will differ. The node signals are defined similarly to the above spatial graph, so only the structure of the graph itself is modified.

2.3. Spectral Graph Filtering and Convolutions

Traditional convolution operators rely on the regular grid-like structure of e.g. 2D and 3D images and it is, therefore, not trivial to generalise the convolution operation to the graph setting. Shuman et al. (2013) showed that this generalisation can be made feasible by the definition of filters in the graph spectral domain. If $A \in \mathbb{R}^{R \times R}$ is the adjacency matrix associated with a graph $\mathcal{G}$ and D the diagonal degree matrix, for which the diagonal elements are given by $d_i = \sum_{i \neq j} A_{ij}$, then the graph Laplacian is defined as $\mathcal{L} = D - A$. Its normalised equivalent is given by $L = I_R - D^{-1/2} A D^{-1/2}$, where I_R is the

identity matrix, and constitutes an essential operator in spectral graph analysis (Shuman et al., 2013). The normalised Laplacian L can be decomposed as $L = U\Lambda U^T$, where U is the matrix of eigenvectors and Λ the diagonal matrix of eigenvalues $\{\lambda_l\}_{l=0,1,\ldots,R-1}$. The eigenvalues represent the frequencies of their associated eigenvectors, *i.e.* eigenvectors associated with larger eigenvalues oscillate more rapidly between connected nodes, and for a connected graph $\mathcal{G}$ they satisfy $0 = \lambda_0 < \lambda_1 \leq \cdots \leq \lambda_{max}$. The graph Fourier transform of a signal $\mathbf{c}$ can, then, be expressed as $\hat{\mathbf{c}} = U^T\mathbf{c}$. This allows to define a convolution on a graph as a multiplication in the spectral domain of the signal $\mathbf{c}$ with a filter $g_\theta = diag(\theta)$ as:

$$g_\theta * \mathbf{c} = U g_\theta U^T \mathbf{c}, \tag{1}$$

where $\theta \in \mathbb{R}^R$ is a vector of Fourier coefficients and g_θ can be regarded as a function of the eigenvalues of L, i.e. $g_\theta(\Lambda)$ (Shuman et al., 2013).

The first formulation of spectral CNNs was directly parametrised on the eigenvectors of the Laplacian (Bruna et al., 2013), which required expensive computations of the eigendecomposition and did not guarantee that the filters represented in the spectral domain would be localised in the graph spatial domain. A polynomial parametrisation of the filters directly on the Laplacian can be used to address these limitations (Hammond et al., 2011), since K-order polynomials are exactly K-localised and define the number of hops around the central node taken into account for the convolution. Defferrard et al. (2016) proposed to approximate the filters by a truncated expansion in terms of Chebyshev polynomials to further reduce computational complexity. The Chebyshev polynomials are recursively defined as $T_k(c) = 2cT_{k-1}(c) - T_{k-2}(c)$, with $T_0(c) = 1$ and $T_1(c) = c$, which essentially reduces the computational complexity of the filtering operation. Filtering of a signal $\mathbf{c}$ with a K-localised filter can, then, be performed using:

$$y = g_\theta(L) * \mathbf{c} = \sum_{k=0}^{K} \theta_k T_k(\tilde{L})\mathbf{c}, \tag{2}$$

with $\tilde{L} = \frac{2}{\lambda_{max}} L - I_R$, where λ_{max} denotes the largest eigenvalue of L. The output of the j^{th} layer for a sample s in a Graph Convolutional Network (GCN) is then given by:

$$y_{s,j} = \sum_{i=1}^{F_{in}} g_{\theta_{i,j}}(L) c_{s,i} \in \mathbb{R}^R, \tag{3}$$

yielding $F_{in} \times F_{out}$ vectors of trainable Chebyshev coefficients $\theta_{i,j} \in \mathbb{R}^K$, where $c_{s,i}$ denotes the input feature maps, F_{in} the number of input filters and F_{out} the number of output filters. Therefore, the total number of trainable parameters per layer is $F_{in} \times F_{out} \times K$.

2.4. Network Architecture

Our siamese network, presented in Fig. 1b, consists of two identical pathways of C graph convolutional layers sharing the same weights, each taking a connectivity graph as input. An inner product layer combines the outputs from the two branches of the network and is followed by a single fully connected (FC) output layer with one output, that corresponds to the similarity estimate. The FC layer accounts for integrating global information about graph similarity from the preceding localised filters. Each convolutional layer is succeeded by a non-linear activation, i.e. Rectified Linear Unit (ReLU), but we avoid a non-linearity in the output layer as this was observed to cause a vanishing gradient problem. Therefore, the learned metric is unbounded.

2.5. Loss Functions

We investigate the performance of three different loss functions for the problem under consideration: the commonly used hinge loss (Zagoruyko and Komodakis, 2015), the global loss function used in our previous work (Ktena et al., 2017b), and a modification of this global loss that we introduce in this paper. The global loss function Kumar et al. (2016), is expected to provide a better performance than the hinge loss due to its increased robustness to outliers and better regularisation. Finally, we provide a modification to this loss that relaxes constraints on the variance of the learned distances, so as to increase the

generalisability on heterogeneous data. It should be noted that an additional l_2 regularisation term on the learned weights is introduced to the loss function in every case. More details about these objective functions that our model is aiming to minimise are provided below.

2.5.1. Hinge loss

The hinge loss is an objective function commonly used for "maximum-margin" classification. It has been employed for similarity metric learning of local image patches (Zagoruyko and Komodakis, 2015) and is described by the following equation:

$$J^{hinge} = \frac{1}{N}\sum_{i=1}^{N} \max(0, 1 - y_i o_i), \tag{4}$$

where y_i is the ground truth label of the pair (i.e. 1 for matching graphs vs. -1 for non-matching graphs) and o_i is the siamese network output. An output o_i higher than 1 for matching graphs or lower than -1 for non-matching graphs is not penalized, whereas outputs within the range (-1, 1) or on the wrong side of the hyperplane $y = 0$ are penalized in a linear fashion compared to their distance from the correct value. It is worth noticing that this loss function does not impose any constraints on the variance of the network outputs for each class.

2.5.2. Global loss

Kumar et al. (2016) proposed a pairwise similarity global loss function that yields superior results in the problem of metric learning for local image descriptors compared to a traditional loss. This global loss maximises the mean similarity μ^+ between embeddings belonging to the same class, minimises the mean similarity between embeddings belonging to different classes μ^- and, at the same time, minimises the variance of pairwise similarities for both matching, σ^{2+}, and non-matching, σ^{2-}, pairs of graphs. The formula of the global loss function suggested by Kumar et al. (2016) is given by:

$$J^{global} = (\ \sigma^{2+} + \sigma^{2-}) \ + \ \lambda \ \max\left(0, m - (\mu^+ - \mu^-)\right), \tag{5}$$

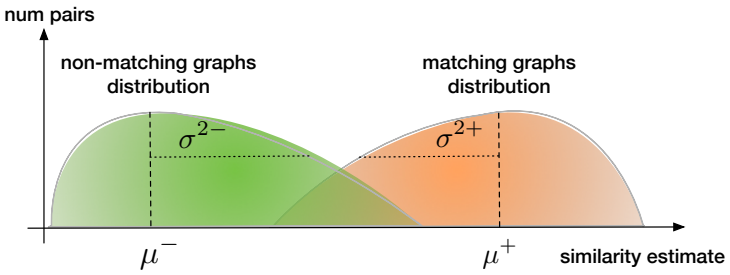


Figure 2: The goal of our loss function is to maximise the mean of the matching graphs similarity μ^+, minimise the mean of the non-matching graphs similarity μ^-, while restricting the variance of matching and non-matching classes, σ^{2+} and σ^{2-}, respectively, below a certain threshold.

where λ balances the importance of the mean and variance terms, and m is the margin between the means of matching and non-matching similarity distributions.

2.5.3. Constrained variance loss

Rather than minimising the variance, an objective that can cause the similarity estimates of the training samples to collapse around the class means, we propose to constrain the variance for each class below a certain threshold. The formulation of our proposed global loss function is, then, the following:

$$J^{convar} = \max(0, \sigma^{2+} - a) + \max(0, \sigma^{2-} - a) + \max\left(0, m - (\mu^+ - \mu^-)\right), \quad (6)$$

where a is the variance threshold. This formulation only penalises the variance when it exceeds the threshold a, allowing similarity estimates to vary around the means and accommodate the diversity inherent in heterogeneous databases such as the multi-site fMRI data. When $a = 0$, the constrained variance loss reduces to the global loss (Figure 2). The loss function is differentiable with

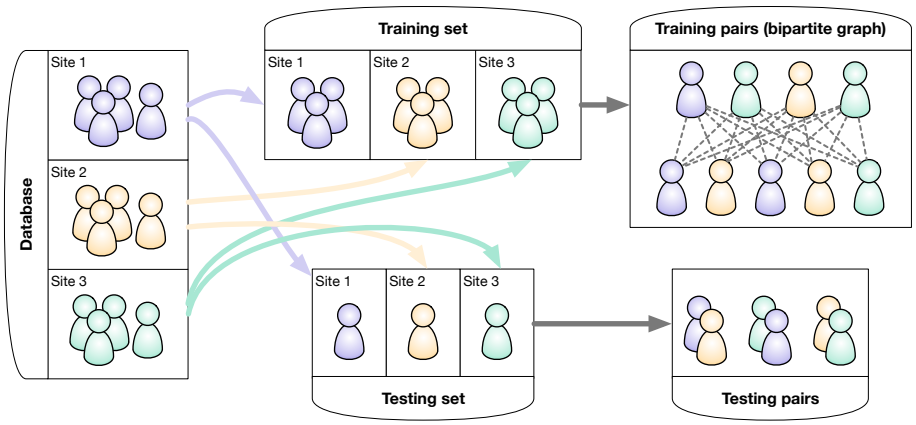


Figure 3: Illustration of the experimental setup used to train and test the siamese graph convolutional network on the *ABIDE database*. For each cross-validation fold, a certain percentage of subjects from each site is preserved for training and the rest for testing. From the training subjects a bipartite graph is constructed (after randomly splitting them into two sets) to obtain the training pairs and reduce the number of combinations between them. Since the number of testing subjects is much lower than the number of training subjects, all possible combinations between them are used as testing pairs.

gradients derived by:

$$\begin{aligned}
\frac{\partial J^{convar}}{\partial o(\mathbf{x}_i, \mathbf{x}_i^+, g_\Theta)} &= \frac{2}{N}[\mathbb{1}_{\{\sigma^{2+}>a\}}(o(\mathbf{x}_i, \mathbf{x}_i^+, g_\Theta) - \mu^+) - \frac{1}{2}\mathbb{1}_{\{(\mu^+-\mu^-)<m\}}(\mu^+ - \mu^-)] \\
\frac{\partial J^{convar}}{\partial o(\mathbf{x}_i, \mathbf{x}_i^-, g_\Theta)} &= \frac{2}{N}[\mathbb{1}_{\{\sigma^{2-}>a\}}(o(\mathbf{x}_i, \mathbf{x}_i^-, g_\Theta) - \mu^-) + \frac{1}{2}\mathbb{1}_{\{(\mu^+-\mu^-)<m\}}(\mu^+ - \mu^-)]
\end{aligned} \tag{7}$$

where $o(\mathbf{x}_i, \mathbf{x}_i^+, g_\theta)$ represents network similarity output for matching graphs, $o(\mathbf{x}_i, \mathbf{x}_i^-, g_\theta)$ the equivalent for non-matching graphs.

3. Results

3.1. Experimental setup

We evaluate the performance of the proposed model in a 5-fold cross-validation setting. In this setting the subjects are randomly partitioned into 5 equal size subsamples. For each run, a single subsample is retained as test data, while the

remaining 4 subsamples are used as training data. For the ABIDE database, we also ensure that subjects from all 20 sites are included in both training and test sets, to account for differences in the acquisition protocol. An illustration of this experimental setup is provided in Figure 3. Similarly to the experimental setup used in Zagoruyko and Komodakis (2015) for image patches, we train the network on matching and non-matching pairs.

Furthermore, for the ABIDE database, matching pairs correspond to graphs representing individuals of the same class, i.e. patients with autism spectrum disorder (ASD) or healthy controls (HC), while non-matching pairs correspond to graphs representing one HC and one subject with ASD. For UK Biobank, matching pairs are defined as graphs representing individuals of the same sex, i.e. both males or both females, while non-matching pairs include one male and one female subject.

We train a siamese network with $C = 2$ convolutional layers consisting of $f = 64$ features each. The different network hyper-parameters are optimised using cross-validation. We use stochastic gradient descent with Adaptive Moment Estimation (ADAM) (Kingma and Ba, 2014) as the optimization algorithm, learning rate 0.001 and polynomial filters of order $K = 3$, meaning that filters at each convolution are taking into account neighbours that are at most K steps away from a node. The number of neighbours for the graph structure is set to $k = 10$ for both datasets. The dropout ratio at the FC layer and regularisation parameter are set to 0.2 and 0.0005, respectively, for the ABIDE database, while for UK Biobank a dropout of 0.5 and a regularisation parameter of 0.05 are used. For the global loss function, the margin m is set to 1.0, while the weight λ is 1.0 as in Kumar et al. (2016). For the constrained variance loss we set $a = m/2$. In each fold the training set contains 720 subjects for the ABIDE database and 2000 subjects for UK Biobank, which we randomly split into two sets to construct a bipartite graph representing the training pairs (see Figure 3) and reduce training time (the number of all possible combinations is much higher). That way we make sure that all subjects are fed to the network the same number of times to avoid biases. The test set consists of all combinations between the

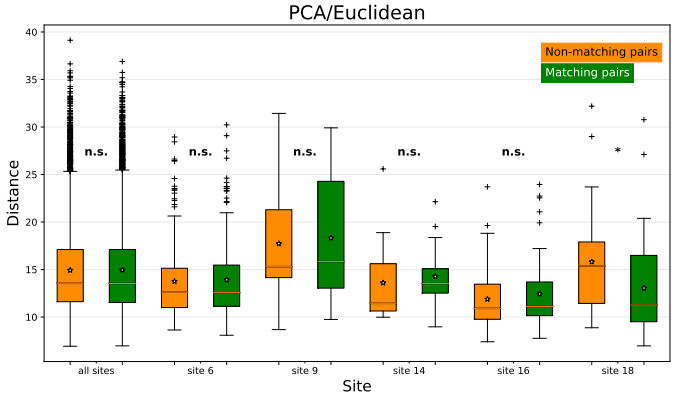
remaining 151 and 500 subjects for ABIDE and UK Biobank, respectively. In the ABIDE task, a binary feature is also introduced at the FC layer indicating whether the pair subjects were scanned at the same site or not.

3.2. Metric learning evaluation

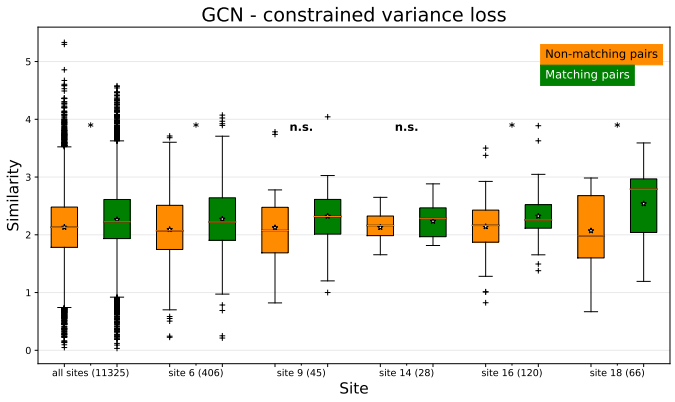
3.2.1. Estimated similarities

Figure 4a shows the pairwise distances between ABIDE functional connectivity matrices for the full test set and the 5 biggest sites after applying dimensionality reduction (PCA) and preserving $R = 110$ components, equal to the number of atlas regions. This figure illustrates that networks are hardly comparable using standard distance functions, even within the same acquisition site. At this point it should be noted that "all sites" refers to all pairs from the test set, even if the subjects were scanned at different sites. Figure 4b, in turn, shows the similarity estimates for the same random split, using the proposed loss function. It can be observed that the proposed metric learning architecture is significantly ($p < 0.05$) improving the separation between matching and non-matching pairs for the total test set, as well as for most individual sites. Additionally, the proposed loss seems to lead to better separation for sites 9 and 14, but the number of pairs is low to lead to statistically significant differences. It should be noted that Euclidean distances for these two sites were initially higher for the matching pairs compared to the non-matching pairs, which is the inverse of the desired behaviour.

Figure 5a illustrates the distances estimated with principal component analysis (PCA) with 55 components, equal to the number of ICA non-artefactual components, for UK Biobank in comparison to the similarities obtained with s-GCN and the constrained variance loss (Figure 5b). These distances/similarities correspond to all pairs between the subjects in the test set. It can be observed that the s-GCN model leads to better separation between matching and non-matching graph pair similarities, similarly to the ABIDE database.

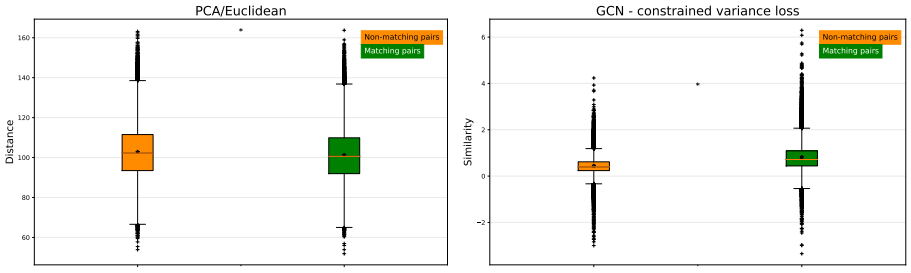


(a)



(b)

Figure 4: **(a)** Boxplots showing the computed euclidean *distance*, after applying dimensionality reduction with PCA on the training data of the *ABIDE database*. This indicates how challenging the problem of metric learning is for brain graphs, given that for certain acquisition sites the estimated distances for the matching graphs have a higher mean than the distances of the non-matching graphs (sites 14 and 16). **(b)** Box-plots showing *similarity* estimates for the same matching and non-matching graphs. Results with the proposed loss function (Eq. 6) are presented for all sites and the 5 largest sites (number of pairs for each case indicated in parentheses). Differences between the similarity distributions of the two classes (matching vs. non-matching) are indicated as significant (*) or non significant (n.s.) using a permutation test with 10000 permutations. Results presented in (a) and (b) correspond to the same set of subjects.

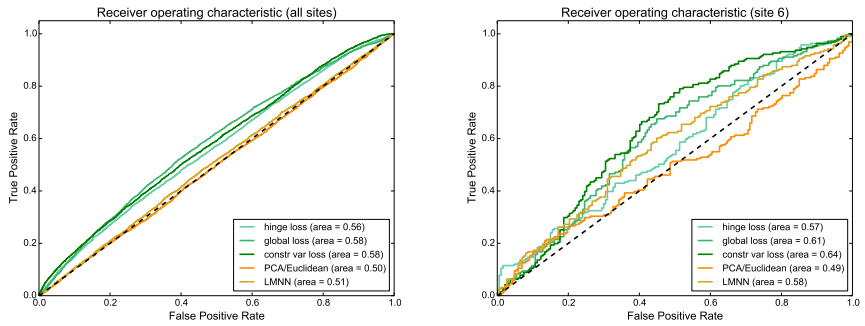


(a) (b)

Figure 5: **(a)** Boxplots showing the computed euclidean *distance*, after applying dimensionality reduction with PCA on the training data of *UK Biobank*. **(b)** Box-plots showing *similarity* estimates for the same matching and non-matching graphs. Results with the proposed loss function (Eq. 6) are given on the right. Differences between the similarity distributions of the two classes (matching vs. non-matching) are indicated as significant (*) or non significant (n.s.) using a permutation test with 10000 permutations. Results presented in (a) and (b) correspond to the same set of subjects.

3.2.2. Pair classification results

Figure 6 illustrates the results on the test set of the ABIDE database through receiver operating characteristic (ROC) curves for the biggest site, as well as across all sites, for the task of matching (same class) vs. non-matching (different class) graphs classification. The estimated area under curve (AUC) for the different loss functions is also indicated in parentheses. The ROC curve illustrates the diagnostic ability of a binary classifier while varying its discrimination threshold. In the case of s-GCN this threshold is compared directly with the similarity estimates, while for Euclidean distances it is compared to the reciprocal of the estimated distances between a pair of subjects. We further compare to the distances learned with a Large Margin Nearest Neighbour (LMNN) classifier, which aims to learn a global linear transformation of the input space that precedes k-nn classification using Euclidean distances (Weinberger and Saul, 2009). In other words, it learns a Mahalanobis distance metric for k-nn classification from labeled examples, but does not take into account the graph structure as it operates on the embedded vectors of the connectivity



(a) (b)

Figure 6: Receiver operating characteristic curves (ROCs) on the test set (a) for all sites and the biggest site (b) from the *ABIDE database* for the task of matching vs. non-matching pair classification. The area under curve (AUC) is indicated for the hinge, global Kumar et al. (2016) and the proposed loss using s-GCN with the spatial graph structure, as well as Euclidean distances after dimensionality reduction and Large-Margin Nearest Neighbour (LMNN) metric learning

.

matrices. Pairs with estimated similarities above the threshold are, then, classified as matching, while pairs below the threshold are classified as non-matching. The ROC curve is obtained by plotting the true positive rate against the false positive rate for different thresholds. The true positive rate, also known as sensitivity, measures the proportion of matching pairs that are correctly identified as such. The false positive rate, in turn, is calculated as the ratio between the number of non-matching pairs wrongly classified as matching and the total number of non-matching pairs. The area under the ROC curve (AUC) is equal to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one. Therefore, AUC values closer to 1 are preferable, while an AUC of 0.5 corresponds to chance level.

It can be observed in figs. 6a and 6b that the AUC obtained with Euclidean distances is at chance level both for the largest site and all sites in the ABIDE database. Improved results are observed for site 6 when using LMNN, which are more comparable to the sGCN with hinge loss, but this performance improve-

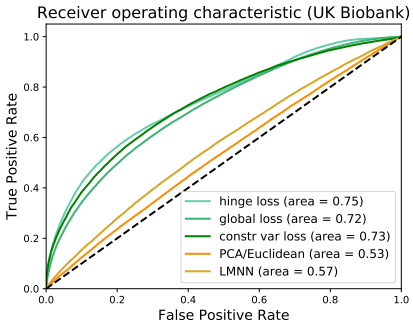
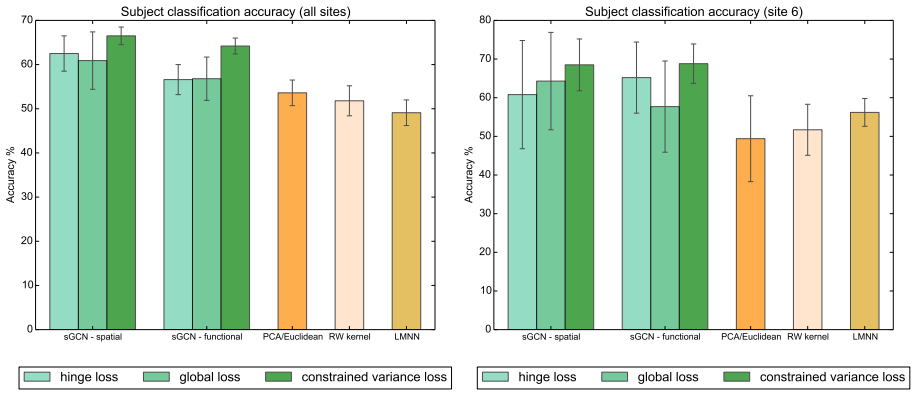


Figure 7: Receiver operating characteristic curves (ROCs) on the test set for all pairs from the *UK Biobank* dataset for the task of matching vs. non-matching pairs classification. The area under curve (AUC) is indicated for the hinge, global Kumar et al. (2016) and the proposed loss using s-GCN with the functional graph structure, as well as Euclidean distances after dimensionality reduction and Large-Margin Nearest Neighbour (LMNN) metric learning

.

ment is not verified for all sites, probably because of the heterogeneity of the data from different sites. For the proposed s-GCN model with a spatial graph structure improved performance is achieved, which is more striking between pairs from the same site. We use as a baseline the AUC obtained with Euclidean distances between pairs after preserving the 110 principal components from the training set. For LMNN all features are used to learn the linear transformation, while the learning rate is set to 10^{-5} and 10^{-7} for ABIDE and UK Biobank, respectively. These parameters are chosen using grid search cross-validation. Among the different loss functions, the proposed loss seems to outperform the global and hinge loss. Results by means of AUC are similar for the functional graph structure and are, thus, omitted in this section.

Figure 7 presents the ROC curves for the different loss functions, PCA/Euclidean with 55 components and LMNN for the task of identifying subject pairs of the same gender vs. pairs of different gender. The improvement in performance with s-GCN is, in this case, more prevalent than the ABIDE database and the ROC curves are smoother, since the size of both training and test sets is larger for UK Biobank. Although the proposed modification to the global loss still out-



(a) (b)

Figure 8: Summary results with 5-fold cross-validation for k-nn classification accuracy with k=9 for all sites and the largest sites of the *ABIDE database* separately. The number of subjects in the test set from the remaining sites is too small to perform this kind of evaluation on a per site basis.

performs the traditional global loss, the hinge loss is leading to slightly improved results. This can attributed to the fact that the proposed loss is advantageous for heterogeneous data, while UK Biobank fMRI data are all acquired with the same scanner. The siamese network still outperforms LMNN, which does not take into account the graph structure.

3.2.3. Subject classification results

In this experiment, the estimated pairwise similarities are used to assign a class membership (ASD or HC for ABIDE, male or female for UK Biobank) for each subject in the test set based on a k-nearest neighbour (k-nn) classifier. The accuracy of k-nn classification is highly dependent on the metric used to compute the similarities between different samples. Therefore, a meaningful metric that reflects the underlying similarities between data samples should be able to facilitate classification with a simple k-nn classifier. In this process, each subject is classified by a majority vote of its neighbours, with the subject being assigned to the most common class among its k nearest neighbours in

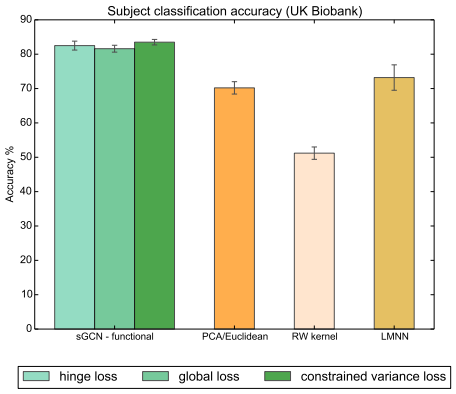


Figure 9: Summary results with 5-fold cross-validation for k-nn classification accuracy with k=9 for *UK Biobank*.

terms of the learned metric, and for this reason k is generally an odd number. We compare the different loss functions and graph structures, spatial and functional, for the ABIDE database by means of k-nn classification accuracy, with $k = 9$ chosen empirically and summarise the results in Figure 8. Results for UK Biobank are presented in Figure 9. We use as a baseline a k-nn classifier based on the euclidean distances between the subjects after applying PCA and further compare to a random walk kernel, which takes into account the graph structure to estimate graph similarities (Vishwanathan et al., 2010) and LMNN as previously. The classification results by means of accuracy/percentage of correctly predicted subjects are obtained with 5-fold cross-validation for both databases. These are reported for subjects from all sites together, as well as the biggest site separately, for the ABIDE database since the rest of the sites did not have enough representatives in the test set to perform such a task.

It can be observed that across all sites, the proposed loss leads to improved performance with the same network architecture compared to the hinge and global losses. This is the case for both graph structures, spatial and functional. For all sites, the metric learning approach leads to improved classification accuracy compared to the standard Euclidean distance after dimensionality reduction and the random walk kernel. This is a fair comparison, since in both cases

only a pairwise (dis)similarity matrix between subjects is provided to the classifier. Compared to the k-nn classifier based on Euclidean distances, the learned metric demonstrates improved performance by 12.9% for all sites, 19.4% for the biggest site. Furthermore, the performance achieved with the learned metric is better by 14.7% and 17.1% compared to the random walk kernel for all sites and the biggest site, respectively. LMNN is, surprisingly, performing marginally worse than Euclidean distances for all sites, although it leads to improved performance for the biggest site. This can be due to the fact that the biggest site dominates the learned LMNN metric, leading to a negative impact on the rest of the sites. As far as the graph structure is concerned, the data-driven functional structure does not seem to improve the classification performance for all sites, but only for the biggest site. This can be attributed to the fact that this site is dominating the training set (with a total of 172 subjects in the database), whereas the remaining sites have much less subjects (maximum 11-86 subjects each) contributing to the estimation of the graph structure.

As illustrated in Figure 9, only the functional structure is available for UK Biobank, since the parcels used to construct the functional connectivity networks are obtain with spatial-ICA and are, therefore, not spatially contiguous. Differences between the different loss functions are marginal, i.e. the proposed loss leads to 1.0% improved performance compared to the hinge loss and 1.9% compared to the global loss. However, the comparison to alternative methods is more striking, mostly due to the larger size of the available dataset (13.3% better than PCA/Euclidean and 10.3% better than LMNN). Notably, the random walk kernel performs very poorly for this dataset.

3.2.4. Manifold learning

Another useful application of a meaningful similarity metric is manifold learning. Manifold learning techniques aim to discover a low-dimensional embedding of high-dimensional data by employing non-linear dimensionality reduction, and aim to facilitate visualisation and interpretation of the data. Locally linear embedding (LLE) is one of the most common manifold learning methods,

which seeks a lower-dimensional projection of the data which preserves distances within local neighbourhoods (Roweis and Saul, 2000). We, therefore, use the inferred pairwise dissimilarities, i.e. the distance matrix of the subjects in the test set, to learn a lower dimensional non-linear embedding of the data and use it as a qualitative means of evaluation. LLE uses an eigenvector-based optimization technique to find the low-dimensional embedding of points, such that each point is still described with the same linear combination of its neighbors.

Figure 10 visualises the embeddings of the two databases (ABIDE and UK Biobank) in two dimensions using LLE with $k = 9$ neighbours and aims to compare the learned metric to the baseline in a qualitative manner. It can be observed that for both databases the learned metric leads to better separation of the two classes (ASD vs HC for ABIDE and male vs female for UK Biobank) compared to the Euclidean distances. This is further highlighted by the centre of mass for each class in this low-dimensional embedding of the data. These are located further apart in figs. 10e and 10f compared to figs. 10a and 10b, respectively, indicating that the learned metric leads to more meaningful local distances between subjects, a result more pronounced for UK Biobank data. LMNN figs. 10c and 10d is also performing better than Euclidean distances in this qualitative evaluation, with visual results being more favourable with UK Biobank as a more homogeneous database, but the improvement gained with sGCN is still noticeable, especially for ABIDE.

4. Discussion

In this work, we use a siamese graph convolutional neural network, which employs the polynomial filters formulated in Defferrard et al. (2016), to learn a similarity metric between brain connectivity graphs for classification and manifold learning. We extend the preliminary work on metric learning for irregular graphs (Ktena et al., 2017b) and perform a more thorough evaluation of the proposed method. We leverage the recent concept of graph convolutions through a siamese architecture and propose a modification to the global loss function

used previously to accommodate heterogeneous data. The method is extensively evaluated on two large databases, i.e. the functional brain connectivity graphs from the heterogeneous ABIDE database used in the previous work, as well as a larger set of subjects from UK Biobank. We, further, explore the impact of three different loss functions, i.e. the commonly used hinge loss, the global loss and the constrained variance loss introduced in this paper, on the learned metric in different settings, including matching vs. non-matching graph classification, subject classification and manifold learning experiments. We obtain promising quantitative and qualitative results for both datasets and significant improvements over the baselines, which are more obvious for UK Biobank that comprises a larger and more homogeneous dataset. This is a clear indicator that the diversity of acquisition protocols and the limited number of subjects limits the overall performance on the ABIDE database. While applied to brain networks, our proposed method is flexible and general enough to be applied to any problem involving comparisons between graphs, e.g. shape analysis.

One of the main shortcomings of the proposed model is the fixed graph requirement. This is due to the fact that the learned filters are tied to the specific domain/graph Laplacian and cannot be altered between samples or during inference. We, therefore, explore the effect of two different graph structures on the ABIDE database, one based on spatial proximity of the atlas contiguous parcels and one based on purely functional information. Although the spatial structure is less meaningful from a neuroscientific point of view, no significant differences are observed between the two graph structures for the ABIDE database. This can be attributed to the fact that the functional structure is biased towards the largest acquisition site, limiting its positive influence on the learned metric, and the redundancy of information within the graph structure that is already existing in the node features. However, the potential of the functional graph structure is validated with the UK Biobank dataset. Another interesting option would be to use the structural connectivity as the graph structure, when this is available, while preserving functional connectivity information as node features. The proposed loss function, leads to an improvement over the more traditional

hinge and global losses for the classification tasks, which is more prominent for the heterogeneous ABIDE database.

Although this approach is different from the traditional one used in connectomics research, where brain networks are treated as complete graphs, i.e. every node is connected to every other node with a different weight, the edge weight information is still modelled in the form of a node's "connectivity profile" as its corresponding feature vector. The proposed modelling approach allows for the extension and integration of additional, potentially multimodal, information characterising each graph node, like e.g. cortical thickness. This is particularly useful in the study of neuropsychiatric and neurodegenerative diseases and such features have already been investigated in conjunction with autism (Chung et al., 2005) and Alzheimer's disease (Querbes et al., 2009) among others. Alternative frameworks that operate purely on connectivity matrices would not be able to accommodate and model multimodal information regarding structural and functional connectivity, as well as anatomical features, as part of a node's feature vector. Furthermore, one could argue that the raw time series could be used instead, as a more intuitive feature vector to represent the graph nodes, however there is no guarantee that the time series of different subjects are temporally aligned in a way that leads to cross-subject feature correspondence.

The choice of this polynomial parametrisation further provides a principled way of applying graph convolutions, while it addresses the challenging problems of defining a local neighbourhood and transferring the notion of translation from the Euclidean domain. Modelling brain connectivity as an irregular graph is more representative of its inherent architecture and can address the increased noise that arises when estimating brain connectivity networks, especially based on functional data. Last but not least, from a technical standpoint, the polynomial parametrization of the spectral filters adopted in this work allows to capture the graph signal at a varying neighbourhood size/proximity around each node, i.e. k-hops away for a k^{th} order polynomial. Alternative approaches, like the work of (Kawahara et al., 2017) require the introduction of a new layer to capture a larger "field of view" and learn a specific filter bank

for each edge of the graph, instead of frequency filters that are applicable to the whole graph. This leads to a larger set of parameters and may give rise to overfitting problems, especially when working with relatively small neuroimaging datasets. The framework used in this work has also been shown to allow pooling operations (Defferrard et al., 2016), which paves the way for the exploration of hierarchical parcellations (e.g. Blumensath et al. (2013)) in future studies on brain connectivity.

Generating and visualising discriminative network regions or connections with these graph convolutional network frameworks, in a similar manner that saliency maps have recently been proposed for natural images (Zintgraf et al., 2017) would be particularly beneficial for the neuroscience community. These would shed light on the underlying mechanisms of disconnection syndromes as well as network patterns giving rise to population differences in brain connectivity. The spectral methods aim to yield more interpretable patterns compared to spatial approaches and hierarchical parcellations would assist in exploring these patterns at multiple scales. Another important aspect of this work is the ability to handle heterogeneous data aggregated at multiple imaging sites, with different acquisition protocols, using a tailored architecture, which could potentially be improved with adversarial training.

One of the main advantages of this work is that the same model and network architecture are able to address very diverse classification tasks given as input datasets preprocessed with different pipelines and graphs representing different types of functional connectivity (e.g. full correlation and partial correlation). A particularly exciting prospect would be to use autoencoders and adversarial training to learn lower dimensional representations of the connectivity networks that are site independent and able to handle heterogeneity that arises from diverse acquisition protocols or other factors, such as age or gender. Additionally, exploring the use of generalisable GCNs defined in the graph spatial domain Monti et al. (2016) would allow to train similarity metrics between graphs of different structures. Last but not least, alternative applications of the learned metric to other biomedical image analysis tasks, which can be

modelled as graphs Paragios et al. (2016), including semi-supervised clustering or graph-based label propagation techniques Parisot et al. (2017) could benefit from the integration of similarity metric learning in their pipeline and are yet to be explored.

Acknowledgements

This research has been conducted using the UK Biobank Resource under Application Number 12579. Sofia Ira Ktena is supported by the EPSRC Centre for Doctoral Training in High Performance Embedded and Distributed Systems (HiPEDS, Grant Reference EP/L016796/1). Enzo Ferrante is beneficiary of an AXA Research Fund postdoctoral grant. The support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research is gratefully acknowledged.

References

Abraham, A., Milham, M.P., Di Martino, A., Craddock, R.C., Samaras, D., Thirion, B., Varoquaux, G., 2017. Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example. NeuroImage 147, 736–745.

Achard, S., Salvador, R., Whitcher, B., Suckling, J., Bullmore, E., 2006. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. Journal of Neuroscience 26, 63–72.

Andersson, J.L., Skare, S., Ashburner, J., 2003. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. NeuroImage 20, 870–888.

Arslan, S., Ktena, S.I., Makropoulos, A., Robinson, E.C., Rueckert, D., Parisot, S., 2017. Human brain mapping: A systematic comparison of parcellation methods for the human cerebral cortex. NeuroImage .

Bannister, P.R., Brady, J.M., Jenkinson, M., 2007. Integrating temporal information with a non-rigid method of motion correction for functional magnetic resonance images. Image and Vision Computing 25, 311–320.

Blumensath, T., Jbabdi, S., Glasser, M.F., Van Essen, D.C., Ugurbil, K., Behrens, T.E., Smith, S.M., 2013. Spatially constrained hierarchical parcellation of the brain with resting-state fMRI. Neuroimage 76, 313–324.

Boscaini, D., Masci, J., Rodolà, E., Bronstein, M., 2016. Learning shape correspondence with anisotropic convolutional neural networks, in: Advances in Neural Information Processing Systems, pp. 3189–3197.

Bruna, J., Zaremba, W., Szlam, A., LeCun, Y., 2013. Spectral networks and locally connected networks on graphs. arXiv preprint arXiv:1312.6203 .

Catani, M., ffytche, D.H., 2005. The rises and falls of disconnection syndromes. Brain 128, 2224–2239.

Chung, M.K., Robbins, S.M., Dalton, K.M., Davidson, R.J., Alexander, A.L., Evans, A.C., 2005. Cortical thickness analysis in autism with heat kernel smoothing. NeuroImage 25, 1256–1265.

Craddock, C., Sikka, S., Cheung, B., Khanuja, R., Ghosh, S.S., Yan, C., Li, Q., Lurie, D., Vogelstein, J., Burns, R., et al., 2013. Towards automated analysis of connectomes: The configurable pipeline for the analysis of connectomes (C-PAC). Front Neuroinform 42.

Defferrard, M., Bresson, X., Vandergheynst, P., 2016. Convolutional neural networks on graphs with fast localized spectral filtering, in: Advances in Neural Information Processing Systems, pp. 3837–3845.

Desikan, R., Ségonne, F., Fischl, B., Quinn, B., Dickerson, B., Blacker, D., Buckner, R., Dale, A., Maguire, R., Hyman, B., et al., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage 31, 968–980.

Di Martino, A., Yan, C., Li, Q., Denio, E., Castellanos, F., Alaerts, K., Anderson, J., Assaf, M., Bookheimer, S., Dapretto, M., et al., 2014. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular Psychiatry 19, 659–667.

Dosenbach, N.U., Nardos, B., Cohen, A.L., Fair, D.A., Power, J.D., Church, J.A., Nelson, S.M., Wig, G.S., Vogel, A.C., Lessov-Schlaggar, C.N., et al., 2010. Prediction of individual brain maturity using fMRI. Science 329, 1358–1361.

Fornito, A., Zalesky, A., Breakspear, M., 2015. The connectomics of brain disorders. Nature Reviews Neuroscience 16, 159–172.

Geerligs, L., Renken, R.J., Saliasi, E., Maurits, N.M., Lorist, M.M., 2014. A brain-wide study of age-related changes in functional connectivity. Cerebral Cortex 25, 1987–1999.

Hagmann, P., Sporns, O., Madan, N., Cammoun, L., Pienaar, R., Wedeen, V.J., Meuli, R., Thiran, J.P., Grant, P., 2010. White matter maturation reshapes structural connectivity in the late developing human brain. Proceedings of the National Academy of Sciences 107, 19067–19072.

Hammond, D., Vandergheynst, P., Gribonval, R., 2011. Wavelets on graphs via spectral graph theory. Applied and Computational Harmonic Analysis 30.

Henaff, M., Bruna, J., LeCun, Y., 2015. Deep convolutional networks on graph-structured data. arXiv preprint arXiv:1506.05163 .

Jie, B., Zhang, D., Wee, C.Y., Shen, D., 2014. Topological graph kernel on multiple thresholded functional connectivity networks for mild cognitive impairment classification. Human Brain Mapping 35, 2876–2897.

Kawahara, J., Brown, C.J., Miller, S.P., Booth, B.G., Chau, V., Grunau, R.E., Zwicker, J.G., Hamarneh, G., 2017. BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment. NeuroImage 146, 1038–1049.

Kingma, D., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 .

Kipf, T., Welling, M., 2016. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 .

Konrad, K., Eickhoff, S.B., 2010. Is the adhd brain wired differently? a review on structural and functional connectivity in attention deficit hyperactivity disorder. Human Brain Mapping 31, 904–916.

Ktena, S.I., Arslan, S., Parisot, S., Rueckert, D., 2017a. Exploring heritability of functional brain networks with inexact graph matching, in: 12th International Symposium on Biomedical Imaging (ISBI), IEEE. pp. 354–357.

Ktena, S.I., Parisot, S., Ferrante, E., Rajchl, M., Lee, M., Glocker, B., Rueckert, D., 2017b. Distance metric learning using graph convolutional networks: Application to functional brain networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 469–477.

Kumar, B., Carneiro, G., Reid, I., et al., 2016. Learning local image descriptors with deep siamese and triplet convolutional networks by minimising global loss functions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5385–5394.

Levie, R., Monti, F., Bresson, X., Bronstein, M.M., 2017. Cayleynets: Graph convolutional neural networks with complex rational spectral filters. arXiv preprint arXiv:1705.07664 .

Livi, L., Rizzi, A., 2013. The graph matching problem. Pattern Analysis and Applications 16, 253–283.

Masci, J., Boscaini, D., Bronstein, M., Vandergheynst, P., 2015. Geodesic convolutional neural networks on riemannian manifolds, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 37–45.

Masci, J., Rodolà, E., Boscaini, D., Bronstein, M.M., Li, H., 2016. Geometric deep learning, in: SIGGRAPH ASIA 2016 Courses, ACM. p. 1.

Miller, K.L., Alfaro-Almagro, F., Bangerter, N.K., Thomas, D.L., Yacoub, E., Xu, J., Bartsch, A.J., Jbabdi, S., Sotiropoulos, S.N., Andersson, J.L., et al., 2016. Multimodal population brain imaging in the uk biobank prospective epidemiological study. Nature neuroscience 19, 1523–1536.

Mokhtari, F., Hossein-Zadeh, G.A., 2013. Decoding brain states using backward edge elimination and graph kernels in fMRI connectivity networks. Journal of neuroscience methods 212, 259–268.

Monti, F., Boscaini, D., Masci, J., Rodolà, E., Svoboda, J., Bronstein, M., 2016. Geometric deep learning on graphs and manifolds using mixture model CNNs. arXiv preprint arXiv:1611.08402 .

Neuhaus, M., Bunke, H., 2005. Self-organizing maps for learning the edit costs in graph matching. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 35, 503–514.

Neuhaus, M., Bunke, H., 2007. Automatic learning of cost functions for graph edit distance. Information Sciences 177, 239–247.

Niepert, M., Ahmed, M., Kutzkov, K., 2016. Learning convolutional neural networks for graphs, in: Proceedings of the 33rd Annual International Conference on Machine Learning. ACM.

Paragios, N., Ferrante, E., Glocker, B., Komodakis, N., Parisot, S., Zacharaki, E.I., 2016. (hyper)-graphical models in biomedical image analysis. Medical Image Analysis 33, 102–106.

Parisot, S., Ktena, S.I., Ferrante, E., Lee, M., Moreno, R.G., Glocker, B., Rueckert, D., 2017. Spectral graph convolutions on population graphs for disease prediction, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 177–185.

Querbes, O., Aubry, F., Pariente, J., Lotterie, J.A., Démonet, J.F., Duret, V., Puel, M., Berry, I., Fort, J.C., Celsis, P., et al., 2009. Early diagnosis of alzheimer's disease using cortical thickness: impact of cognitive reserve. Brain 132, 2036–2047.

Raj, A., Mueller, S., Young, K., Laxer, K., Weiner, M., 2010. Network-level analysis of cortical thickness of the epileptic brain. NeuroImage 52, 1302–1313.

Richiardi, J., Eryilmaz, H., Schwartz, S., Vuilleumier, P., Van De Ville, D., 2011. Decoding brain states from fMRI connectivity graphs. Neuroimage 56, 616–626.

Roweis, S.T., Saul, L.K., 2000. Nonlinear dimensionality reduction by locally linear embedding. science 290, 2323–2326.

Rubinov, M., Sporns, O., 2010. Complex network measures of brain connectivity: uses and interpretations. NeuroImage 52, 1059–1069.

Rudie, J.D., Brown, J., Beck-Pancer, D., Hernandez, L., Dennis, E., Thompson, P., Bookheimer, S., Dapretto, M., 2013. Altered functional and structural brain network organization in autism. NeuroImage: clinical 2, 79–94.

Satterthwaite, T.D., Wolf, D.H., Roalf, D.R., Ruparel, K., Erus, G., Vandekar, S., Gennatas, E.D., Elliott, M.A., Smith, A., Hakonarson, H., et al., 2014. Linked sex differences in cognition and functional connectivity in youth. Cerebral cortex 25, 2383–2394.

Shervashidze, N., Schweitzer, P., Leeuwen, E.J.v., Mehlhorn, K., Borgwardt, K.M., 2011. Weisfeiler-lehman graph kernels. Journal of Machine Learning Research 12, 2539–2561.

Shervashidze, N., Vishwanathan, S., Petri, T., Mehlhorn, K., Borgwardt, K., 2009. Efficient graphlet kernels for large graph comparison, in: Artificial Intelligence and Statistics, pp. 488–495.

Shuman, D., Narang, S., Frossard, P., Ortega, A., Vandergheynst, P., 2013. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. IEEE Signal Processing Magazine 30, 83–98.

Simonovsky, M., Komodakis, N., 2017. Dynamic edge-conditioned filters in convolutional neural networks on graphs, in: Proceedings of the IEEE International Conference on Computer Vision.

Smith, S.M., Hyvärinen, A., Varoquaux, G., Miller, K.L., Beckmann, C.F., 2014. Group-pca for very large fMRI datasets. NeuroImage 101, 738–749.

Smith, S.M., Miller, K.L., Salimi-Khorshidi, G., Webster, M., Beckmann, C.F., Nichols, T.E., Ramsey, J.D., Woolrich, M.W., 2011. Network modelling methods for fMRI. Neuroimage 54, 875–891.

Smith, S.M., Nichols, T.E., Vidaurre, D., Winkler, A.M., Behrens, T.E., Glasser, M.F., Ugurbil, K., Barch, D.M., Van Essen, D.C., Miller, K.L., 2015. A positive-negative mode of population covariation links brain connectivity, demographics and behavior. Nature Neuroscience 18, 1565–1567.

Sporns, O., 2011. The human connectome: a complex network. Annals of the New York Academy of Sciences 1224, 109–125.

Sporns, O., 2013. Structure and function of complex brain networks. Dialogues Clin Neurosci 15, 247–262.

Sporns, O., Chialvo, D.R., Kaiser, M., Hilgetag, C.C., 2004. Organization, development and function of complex brain networks. Trends in cognitive sciences 8, 418–425.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al., 2015. Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS medicine 12, e1001779.

Takerkart, S., Auzias, G., Thirion, B., Ralaivola, L., 2014. Graph-based inter-subject pattern analysis of fMRI data. PloS one 9, e104586.

Vishwanathan, S.V.N., Schraudolph, N.N., Kondor, R., Borgwardt, K.M., 2010. Graph kernels. Journal of Machine Learning Research 11, 1201–1242.

Wauquier, P., Keller, M., 2015. Metric learning approach for graph-based label propagation. arXiv preprint arXiv:1511.05789 .

Weinberger, K.Q., Saul, L.K., 2009. Distance metric learning for large margin nearest neighbor classification. Journal of Machine Learning Research 10, 207–244.

Zagoruyko, S., Komodakis, N., 2015. Learning to compare image patches via convolutional neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4353–4361.

Zeng, L.L., Shen, H., Liu, L., Wang, L., Li, B., Fang, P., Zhou, Z., Li, Y., Hu, D., 2012. Identifying major depression using whole-brain functional connectivity: a multivariate pattern analysis. Brain 135, 1498–1507.

Zintgraf, L.M., Cohen, T.S., Adel, T., Welling, M., 2017. Visualizing deep neural network decisions: Prediction difference analysis, in: International Conference on Learning Representations (ICLR).

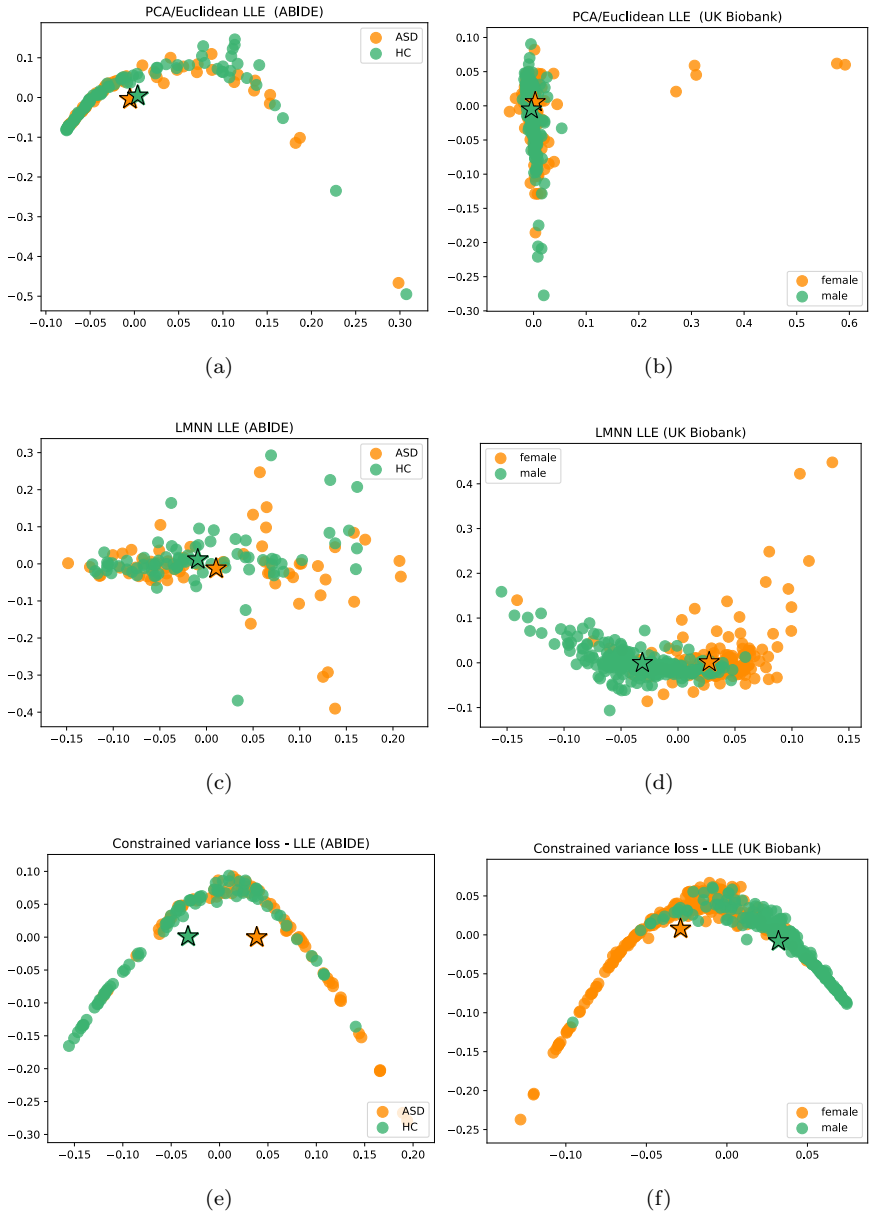


Figure 10: Locally linear embedding (LLE) for the *ABIDE database* (left) and *UK Biobank* (right) using Euclidean distances (a-b) after dimensionality reduction, metric learning for large margin nearest neighbour (LMNN) classification (c-d) and similarities learned with s-GCN and the proposed loss function (e-f). The star markers indicate the centre of mass for each class. For both databases the learned metric leads to better separation between the two classes.