Abstract

Comprehending the interplay between spatial and temporal characteristics of neural dynamics can contribute to our understanding of information processing in the human brain. Graph neural networks (GNNs) provide a new possibility to interpret graph structured signals like those observed in complex brain networks. In our study we compare different spatio-temporal GNN architectures and study their ability to replicate neural activity distributions obtained in functional MRI (fMRI) studies. We evaluate the performance of the GNN models on a variety of scenarios in MRI studies and also compare it to a VAR model, which is currently predominantly used for directed functional connectivity analysis. We show that by learning localized functional interactions on the anatomical substrate, GNN based approaches are able to robustly scale to large network studies, even when available data are scarce. By including anatomical connectivity as the physical substrate for information propagation, such GNNs also provide a multimodal perspective on directed connectivity analysis, offering a novel possibility to investigate the spatio-temporal dynamics in brain networks.

Keywords: brain connectivity, graph neural networks, structure - function relationship, directed functional connectivity

1 Introduction

Distinct concepts of brain connectivity can provide different but complementary aspects of information processing in the human brain [44]. On the one hand, imaging modalities like functional magnetic resonance imaging (fMRI) allow us to temporally resolve dynamic neural activity patterns in distinct spatial locations in the human brain. Different statistical approaches that describe the coherency of activity profiles in brain networks have been proposed based on the notion of
functional connectivity (FC). On the other hand, diffusion tensor imaging (DTI) can provide insights into the structural and relatively static aspects of the brain. By reconstructing white matter tracks from DTI data, the anatomical or structural connectivity (SC) between different brain areas can be estimated. Also directed and potentially causal relationships between regions become of interest in fMRI and are studied with respect to directed functional or effective connectivity. The latter is most often inferred from Granger causality or dynamic causal modeling [26].

Based on these concepts, spatial and temporal relationships between brain areas can be represented by graphical models, which have recently received increasing attention in the field of machine learning [81, 13]. So-called graph neural networks (GNNs) allow us to effectively process signals in the non-Euclidean geometry of graphs, providing also novel possibilities for applications in brain connectivity research [57, 79, 43, 7, 45, 41]. Given a decomposition of the brain into specified areas, their spatio-temporal neural activity patterns can be interpreted as graph structured signal distributions. Nodes in brain networks can be associated with variables like the temporal neuronal activity of neuron pools, while edges in such networks would reflect the strength of interactions between neural populations [15]. As proposed in our recent study [79], these complex signals exhibited in non-Euclidean geometries can be processed with a variant of GNN denoted as spatio-temporal graph neural network (STGNN), which simultaneously models spatial and temporal dependencies in such graph structured signals. We investigated such spatio-temporal brain activity distributions from a multimodal perspective by employing fMRI in combination with DTI data. In this framework the time-varying signal in the network nodes is represented as the average functional activity in the brain regions, while the static edge strength between nodes/regions is given by their anatomical connectivity strength.

Recently, several different GNN architectures have been proposed to model information propagation across time and space in graphical signals [81]. Convolution operations, often applied in deep learning, have recently been extended successfully to graphical models and allow us to capture inherent spatial dependencies on non-Euclidean network structures [22]. They were later combined with recurrent neural networks (RNNs) [58], which can detect sequential relations in signals. This combined spatio-temporal GNN framework was proposed in the notion of diffusion convolution recurrent neural network (DCRNN) [46]. However RNNs can have problems with long time series and, when combined with graph convolution operations, their gradients are more likely to explode or vanish [61, 46]. This was the motivation for introducing approaches that combine spatial graph convolutions with standard one-dimensional temporal convolutions [81]. But their receptive field size can only grow if many hidden layers are used (linear growth) or global pooling is applied. To alleviate such shortcomings so-called WaveNets (WNs) have been introduced that employ stacked dilated temporal convolutions, which provide a long-term temporal memory [75]. They have recently been combined with graph convolution operations in an architecture denoted as Graph WaveNet (GWN) [82]. As an alternative method for the temporal processing, also attention mechanisms have been recently included in STGNN architectures [87]. Attention mechanisms select, from all inputs, information that is critical to the task at hand and modify edge connection strengths accordingly. They have been applied already to natural language processing, speech recognition and image processing [84, 77, 48], but applications to analyze dynamics in neural signals are still missing. In this study we compare these different STGNN architectures with each other and evaluate their effectiveness in replicating functional ac-
tivity distributions observed in brain networks. In addition to the different temporal models we study different approaches to model the spatial information exchange between brain regions. At first we employ the structural connectivity as the substrate for information propagation between brain regions. Further we evaluate the effectiveness of employing connectome embeddings of the anatomical network to characterize the node relations. At first, in a recent study Rosenthal et al. [57] have shown that embeddings of nodes in the anatomical network inherently capture higher order topological relations between nodes. Finally we compare it to the case when we incorporate no predefined spatial layout into the GNN models, trying to learn the spatial structure by gradient descent based optimization during model training. Based on our comparisons we try to identify the most efficient STGNN architectures to investigate spatial and temporal dynamics in brain networks.

We then compare this STGNN based approach to the currently popular data-driven approach for modeling directed relations between neuron pools. Granger causality is based on the idea that if one event $A$ would cause another event $B$, then $A$ should precede $B$ and the occurrence of event $A$ should contain information about the occurrence of event $B$ [26]. In the context of neuroimaging this is realized in a predictive framework, by testing if adding information on activity in a region $A$ improves the prediction of activity in region $B$. The underlying predictive model in Granger causality is usually based on a vector auto regression (VAR) for multivariate timeseries inference [26, 8]. In a brain network with $N$ regions of interest (ROIs) the parameters in a VAR model grow with $N^2$, so for larger brain networks it can be challenging to accurately fit the model if only limited data are available. This can be problematic in fMRI, where the temporal sampling rate is relatively low, while its good spatial resolution would allow for a detailed, high-resolution network analysis. Therefore it would be desirable to have a predictive model that can learn interactions between all brain areas of interest, and in addition naturally scales to larger brain networks. In our study we compare the STGNN approaches to a classical VAR model and test their accuracy on a variety of network sizes and data set sizes. We show that by learning localized functional interactions based on the anatomical network, the STGNN models are able to accurately replicate functional dynamics even when networks become very complex and only few data are available. This demonstrates that the STGNN approaches are robust among a larger variety of MRI study scenarios, and are also suitable for the analysis of smaller subject cohorts, like in studies of patients with rare neurological diseases.

Finally we study the spatial interactions between brain regions, which are learned by the STGNN models. By integrating prior knowledge on the brain anatomy in form of structural connectivity or based on connectome embeddings, such models can provide multimodal perspective on directed relations between brain areas. Recent approaches that investigate the structure-function relation in the brain are based on computational modeling [36, 21, 52, 17, 53], graph theory [47, 1, 9, 49] and machine learning [23, 3, 57]. These approaches have already contributed numerous insights into the relationship between functional and structural brain networks, like those that explain how functional coherency patterns emerge in biophysically inspired models constrained by anatomical connections [36, 53], or how indirect structural connections contribute to the inference of FC strength [47, 9]. Therefore the vast majority of studies focuses on inferring overall FC patterns from their SC, although static coherency based measures of FC might have limitations in their ability to capture the rich nature of dynamic brain activity [78]. To the con-
trary, STGNNs are able to directly predict the measured BOLD dynamics, and their interactions between brain regions, without relying on the indirect representation of functional dynamics based on coherency. This characteristic of STGNNs allows us to additionally investigate the structure-function coupling in brain networks from a novel perspective.

2 Results

2.1 Graph neural network models

In our context of MRI the goal of the spatio-temporal GNNs will be to replicate the observed BOLD signal as accurately as possible in order to accurately describe the spatio-temporal dynamics of the underlying mechanisms in the brain. The learning objective can be formalized by introducing a graph signal $x(t) \in \mathbb{R}^N$, representing the BOLD signal measured at timestep $t$ in $N$ different brain regions. The goal of the models is to predict from an input sequence of $T_p$ past neural activity states $t = 1, \ldots, T_p$ a sequences of future states $t = 1, \ldots, T_f$. In addition to the temporal information, also spatial dependencies are included in the GNN architectures. The spatial relations between the $N$ brain regions can be represented in the notion of a graph $G = (V, E, A_w)$, containing vertices (nodes) $V$, with $|V| = N$, and edges $E$. The structure of the graph is characterized by a weighted adjacency matrix $A_w \in \mathbb{R}^{N \times N}$, where an entry $w_{nn'}$ describes the connection strength between brain region $n$ and $n'$. An overview of the graphical representation of a dynamic brain state is provided in figure 1 (A). Based on this concept, the task of the GNN models is to derive a function $h(\cdot)$ that best predicts $T_f$ future activity states from an input sequence of $T_p$ past states:

$$\left[ x^{(1)}, \ldots, x^{(T_p)}; G \right] \xrightarrow{h(\cdot)} \left[ x^{(T_p+1)}, \ldots, x^{(T_p+T_f)} \right]$$

Until now various spatio-temporal GNN architectures have been proposed to account for spatial and temporal dependencies of such graph structured signals [81]. In timeseries analysis, recurrent neural networks (RNN) [58] provide one efficient way to detect patterns in sequential data structures, like in our context the BOLD signal subsequently sampled at different timesteps $t$. In a RNN based sequence-to-sequence architecture an encoder recursively processes an input sequence of $T_p$ past neural activity states $x^{(t)}$ and encodes the temporal information into a hidden state $H(T_p)$ [69]. Next a decoder network uses the information in $H(T_p)$ to generate a prediction for $T_f$ future activity states. To account for vanishing gradients during training, the encoder and decoder consist of gated recurrent unit (GRU) cells [18]. An alternative for detecting repetitive patterns in sequential data is provided by convolutional neural networks (CNNs). By employing one-dimensional convolutions in the time domain they are used in our context to process temporal dynamics of neural activity. To efficiently capture long-term dependencies in temporal data the WaveNet (WN) architecture has been proposed [75]. This model introduces dilated causal convolution operations to generate a large receptive field when using only relatively few network layers, which alleviates the processing of long temporal input horizons. More recently also attention mechanisms have been proposed to detect underlying hidden correlations in sequential data structures [77]. The idea is to adaptively focus on the most important feature in a sequence.
These different fundamental approaches for temporal dependency modeling have been recently combined with techniques to additionally capture spatial relationships in graph structured signals [46, 82, 87]. Graph convolutional neural networks can be incorporated to model the propagation of information between adjacent nodes in the graphical representation of the signal [22]. The neighborhoods of the vertices/nodes \( V \) in the network are characterized by the adjacency matrix \( A \). In our study we additionally investigate different possibilities for defining the spatial layout for the information propagation between brain regions. As a first choice for the adjacency matrix we will employ the structural connectivity \( A_{SC} \) between the \( N \) brain areas, as it could be reconstructed from DTI data. This choice is motivated by the idea that white matter connections obtained from this modality would establish the anatomical backbone for information exchange between brain areas. In a recent study Rosenthal et al. [57] demonstrated that connectome embeddings (CE) can be utilized for projecting the structural connectome into a continuous vector space, which captures meaningful correspondences between different brain areas. This technique allowed to additionally account for long range and inter-hemispheric homotopic connections, which are usually only weakly expressed in DTI based anatomical connectivity [70]. In our study we utilized this technique to represent the edge weight \( w_{nn'} \) in the adjacency matrix as the similarity between the vector representations of two nodes \( n \) and \( n' \), which will be denoted as \( A_{CE} \). The information is accordingly propagated between brain regions which possess high similarity based on their neighborhood role within the anatomical network. Finally we compare these techniques to the case when the model is given the freedom to learn spatial dependencies between the \( N \) regions itself. In this setup the adjacency matrix is represented by a self-adaptive matrix \( A_{Adap} \in \mathbb{R}^{N \times N} \), which is learned during the model optimization. In this study we compare these different temporal and spatial techniques with each other as illustrated figure 1 (B). A detailed formal description of the model architectures and the training involved is outlined in section 4 ‘Materials and methods’. In the following we will assess the effectiveness of the different spatial and temporal modeling approaches by comparing their predictive performance on a MRI dataset from the Human Connectome Project (HCP) [76].

2.2 Data description

For the different evaluations in this study, resting-state fMRI data provided by the HCP S1200 release was incorporated [28]. To define the nodes of the brain network, the multimodal parcellation proposed by Glasser et al. [27] was applied, which is composed of 180 segregated regions within each hemisphere. The average of the BOLD signal was computed within each brain region, so for each resting state session, \( N = 360 \) time courses were obtained (180 per hemisphere). During one session \( T = 1200 \) fMRI images were collected, so the ROI timeseries can be collected in a data matrix \( X \in \mathbb{R}^{N \times T} \). We filtered the resting state fMRI timeseries data with a \( 0.04 - 0.07 \) Hz narrow band bandpass filter, because it has shown to be reliable and functionally relevant for gray matter activity [29, 14, 20, 12, 2].

For learning the predictions of the BOLD signal, samples of input and output sequences were generated from the timeseries data in \( X \) [78]. This was achieved by selecting windows of length \( T_p \) to obtain input sequences of neural activity states \( [x^{(1)}, \ldots, x^{(T_p)}] \), and respective target sequences of length \( T_f \) denoted as \( [x^{(T_p+1)}, \ldots, x^{(T_p+T_f)}] \). The time index \( t \) was propagated through each fMRI session, where in total \( T - T_p - T_f + 1 \) input-output pairs were generated per session. For
Figure 1: In (A) the spatio-temporal representation of a signal in a brain network is illustrated. The temporal component is given by the BOLD signal $x(t) \in \mathbb{R}^N$ in $N$ brain regions sampled at different timesteps $t$. The strength of the edges in the brain network are defined by the weighted adjacency matrix $A_w \in \mathbb{R}^{N \times N}$, which characterizes the spatial relation between all $N$ regions. In our study we compare different approaches in graph neural network (GNN) architectures to learn temporal and spatial dynamics in brain networks. Like illustrated in (B), temporal mechanisms are used to model dependencies in the signal between different timesteps $t = 1, 2, 3, \ldots$ in a certain region $i$. For this purpose we compare the performance of recurrent neural networks, convolutional neural networks and temporal attention networks to each other. Graph convolutional operations allow us to model the propagation of information between a region $i$ and other regions $j, k, l, \ldots$ in the network. We evaluate the model performances when using either the structural connectivity ($A_{SC}$), the connectome embedding similarity ($A_{CE}$) or a self-adaptive matrix ($A_{Adap}$) as an adjacency matrix in the GNN models.

For the following comparisons, the length of the input and output sequences were selected to be $T_p = T_f = 60$, which corresponds to a time span of roughly 43 s, based on a sampling interval
of $TR = 0.72 \text{ s}$ [74]. This time window has been shown to be long enough to be sufficiently challenging for the models and to make clear the differences in their performance. Likewise, the time window of 60 timepoints is short enough for them to make reasonable non-random forecasts of the BOLD signal.

In addition to the functional dynamics in the different brain regions derived from fMRI, the structural connectivity between those regions was reconstructed from DTI data. For this purpose the DTI dataset in the HCP S1200 release was processed using the multi-shell, multi-tissue constrained spherical deconvolution model [40], made available in the MRtrix3 software package [73]. White matter tractography was performed to estimate the anatomical connection strength between the regions defined by the multimodal parcellation atlas [27]. The number of the streamlines which connect two atlas regions was used to determine the structural connectivity values between the $N$ brain regions, which were then collected in a structural connectivity matrix $A_{SC} \in \mathbb{R}^{N \times N}$. A detailed description of the MRI datasets and their preprocessing is provided in section 4.7 ‘Dataset’. In addition the embeddings of the nodes within the structural network $A_{SC} \in \mathbb{R}^{N \times N}$ were generated using the node2vec algorithm [32]. The parameters for this algorithm are outlined in detail in section 4.3 and further Pearson correlation was used to quantify the degree of similarity of structural nodes in their connectome embedding space. The pairwise similarities between the $N$ nodes were then collected in the matrix $A_{CE} \in \mathbb{R}^{N \times N}$.

2.3 Comparison of GNN architectures

Before evaluating the performance of different models on a larger variety of MRI study scenarios, we will first focus on the effects of different temporal and spatial modeling techniques. For this purpose a dataset with a sample size of a medium sized fMRI study including 25 subjects will be incorporated. Each resting-state fMRI session was decomposed into pairs of input and output samples, like those described in section 2.2, and the generated training, validation and test samples were then aggregated across the 25 fMRI sessions. The neural signal of regions within the right hemisphere [28], consisting of $N = 180$ ROIs, will be included in the following comparison. At first we evaluate the prediction accuracy of the different temporal modeling strategies. For this purpose we compare the recurrent neural network (RNN) model, with the WaveNet (WN) model and the temporal attention (TAtt) model. The influence of the model hyperparameters, which are used in the following comparisons are described in section 4.5 and discussed in detail in supplement 1. The BOLD signal data was scaled to zero mean and unit variance for the evaluations, to obtain values of a magnitude that is easier to interpret. Figure 2 (A) shows the test mean absolute error (MAE) between the predicted and the true neural activity. The error was computed as the average across all test samples, brain regions and the 60 predicted time points (corresponding to roughly $43 \text{s}$ of activity). The comparison shows that RNN and WN have very similar capabilities in predicting the BOLD signal, while the TAtt model exhibits a worse performance. Despite their conceptual differences this shows that the RNN and WN based approach both recover a comparable and consistent amount of temporal information from the fMRI data. In comparison to these, the TAtt architecture appears to be less able to capture the non-linear characteristics of the BOLD signal with this limited amount of data in fMRI studies. For this reason in the following we will focus on RNN and WN based approaches for identifying suitable models to model functional dynamics in brain networks.
Figure 2: Figure (A) shows a comparison of different modeling strategies for temporal dynamics in the BOLD signal, comparing the test MAE of the recurrent neural network (RNN), the WaveNet (WN) and the temporal attention (TAtt) architecture. The overall test error was computed as an average across samples, brain regions and subject sessions. In comparison to the RNN and GWN, the prediction error of the TAtt is considerably higher, and we therefore focus on RNN and WN based approaches for studying the spatio-temporal dynamics in the following. Spatial relations are added to the temporal models in form of graph convolutions, and the spatio-temporal extension of the RNN and WN models are respectively denoted as diffusion convolution recurrent neural network (DCRNN) and Graph WaveNet (GWN) [46, 81]. Spatial transitions are based on the relations of network nodes captured in a weighted adjacency matrix, which is either based on structural connectivity ($A_{SC}$), connectome embedding similarity ($A_{CE}$), or adapted during model training ($A_{Adap}$). In (B) the adjacency matrix $A_{SC}$ based on structural connectivity within the 180 regions of the right hemisphere is illustrated, together with the adjacency matrix $A_{CE}$ derived from structural connectome embedding similarities. The regions in this illustration are ordered according to the atlas proposed by Glasser et al. [27]. Figure (C), (D) and (E) show the prediction accuracies of the DCRNN and GWN model in dependence of the walk order $K$. Note that $K = 0$ corresponds to the case when there is no spatial information exchange between network regions included. In figure (C) the overall test MAE is shown when incorporating the SC as an adjacency matrix $A_{SC}$, figure (D) illustrates the test MAE when employing CEs in an adjacency matrix $A_{CE}$ to define spatial relationships, and (E) displays the case when using a self-adaptive weight matrix $A_{Adap}$. 

8
In the next step we will study the impact of adding information on spatial relations between the different regions in the brain network. This will be implemented by invoking graph convolution operations in the predictive models, as outlined in detail in section 4.3 'Spatial dependencies'. The definition of a adjacency matrix determines how information is propagated between the different nodes in our brain network, and in our evaluations we investigate three conceptually different possibilities. In the first approach we use the structural connectivity as derived from DTI as the substrate for information exchange between different ROIs. The SC based adjacency matrix $A_{SC}$ is illustrated in figure 2 (B). The information can propagate along direct connections in the graph ($K = 1$), but also higher orders ($K = 2, 3, \ldots$) expressing the influence of indirect connections can considerably contribute to interactions between different areas in the brain [9, 47, 11]. A walk order of $K = 0$ denotes the case when including no spatial information exchange between network areas, exclusively incorporating temporal information for the predictions. Figure 2 (C) depicts the test MAE in dependence of the walk order $K$ when using the SC derived from DTI as a basis for information propagation in space. The RNN based model in combination with graph convolution operations is referred to as DCRNN [46] and the MAE of its predictions, averaged across test samples, brain regions and predicted timepoints is depicted in blue. Figure 2 (C) shows that it has the lowest test MAE when incorporating walks on the structural graph up to a order of $K = 2$. The WN incorporating graph convolution operations is denoted as GWN [82] and its average test MAE is shown in red in figure 2 (C). The influence of the walk order $K$ on the GWN accuracy suggests that its performance can be successively improved by including first-order connections, followed by the second- and third-order connections. As an alternative thereto, the structural similarity between ROIs can be based on their CE similiarity $A_{CE}$, like illustrated in figure 2 (B). The comparison between $A_{CE}$ and the structural connectivity matrix $A_{SC}$ highlights that in the adjacency relation defined by the structural embeddings, long range connections between brain regions are considerably more pronounced. Figure 2 (D) shows then the test MAE of the models when incorporating $A_{CE}$ in the graph convolution operations. In this case we can observe for both models a sharp drop in the error at walk order $K = 1$, what suggests that the similarity of node embeddings already inherently capture higher order relations between nodes in the brain network. Finally in figure 2 (E) the test MAE is shown when treating the connections between nodes as learnable weights. In this case we do not observe an improvement in the test error, what indicates that it is rather challenging to learn all $N^2$ connections between brain regions without prior knowledge. In general both STGNN models could profit the most when using CEs to characterize the spatial layout for functional interactions between brain regions. For the DCRNN the test error was $MAE = 0.1388$ when incorporating no information from other brain regions in the network, and could be reduced to $MAE = 0.1158$ (for $K = 1$) when using CEs to model the information exchange within the brain network. To test the significance of this improvement, the overall test MAE for each subject was computed and based on a paired t-test the impact of structural modeling on the model accuracy was significant with $p \leq 0.0001$. This demonstrates that around 17% more information on functional dynamics can be directly retrieved from nodes with similar context within the structural network. Using the SC to model transitions could only reduce the MAE of the DCRNN by 5% at $K = 1$, which supports the idea that the structural node embeddings can strengthen the relationship between structural data derived from DTI with functional data observed in fMRI [57]. By inherently capturing higher order transitions in $A_{CE}$. 
only a low walk order $K$ is required to capture information from structurally connected ROIs. In this manner, this technique can help to efficiently reduce the number of necessary parameters to account for spatial dependencies in STGNN models.

### 2.4 Model accuracy and network scaling

In this section we study the prediction accuracy of the above introduced STGNN based approaches and compare it to the VAR model, which is currently the predominantly used model for directed functional connectivity analysis [26, 8]. In practicable applications the amount of available fMRI data may vary depending on the project size and on the recruited subject cohort. Also the size of the brain network of interest can range from a few specific areas in a single functional network to a large-scale whole brain analysis. For this purpose we consider different scenarios in our following evaluations, by analyzing the models accuracies in dependence of the brain network size and the fMRI dataset size. We consider one larger subject dataset consisting of resting-state fMRI sessions from 50 different subjects, one medium sized dataset of 25 subjects and one small dataset including data from 10 subjects. In addition we vary the size of the analyzed brain network. The first network consists of 22 ROIs per hemisphere involved in visual processing as defined by the Glasser parcellation [27] (a complete list of selected ROIs is provided in the supplement II). The second network includes the regions within one hemisphere, and for that purpose the 180 regions within the right hemisphere included in the Glasser atlas were selected [27]. Finally the whole brain network of in total 360 regions was incorporated. As discussed in section 2.2 ‘Data description’, windowed input and output time sequence pairs were created from the data and the goal of the different models is accordingly to predict $T_f = 60$ TRs of neural activity from the past $T_p = 60$ activity values. We fitted the VAR model using the ordinary least squares (OLS) method as implemented in the multivariate Granger causality (MVGC) toolbox [8], and for each dataset we selected the VAR model with order $p$ that achieved the best MAE on the test set, as outlined in more detail in section 4.6 ‘Vector auto regressive model’. The hyperparmeters used for the STGNNs are described in section 4.5 ‘Model training’. Further for this comparison the CE similarity $A_{CE}$ with transition order of $K = 1$ was used in the STGNN models, which has shown to improve the GNNs forecasting accuracy with low computational cost, as discussed in section 2.3 ‘Comparison of GNN architectures’.

Figure 3 shows the test accuracy of the VAR, DCRNN and GWN model in dependence on the dataset size and brain network size. It can be observed in figure 3 (A) that if a large dataset of 50 subjects is available, all models are able to accurately predict the BOLD signal with a low test MAE, and a notable increase in the test error only appears for the VAR model, when it is fitted to the whole brain network. Figure 3 (B) shows the test MAE when data from 25 subjects is incorporated. In this case the test error of the VAR model starts to increase noticeably when modeling activity distributions within one hemisphere and becomes quite large when including the whole brain network. In contrast to these, the prediction accuracies of the DCRNN and GWN models remain stable in all cases. Finally when only 10 subject datasets are available the test MAE of VAR model is highly dependent on the analyzed network size, as illustrated in figure 3 (C). The DCRNN and GWN model can still achieve a high accuracy also when sparse data are available and the network size is relatively large.

To illustrate the prediction accuracies of the different models in more detail, an example of
the predictions using the dataset including 25 subjects, and modeling the activity within one hemisphere is shown in figure 4. Figure 4 (A) shows the MAE of the models computed as an average across test samples and ROIs in dependence of the forecasting horizon. In this case within the first 15 predicted timesteps all three models can generate very accurate predictions, but after that period the error of the VAR model starts to accumulate, while the GNN based approaches remain considerably more stable and precise. The predicted BOLD signals of the different models in a few representative samples are shown in figures 4 (B), (C) and (D).

In addition we evaluated in more detail how the prediction error is distributed across different ROIs across the cortical surface. Figure 5 shows the test MAE of the DCRNN, GWN and VAR model in dependence of the location within the right hemisphere. For all three models we observe a consistently greater prediction error in the posterior cingulate cortex and medial orbitofrontal cortex, which could point towards a more complex BOLD dynamic in those regions. Alternatively, the prediction accuracy might be also affected by a lower signal-to-noise ratio observed in medial brain regions [56].

![Figure 3](image-url)

**Figure 3**: The figure shows a comparison of the model performances when varying the amount of data and the size of the network. The test MAE of the VAR is here depicted in orange, the MAE of the DCRNN in blue and the error of the GWN in red. The overall error was computed as an average across brain regions, timesteps and test samples. In (A) the test MAE using a dataset of 50 subjects is shown for the visual network, the network within the right hemisphere and the whole brain network [27]. Figure (B) and (C) show the test performances in dependence of the network size using the 25 and 10 subject dataset, respectively.
Figure 4: The prediction accuracy of the different models is presented in more detail for the 25 subject dataset and the brain network including the ROIs within the right hemisphere [27]. In (A) the test MAE in dependence of the forecasting horizon is shown, computed as an average across test samples and brain regions. Figure (B) shows a representative example of predictions generated by the VAR model, and the error of the predictions in this example are with $MAE = 0.376$ slightly below its overall test MAE. Figure (C) shows an example of GWN predictions and the error in this example is with $MAE = 0.137$ slightly higher than its average MAE. Finally (D) shows the predictions of the DCRNN and the error is with $MAE = 0.120$ slightly higher than the average error.
Figure 5: The distribution of the test error across the cortical surface is shown. In (A) the MAE across brain regions of the DCRNN is first visualized in a boxplot on the left side. Additionally on the right side of the figure, the MAE values are projected onto the cortical surface within the right hemisphere, where the colormap was linearly scaled between 0 and 0.18. In (B) the MAE of the GWN is shown across regions and in (C) the MAE values of the VAR model. For the VAR the colormap was adjusted to account for larger error values by scaling it between 0 and 0.6.
2.5 Multimodal directed connectivity

In the section 2.3 ‘Comparison of GNN architectures’ different approaches have been investigated to model functional interactions between segregated regions in the brain network. The results showed that incorporating information on the spatial relation between ROIs in the form of the structural connectivity or connectome embedding similarity could considerably improve the prediction accuracy of the GNNs models. This points out that the GNN are able to learn relevant and functional informative transitions of neural activity on their structural spatial layout. Based on the idea of Granger causality \[30\] that the observation of one event \(A\) carries information about about the occurrence of a future event \(B\), this might represent initial evidence for a potentially causal relation between \(A\) and \(B\). In this spirit, propagating the information between ROIs based on their SC or structural CE similarity can give us a multimodal perspective of such a directed relationship between brain areas. In the following we choose a perturbation base approach to reconstruct the amount of information one ROI carries about other ROIs \[85, 79\]. By learning a function \(h(\cdot)\), the GNN models try to infer from an input sequence of neural activity states \([x^{(1)}, \ldots, x^{(T_p)}]\) a sequence of future activity states \([\hat{x}^{(T_p+1)}, \ldots, \hat{x}^{(T_p+T_f)}]\), whereby \(x^{(t)} \in \mathbb{R}^N\) denotes the activity at timestep \(t\) in all regions \(n = 1, \ldots, N\). To induce a perturbation into the system of neural dynamics, we remove all activity in a certain ROI \(n'\) by setting its activity values to the sample mean \(x_{n'} = 0\). By using the perturbed timeseries as an input for our trained model \(h(\cdot)\) the model generates then a prediction \([\hat{x}^{(T_p+1)}, \ldots, \hat{x}^{(T_p+T_f)}]\). To reconstruct the directed influence of ROI \(n'\) on ROI \(n\) we evaluate the overall difference between the original prediction and the prediction with perturbation in the input:

\[
I_{n}(n') = \frac{1}{S} \sum_{s=0}^{S} \frac{1}{T_f} \sum_{t=0}^{T_f} |\hat{x}^{(t)}_{n}(s) - \hat{x}'^{(t)}_{n}(s)|
\]

where \(I_{n}(n')\) denotes the impact of ROI \(n'\) on \(n\). Further \(\hat{x}^{(t)}_{n}(s)\) and \(\hat{x}'^{(t)}_{n}(s)\) denote the predictions in ROI \(n\) with and without the perturbation in \(n'\) of one test sample \(s\) at time step \(t\).

In the following we compare this proposed measure of directed influence \(I(n')\) to the classical undirected types of brain connectivity. First we compare it to structural connectivity as derived from DTI, characterized by the number of fiber tracks connecting two brain regions. Then we incorporate functional connectivity, defined as the Pearson correlation of functional activity time-courses between two areas. We employ the above introduced GWN model to obtain a multimodal measure of directed connectivity \(I(n')\), first using the SC as substrate for information propagation, captured in \(A_{SC}\), and then also employing the similarity of CEs, represented by \(A_{CE}\). In the following example we study the connectivity of V1 within the right hemisphere by incorporating data of 25 subjects. For the comparison all connectivity values were rescaled by normalizing them between 0 and 100, and then were visualized by projecting them onto the cortical surface. In figure 6 (A) the structural connectivity is depicted and the target region V1 is marked here in light blue. The strength of connectivity to all other regions is encoded in red. Figure 6 (A) shows that we can mainly observe a pronounced structural connectivity between V1 and V2 and some structural connections leading to V3. Figure 6 (B) shows the undirected functional connectivity in resting state. In this type of connectivity we can observe predominantly correlations to the functional activity in V2 and V3, but also a considerable connection strength to V3, V4 and V6. In figure
In comparison to the SC, in this variant of brain connectivity we can observe in addition to V2 also a more pronounced relationship to areas V3 and V4, and to some anatomically more distant areas like V6 and the ventromedial visual area VMV1. This multimodal type of connectivity also reflects the role of indirect structural connections by modeling higher order transitions on the structural scaffold captured by the STGNN model. As an alternative to the SC, in figure 6 (D) the directed connectivity patterns when using CE similarity as the spatial layout in the GWN are displayed. Here we can see an even stronger integrity of V1 within the visual network, which is in agreement with the observation that CEs capture higher order topological information of anatomical connectivity [57]. Supplement III shows additionally the spatial relations learned by the DCRNN model. Here we can observe a pronounced similarity to the directed connectivity pattern learned by the GWN architecture, showing additionally strong relations to areas like V3 and V4. Based on this observation, such a GNN based connectivity approach can serve as a link between structural and functional connectivity and as such they can provide a multimodal perspective on directed influences between individual areas in brain networks.
Figure 6: Different types of connectivity are illustrated between V1 and all other regions within the right brain hemisphere (here only regions in the medial wall of the right hemisphere are illustrated). In (A) the structural connectivity is shown, whereby the target region V1 is marked in light blue and the connectivity strength is encoded in red. In (B) the correlation based functional connectivity is illustrated, which was computed as an average across subjects. Further (C) shows the measures of influence $I(n')$, derived from the GWN model using the SC for information propagation and figure (D) depicts the influence when incorporating CEs for the information exchange between ROIs. The values of the connectivity measure were linearly mapped between 0 and 100 (and between $-100$ and 100 for FC). The default scaling of the color values provided by the connectome workbench (version 1.4.2) was used, adjusting the colormap between the $2th$ and $98th$ percentile of the values respectively.
3 Conclusion

In this study we have compared different STGNN architectures for learning the spatio-temporal dynamics in brain networks. First in section 2.3 we studied different mechanisms for learning the temporal dynamics in the BOLD signal. We could show that a RNN based model and a WN based model exhibit very similar capabilities in learning the temporal characteristics in neural activity timeseries. Despite their conceptual differences in their architectures, they demonstrated almost the exact same prediction accuracy, which indicates that they are both very consistent in capturing the temporal information in the data. As an alternative, we also studied TAtt mechanisms to learn temporal characteristics of neural signals. The TAtt model showed to be less suitable to model the dynamics in the BOLD signal with a limited amount of fMRI data. Despite incorporating techniques into the TAtt model that in general stabilize the learning, like residual connections and batch normalization [34, 37], its prediction error was considerably higher in comparison to the RNN and WN based approach. This indicates that the geometric assumptions which are realized by the temporally structured inference in the RNNs and WNs based on either recurrent computations or causal convolutions can contribute to the learning of the temporal characteristics of the BOLD signal. We then studied the impact of adding spatial dependencies to the temporal models, realized by invoking graph convolution operations. We have compared different spatial layouts for information propagation between ROIs, and therefore included either the structural connectivity (A_{SC}), the CE similarity (A_{CE}), or a self-adaptive adjacency matrix (A_{Adap}) into the STGNN models. While the model performance of the GWN and DCRNN steadily improved with higher walk orders $K$ on the anatomical substrate, we could observe a more pronounced improvement already when using CEs with a walk order of only $K = 1$. This embedding strategy turns out to be therewith also interesting in applications of STGNNs, because it helps to effectively incorporate indirect structural connections with low computational cost. In addition, the observed characteristics of CEs in our application support the ideas of Rosenthal et al. [57], which showed in their study that embeddings of the structural network can naturally capture higher order topological relations between ROIs within the structural layout. In our context of modeling spatio-temporal dynamics this method also proved to strengthen the relationship between brain structure and functional dynamics.

In section 2.4 we then compared the STGNN models to a VAR model, which is currently the predominant model used in Granger causality analysis for inference of directed relationships between brain regions [8]. We evaluated the accuracy of the different approaches on a variety of brain network sizes and data set sizes to account for different possible scenarios in their application in fMRI studies. The results showed that if a sufficiently large cohort of 50 subjects is available a VAR model is able to make very reliable long-term predictions, and only for a large network consisting of $N = 360$ there is a notable increase in the prediction error. But the dependency of the accuracy on the network size $N$ becomes more apparent when data from only 25 subjects are used to fit the VAR model, and when only 10 subjects are available, the error grows strongly with $N$. This demonstrates that a VAR is a very reliable and fast model for fMRI studies with a sufficiently large test subject size and for connectivity studies including a limited amount of predefined regions. However it can be desirable in some cases to include a larger amount of brain areas into the connectivity analysis, in order to avoid omitting relevant areas in the network of
interest. Also in MRI studies it can be very costly and time-consuming to collect a large amount of data, which is, for example, especially challenging in studies on rare neurological disorders. We could show in our study that the STGNN based approaches are able to make robust inferences also on large networks and when only limited data are available, thereby providing a considerably more flexible method for different network analysis scenarios. While the number of parameters in a VAR based approach scale with $N^2$ the spatial modeling in STGNNs based on localized graph convolutions becomes independent of $N$, allowing for an efficient scaling to large graphs.

Finally in section 2.5 we studied the spatial interactions within the brain network which were learned by the STGNN models. By integrating information on the anatomical connectivity into the GNN based models, we could derive a multimodal connectivity measure for directed and potentially causal relationships between brain regions. When comparing this measure of influence to the original structural connectivity, we can observe that STGNN have learned to include transitions along higher order structural connections in the network. The models could infer links between $V_1$ and $V_2$, but additionally strong connections to $V_3$ and $V_4$. Especially when incorporating the CE based similarity $A_{CE}$ to define spatial node relations in the STGNN models, we can observe a high integration of $V_1$ within the visual system. However due to the relatively low temporal sampling rate in fMRI, and the indirect measurement of neural signals based on their hemodynamic response, one should also be aware of these limitations of direct or causal relations in fMRI [26]. Still such measures can provide initial evidence for a causal relation between brain areas and in general supplement the analysis on connectivity in brain networks.

In conclusion we found the DCRNN and GWN architecture are both suitable for the task of functional dynamics inference. Using CEs to characterize the structural similarities between brain regions could further improve their prediction accuracy. Their robust scaling properties and the possibility to combine the information in structural and functional MRI data reveal the potential of STGNNs in the field of brain connectivity analysis. Besides their applications in fMRI, other functional neuroimaging techniques like electroencephalography (EEG) or magnetoencephalography (MEG) might be interesting for analyzing temporal dynamics with STGNN in the high frequency range. Also alternative structural imaging techniques like neurite orientation dispersion and density imaging (NODDI) [86] might capture additional aspects the brain structure, which could be included as structural information in STGNN based models. Still research on GNNs is a relatively new field in machine learning and recent developments in this field can make interesting contributions to our understanding of information processing in brain networks [19, 60].

4 Materials and methods

4.1 Graph neural networks

Different brain areas communicate via bioelectrical signals transmitted along neuronal axons and collected by neuronal dendrites. Spatio-temporal GNNs provide a novel possibility to incorporate such a structural scaffold into a graph-based prediction model [79]. Due to cognitive information processing in the brain, the spatial interactions of the activity distribution changes dynamically. Spatio-temporal GNNs thus encompasses both the information about the layout of the physical scaffold encoded by the graph structure and the dynamical information about temporal activity
correlations. Recently we used a DCRNN architecture to model the spatio-temporal brain dynamics in resting state fMRI [79]. In this study, spatial dependencies of brain activities were modeled via diffusion convolution operations based on the anatomical connectivity and the temporal dynamics of the graph signal were captured in an RNN based model architecture [46]. In our current study we evaluate some alternative spatial and temporal approaches to model dynamics in brain networks. In addition to RNNs, a CNN based architecture for temporal modeling has been introduced by Wu et al. [82]. These authors built upon the WaveNets [75] and stack dilated causal convolution layers to capture long-range temporal dependencies. Dilated convolutions support exponentially growing receptive fields in deeper layers of the network and allow us to handle long-range temporal sequences efficiently [75]. In addition to the temporal processing based on RNNs and CNNs, we also follow ideas expressed in attention networks and incorporated a relevance score that was computed in temporal attention layers [77, 87].

Based on these temporal approaches, we further study different concepts for representing the spatial dependency between brain regions. First we integrated the SC reconstructed from DTI to represent the anatomical substrate for information propagation in graph convolution operations. Then we additionally incorporated CEs of the structural graph to inherently capture higher order relations between ROIs. Finally we used no predefined spatial layout and treated the spatial connection strengths between ROIs as free parameters. These different spatio-temporal GNN architectures have to the best of our knowledge not been applied yet to analyze the dynamics of brain networks, and in our study we investigate their effectiveness in spatio-temporal modeling of functional MRI.

4.2 Preliminaries

Let us represent the brain network as a graph. Every specific brain area or region of interest (ROI), then forms a node in the graph. Let these \( N \) ROIs form an graph \( G = (\mathcal{V}, \mathcal{E}, A_w) \) encompassing \( N \) vertices, i.e. the meta-voxels or ROIs, and a set \( \mathcal{E} \) of edges connecting the vertices \( v_n, v_{n'} \). The graph structure can then be captured in a weighted adjacency matrix \( A_w \in \mathbb{R}^{N \times N} \), whose entries \( w_{nn'} \) provide the connection strengths between nodes \( v_n \) and \( v_{n'} \) and implicitly define the spatial structure of the graph. As introduced above, in our study we compared three different variants to define the spatial relationship between ROIs. Once we incorporated the SC derived from DTI data as an adjacency matrix \( A_{SC} \), we next employed CE to additionally capture higher order topological features in SC represented by \( A_{CE} \), and finally we treated the spatial relations as adaptive learnable parameters \( A_{Adap} \) in the GNN models. The dynamics of the graph signal is then represented by the time-varying neural activity obtained from functional imaging data. Let us first assume that each node of the graph is associated with a single feature represented by the BOLD activity. By considering voxel time series of brain activity maps, then all data can be collected into a data matrix \( X = (x^{(1)}, \ldots, x^{(T)}) \in \mathbb{R}^{N \times T} \) with \( x^{(i)} \in \mathbb{R}^N \). Given \( N \) ROIs, taken from a brain atlas and each represented by a meta-voxel, and considering \( T \) time points for each meta-voxel time series, which represents the activation time course of one of the ROIs, then we have, for the BOLD feature represented at node \( n \) a related graph signal matrix or BOLD feature matrix:
\[ \mathcal{X}_{:m} \equiv \mathbf{X}^{(m)} = (\mathbf{x}^{(m)}_1 \ldots \mathbf{x}^{(m)}_T) = \begin{pmatrix} x^{(m)}_{11} & \cdots & x^{(m)}_{1T} \\ \vdots & \ddots & \vdots \\ x^{(m)}_{N1} & \cdots & x^{(m)}_{NT} \end{pmatrix} \in \mathbb{R}^{N \times T} \] (3)

Note that the columns \( \mathbf{x}^{(m)}_t \in \mathbb{R}^N \) of the data matrix describe the activation of all ROIs at any given time point \( 1 \leq t \leq T \), while its rows \( \hat{\mathbf{x}}^{(m)}_n \) represent the meta-voxel time course of every single ROI \( 1 \leq n \leq N \). More generally, if nodes not only represent a single feature \( m \), like the input BOLD signal, but an \( M \)-dim feature vector \( \mathcal{X}_{nt} \in \mathbb{R}^M \), then we obtain a feature tensor \( \mathcal{X} \in \mathbb{R}^{N \times T \times M} \), whose frontal, lateral (vertical) and horizontal slices, respectively, read

\[ \mathcal{X}_{::m} \equiv \mathbf{X}^{(m)} \] of the data tensor \( \mathcal{X} \), we thus have the lateral tensor slices:

\[ \mathcal{X}_{:t} \equiv \mathbf{X}^{(t)} = (\mathbf{x}^{(t)}_1 \ldots \mathbf{x}^{(t)}_M) = \begin{pmatrix} x^{(t)}_{11} & \cdots & x^{(t)}_{1M} \\ \vdots & \ddots & \vdots \\ x^{(t)}_{N1} & \cdots & x^{(t)}_{NM} \end{pmatrix} \in \mathbb{R}^{N \times M} \] (4)

and the horizontal tensor slices:

\[ \mathcal{X}_{n::} \equiv \mathbf{X}^{(n)} = (\mathbf{x}^{(n)}_1 \ldots \mathbf{x}^{(n)}_M) = \begin{pmatrix} x^{(n)}_{11} & \cdots & x^{(n)}_{1M} \\ \vdots & \ddots & \vdots \\ x^{(n)}_{T1} & \cdots & x^{(n)}_{TM} \end{pmatrix} \in \mathbb{R}^{T \times M} \] (5)

Note that the column fibers of the data tensor \( \mathcal{X}_{:t,m} \), denoted as \( \mathbf{x}^{(m)}_t \), represent, at every time point \( t \), the distribution of the activity of feature \( m \) across all nodes \( n \) of the graph. Correspondingly, the row fibers of the tensor \( \mathcal{X}_{n:,m} \), denoted as \( \mathbf{x}^{(m)}_n \), represent the time course of every feature \( m \) at node \( n \). Finally the tube fibers of the tensor \( \mathcal{X}_{:t,n} \), denoted as \( \mathbf{x}^{(n)}_t \), represent the distributions of features at every node \( n \) and time point \( t \). This notation will in the following provide the framework to introduce the different techniques to model either dependencies between nodes \( n \), time \( t \) or features \( m \).

4.3 Spatial dependencies.

**Diffusion convolution:** In the following we provide a short introduction on a variant of graph convolution denoted as diffusion convolution in the context STGNNs [46, 81]. The information flow in the underlying graph \( G = (\mathcal{V}, \mathcal{E}, \mathbf{A}_w) \) is considered as a stochastic random walk process modeled by a state transition matrix \( \mathbf{T} = \mathbf{D}^{-1} \mathbf{A}_w = (\tilde{\mathbf{w}}_1 \ldots \tilde{\mathbf{w}}_N) \) where \( \mathbf{A}_w \) represents a weighted adjacency matrix. The diagonal node degree matrix is given by:

\[ \mathbf{D} = diag(\mathbf{A}_w \mathbf{1}) \] (6)
where \( \mathbf{w}_n = (\mathbf{w}_{1n}, \ldots, \mathbf{w}_{Nn})^T \in \mathbb{R}^N \forall n = 1, \ldots, N \) with \( \hat{w}_{nn'} = w_{nn'}/\sum_{n'} w_{nn'} \) denoted normalized edge strengths. State transitions were modeled as a diffusion process on an unstructured graph. The former was represented by a random walk Laplacian:

\[
L_{rw} = \mathbf{I} - \mathbf{T} = \mathbf{U} \hat{\mathbf{A}} \mathbf{U}^T = \mathbf{U} (\mathbf{I} - \mathbf{\Lambda}) \mathbf{U}^T = \mathbf{I} - \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T.
\]

(7)

where the transition operator \( \mathbf{T} \) was replaced by its eigen-decomposition with \( \mathbf{U} \) the matrix of eigenvectors and \( \mathbf{\Lambda} \) the diagonal matrix of eigenvalues. Hence, the set of eigenvectors \( \mathbf{u}_n \) provided an orthogonal basis system for the spatial representation of the brain graph. With the help of these eigenvectors \( \mathbf{u}_n \) the spatial structure of the graph could be implemented. A spectral representation in combination with the convolution theorem then provided a definition of the graph convolution operator \( G_C [63] \), which served to compute the spatial convolution of the input signal and a spatial filter kernel to yield the output of the \( \ell \)-th convolution layer as:

\[
y^{(q)}_t = \mathbf{U} \Theta^{(q)}_x \mathbf{x}_t^{(m)} = \mathbf{U} \Theta^{(q)} \mathbf{U}^T \mathbf{x}_t^{(m)} \approx \sum_{k=0}^{K-1} \phi^{(q)}_k(\omega)^T \mathbf{x}_t^{(m)}
\]

(8)

Here the approximation resulted from a power series expansion of the convolution kernel with respect to the eigenvalue matrix \( \mathbf{\Lambda} \) of the transition operator \( \mathbf{T} \) [22]. Finally considering a CNN architecture and applying the graph convolution operator \( G_C \), the filtered input signal \( y^{(q)}_t \) was transformed with an activation function \( \sigma(\cdot) \) to yield the output \( h^{(q)}_t \) of each of the \( q \in \{1, \ldots, Q\} \) graph convolution layers as follows:

\[
h^{(q)}_t = \sigma (y^{(q)}_t) = \sigma \left( \sum_{k=0}^{K-1} \phi^{(q)}_k(\omega)^T \mathbf{x}_t^{(m)} \right)
\]

(9)

Hereby \( \mathbf{x}_t^{(m)} \in \mathbb{R}^N \) denotes the \( m \)-th input feature component at time \( t \), \( h^{(q)}_t \in \mathbb{R}^N \) the corresponding output component of the \( q \)-th convolution channel, \( \Theta^{(q)}_k \in \mathbb{R}^N \) parameterizes the \( q \)-th convolutional kernel of order \( k \) and \( \sigma(\cdot) \) denotes any suitable activation function. Note that for deeper convolution layers \( \ell > 1 \), \( \ell = 1, \ldots, L \), the input to the convolution \( \ell \)-layer is given by the output component of the convolution layer \( \ell - 1 \) instead of the input signal. In summary, these graph convolution layers can learn to represent graph structured data and could be trained with gradient descent based optimization techniques.

**Structural connectivity:** One possibility to define the spatial layout of the brain network characterized by the weighted adjacency matrix \( \mathbf{A}_w \) is to directly incorporated the structural connection strength as reconstructed from DTI data. The weights \( w_{nn'} \) in our adjacency matrix would accordingly reflect the number of fibers connecting two brain regions \( n \) and \( n' \), derived from probabilistic fiber tracking [72]. This type of structural adjacency relation is denoted as \( \mathbf{A}_{SC} \in \mathbb{R}^{N \times N} \). The acquisition parameters of the DTI data and the structural connectome generation are outlined in detail in section 4.7.
Connectome embeddings: As an alternative to the original SC, connectome embeddings (CEs) can generate node embeddings that capture also higher order topological features of the structural layout [57]. The idea of such a graph embedding is to represent each node in the graph by a $M$-dimensional feature vector. This technique is originally inspired by the word2vec algorithm introduced by Mikolov et al. [54] who proposed a technique to learn vector-valued representations for words in a text which preserve linguistic regularities in their embedding space. Similarly the node2vec algorithm can be used to embed vertices of a graph into a subspace where similar embeddings capture the $k$-step ($k = 1, 2, \ldots, K$) relation between the vertices and their $k$-step neighbors [57, 32]. We used this technique to embed each brain region $n$ in the SC graph into a 64-dimensional vector representation. We therefore employed the gensim python package [89] using the skip-gram model to learn the node representations [54]. Briefly, in this context the idea of the skip-gram model is to predict from a target node in a network its neighboring nodes, whereby a sequence of neighboring nodes is created by performing a biased random walk on the structural graph [32]. To generate the node sequences in total 100 random walks were performed for each node with walk a length of 80 nodes. The return parameter of the random walk was set to $p = 2$ and the in-out parameter to $q = 1$. The similarity between the $N$ brain regions in their embedding space was computed using the Person correlation coefficient, yielding a connectivity matrix denoted with $A_{CEF} \in \mathbb{R}^{N \times N}$. As illustrated in figure 2 (B) the embeddings could yield meaningful representations that revealed long-range connections between regions which were not present in the original SC [57].

Adaptive adjacency matrix: So far the spatial layout of the brain graph has been represented with the help of the orthogonal eigenbasis system $U$ of the transition operator proportional to the random walk Laplacian. This presupposed a thorough knowledge about the spatial structure of the underlying brain network that entered the related adjacency matrix. Remember that the weights of the adjacency matrix were deduced from DTI measurements based on SC or their CE similarity. However there may exist hidden activity correlations that are not represented in the original adjacency matrix used to construct the random walk Laplacian. Hence, one may wish to introduce an additional self-adaptive, normalized adjacency matrix $A_{Adap} \in \mathbb{R}^{N \times N}$ [82]. The latter has been constructed as a matrix of trainable weights $V_{Adap} \in \mathbb{R}^{N \times N}$, which were at first initialized as zero and then again optimized via gradient descent [42]. Inspired by the study of Wu et al. [82], a normalized self-adaptive adjacency matrix was computed as:

$$A_{Adap} = \frac{\sigma_n (V_{Adap})}{N}$$

(10)

The transformation function $\sigma(\cdot) \equiv \tanh(\cdot)$ confined the adaptive weights to the range $[-1, 1]$, which then were normalized by the number of nodes $N$ in the network. This self-adaptive adjacency matrix can help to uncover any hidden, still unknown dependencies between ROIs of a given graph structure. Thus it may extend any graph diffusion convolution layer to yield its output activity as:

$$h_{t}^{(q)} = \sigma \left( y_{t}^{(q)} \right) = \sum_{k=0}^{K} \left[ \theta_{k}^{(q)} (\omega) T^{k} + \theta_{k}^{(q), Adap} (A_{Adap})^{k} \right] x_{t}^{(m)}$$

(11)
Note that the normalized self-adaptive adjacency matrix $A_{Adap}$ may be considered as an additional transition operator here. In an attempt to decouple the temporal processing from any underlying spatial layout of the graph connectivity, the first term within parentheses may be skipped and the self-adaptive adjacency matrix may possibly identify the underlying graph structure from the data alone. This may be applicable to situations, where no predefined graph structure is known or involved. The output of the $q$-th convolution channel of can in this case be obtained with:

$$h_t^{(q)} = \sigma \left( \sum_{k=0}^{K} \theta_k^{(q)} (A_{Adap})^k x_t^{(m)} \right)$$

(12)

4.4 Temporal dependencies.

Recurrent neural networks: In the DCRNN model, the temporal variations of the signal $x_t^{(m)} \in \mathbb{R}^N$ in $N$ brain regions at $T_p$ past time points were explored with sequence to sequence learning in RNNs [69], where an encoder network compresses the information into a compact new representation. The latter is fed into a decoding network, which generates predictions of the graph signal at $T_f$ future time points representing the intended prediction horizon, as illustrated in figure 7 (A).

Given that the graph convolution operation to effect the spatial layout of graph structure at any time point $t$, temporal dynamics on the graph can be modeled in the DCRNN via GRUs [18]. The idea is to replace convolution operations in the spatial domain by corresponding matrix multiplications in the conjugate spatial-frequency domain employing the diffusion convolution operator. This leads to the diffusion convolution gated recurrent unit (DCGRU) [46]:

$$r_t^{(q)} = \sigma \left( G_C \left( \Theta_r^{(q)}, [x_t^{(m)} || h_{t-1}^{(q)}] \right) + b_r \right)$$

$$u_t^{(q)} = \sigma \left( G_C \left( \Theta_u^{(q)}, [x_t^{(m)} || h_{t-1}^{(q)}] \right) + b_u \right)$$

$$c_t^{(q)} = \tanh \left( G_C \left( \Theta_c^{(q)}, [x_t^{(m)} || (r_t^{(q)} \odot h_{t-1}^{(q)})] \right) \right) + b_c$$

$$h_t^{(q)} = u_t^{(q)} \odot h_{t-1}^{(q)} + (1 - u_t^{(q)}) \odot c_t^{(q)}$$

(13)

where $x_t^{(m)}$, $h_t^{(q)}$ denote the $m$-th input and $q$-th output graph signal feature component of the GRU, respectively, at time $t$ and $[x_t^{(m)} || h_{t-1}^{(q)}]$ denotes their concatenation. Also $r_t^{(q)}$, $u_t^{(q)}$ represent reset and update gates at time $t$, and $b_r$, $b_u$, $b_c$, respectively, denote bias terms. Furthermore, $\Theta_r^{(q)}$, $\Theta_u^{(q)}$, $\Theta_c^{(q)}$ denote the parameter sets of the corresponding filters. An illustration of the complete sequence to sequence architecture incorporating DCGRU cells is provided in figure 7.
Figure 7: The overview of the complete DCRNN model is provided in (A). The RNN architecture consists of an encoder and decoder, which recursively process the graph structured signals. The encoder receives a sequence of inputs \([x^{(1)}, \ldots, x^{(T_p)}]\) and iteratively updates the hidden state \(h^{(t)}\). The final state of the encoder \(h^{(T_p)}\) is passed to the decoder branch, which then recursively predict the output sequence of future signals \([x^{(T_p+1)}, \ldots, x^{(T_p+T_f)}]\). The encoder, as well as the decoder (B) consists of multiple diffusion convolution gated recurrent unit cells (DCGRU). The first DCGRU cell receives the input graph signal, and then passes its hidden state to the subsequent cell. During decoding, the final cell of the decoder then generates the predictions for the signal. For testing and validation, the decoder uses its own prediction as input for generating the subsequent prediction. The first input of the decoding branch (\(<\text{GO}\>\) symbol) is simply a vector of zeros. The processing steps in an individual DCGRU cell are shown in (C). The input \(x^{(m)}\), as well as the previous hidden state \(h^{(q)}\) are concatenated and passed to the reset gate \(r^{(q)}\), as well as to the update gate \(u^{(q)}\). The reset gate \(r^{(q)}\) determines the proportion of \(h^{(q)}\), which enters \(c^{(q)}\), together with input \(x^{(m)}\). Then the hidden state \(h^{(q)}\) is updated by \(c^{(q)}\), whereby the amount of new information is controlled by \(u^{(q)}\).
WaveNets: Rather than incorporating diffusion convolution layers into RNNs, dilated causal convolution (DCC) layers [75] have been instead employed in the GWN architecture [82]. The full GWN model is illustrated in figure 8. The DCC was defined through a dilated causal convolution operator $D_C$:

$$D_C \left( \Theta^{(q)}_t, x^{(m)}_t \right) = \sum_{r=0}^{R-1} \Theta^{(q)}_{r} x^{(m)}_{t-d \cdot r}$$  \hspace{1cm} (14)

whereby $d$ denoted the dilation factor and $\Theta^{(q)}_t$ represented the filter kernel. DCC could be implemented by sliding over the input time series $x^{(m)}_t$ while skipping input values while, from layer to layer, increasing step size $d \cdot r$. This procedure leads to an exponential growth of the receptive field with increasing layer depth as is schematically illustrated in figure 8 (C). The information flow was controlled by a gated temporal convolution network (TCN) as shown in figure 8 (B), which is obtained as:

$$h^{(q)}_t = \tanh \left( D_C \left( \Theta^{(q)}_1, x^{(m)}_t + b_1 \right) \right) \odot \sigma \left( D_C \left( \Theta^{(q)}_2, x^{(m)}_t + b_2 \right) \right)$$  \hspace{1cm} (15)

Here $\tanh(\cdot)$ denotes the output activation function, and $\Theta^{(q)}_1$, $\Theta^{(q)}_2$, and $b_1$, $b_2$ represent the convolution filers and bias terms respectively. Further $D_C$ represents the causal convolution operator, $\odot$ the Hadamard product and $\sigma(\cdot)$ denotes the logistic function, which controls the information passed to the next layer. To achieve large receptive fields, the layers in a WN architecture are organized in blocks, whereby in each block the dilation factor $d$ is doubled with $d = 1, 2, 4, \ldots$ up to a certain limit and then repeated in the same manner in the next block [75]. After each such dilated convolution layer a diffusion convolution layer $G_C$ (equation 9) is subsequently applied to account for the spatial dependencies, like illustrated in figure 8 (A).
Figure 8: An overview of the complete GWN model is provided in (A). The GWN model consists of $L$ layers. For the temporal modeling the GWN applies first the gated TCN mechanism and then for the spatial aspects utilizes graph convolution operations ($G_C$) in each layer. Each layer additionally incorporates residual connections to stabilize the gradient during learning [34]. The information in each layer is combined by using skip connections, and the final predictions are generated by passing the output of the skip connections through two fully connected layers. The gated temporal convolution network (TCN) mechanism (B) applies a dilated causal convolution ($D_C$) in combination with a $\tanh(\cdot)$ and a $\sigma(\cdot)$ activation function to control the information flow. In (C) the dilated causal convolutions are illustrated. In each layer a temporal convolution is applied whereby the dilation factor can be increased in subsequent layers. This dilations lead to exponentially growing receptive fields for neurons in higher layers. The receptive field of a neuron in layer is highlighted in blue.
Temporal relevance: Yet another approach to solve spatio-temporal time series prediction problems considers attention mechanisms in spatial and temporal domains to capture dynamic correlations [77, 87]. In this study we therefore additionally explore non-linear temporal correlations via a temporal relevance mechanism for modeling temporal fluctuations in the BOLD signal. Let the temporal state of the brain network be represented by the multivariate signal tensor $\mathcal{X} \in \mathbb{R}^{N \times T \times M}$ such that the temporal states of any node $n$ be collected in the signal matrix $\mathcal{X}_n \equiv X^{(n)} \in \mathbb{R}^{T \times M}$. The activity at any node $n$ and at any time $t$ was then represented by the tube fibers $\tilde{x}_t^{(n)} \in \mathbb{R}^M$, where $M$ denoted the number of features characterizing the node activity. Temporal correlations between different node states could be estimated by filtering the multivariate signals in a cascade of temporal relevance blocks, like illustrated in figure 9. The queries and keys are computed from the input in the $\ell$-th block at time point $t$ with a simple non-linear transformation $g_r(\tilde{x}_t^{(n)}) = \text{ReLU}(W_r \tilde{x}_t^{(n)} + b_r)$ with parameters $W_r \in \mathbb{R}^{D \times M}$ and $b_r \in \mathbb{R}^D$. For any node $n$ and any time point $t_i$ the relevance of its states $\tilde{x}_t^{(n)}$ at time points $t_j < t_i$ with respect to the considered state $\tilde{x}_t^{(n)}$ could then be assessed by computing the inner product between the queries and keys:

$$\delta_{t_i,t_j}^{(n)} = \frac{\langle g_r(\tilde{x}_t^{(n)}), g_r(\tilde{x}_t^{(n)}) \rangle}{\sqrt{D}}. \quad (16)$$

A normalized temporal relevance score $\hat{\delta}_{t_i,t_j}^{(n)}$ could then be computed according to

$$\hat{\delta}_{t_i,t_j}^{(n)} = \frac{\exp(\delta_{t_i,t_j}^{(n)})}{\sum_{t_j < t_i} \exp(\delta_{t_i,t_j}^{(n)})}. \quad (17)$$

Finally, $t_j < t_i, j \in \{1, \ldots, T_p\}$ denoted a set of time steps before time point $t_i$. After computing the temporal relevance score $\hat{\delta}_{t_i,t_j}^{(n)}$, the hidden state of node $n$ at time $t_i$ could be computed according to

$$\hat{h}_{t_i} = g_r \left( \sum_{t_j < t_i} \hat{\delta}_{t_i,t_j}^{(n)} \cdot g_r(\tilde{x}_t^{(n)}) \right). \quad (18)$$

Whereby $g_r(\cdot)$ denotes a non-linear projection again. Note that all parameters $W_r$ and $b_r$ to be learned were shared across all nodes and time steps. In total $L$ layers of temporal attention mechanisms were stacked to generate a final prediction for the BOLD signal. After each layer batch normalization was applied and additionally residual connections were incorporated to stabilize the gradient [34].
Figure 9: An overview over the temporal relevance or attention model. The single feature input $X^{(in)}$, representing the BOLD signal, is first projected by a fully connected layer onto $M$ output features. Then the temporal relevance scores are computed in the $L$ attention layers, which additionally account for vanishing gradients, additionally residual connects are incorporated [34]. The output of the final layer $L$ is then projected back onto a single feature, representing the predicted neural signal.
### 4.5 Model training

In this section we outline the training procedures which were used for the different neural network models to learn the temporal and spatial dynamics in the BOLD signal. Before training, the fMRI data of each session was scaled between 0 and 1 to confine the gradients in reasonable range. For all models, the mean absolute error (MAE) was used as an objective function to quantify the overall difference between the true BOLD signal \( x(t) \) and predicted signal \( \hat{x}(t) \) in all \( N \) brain regions:

\[
\text{MAE}(x, \hat{x}) = \frac{1}{N \cdot T_f} \sum_{n=1}^{N} \sum_{t=1}^{T_f} |x_n(t) - \hat{x}_n(t)|
\]  

(19)

**DCRNN** The DCRNN model, based on a RNN architecture, was trained with backpropagation through time [80], with the objective to maximize the likelihood of generating the target timeseries. To additionally account for a mismatch between training and testing distributions of stimuli, a scheduled sampling strategy was used [10]. The probability of using a true label as a decoder input decayed according to:

\[
e(i) = \frac{\tau}{\tau + \exp(i/\tau)} \in (0, 1)
\]  

(20)

with \( \tau > 0 \) the decay parameter and \( i \in \mathbb{N} \) counting the iterations. During supervised learning, instances to be predicted were, of course, known. For this optimization problem, the Adam algorithm [42] was employed, and the model was trained for 70 epochs on mini-batches of 32 training samples. To further improve convergence, an annealing learning rate was used, initialized as \( \eta = 0.1 \), and decreased by a factor of 0.1 at epochs 20, 40 and 60, or if the validation error did not improve for more than 10 epochs. Before lowering the learning rate, the weights with lowest validation error were restored, in order to avoid getting stuck in local optima. For the training data set including only 10 subjects used in section 2.4 ‘Model accuracy and network scaling’, the number of training epochs was increased to 140 and the learning rate decay applied at epochs 40, 80 and 120. The influence of the DCRNN model hyperparameters are discussed in supplement 1 (figure S1) and were chosen to yield a reasonable trade-off between accuracy and computational requirements. The encoder and decoder of the sequence-to-sequence architecture consist to two diffusion convolution GRU layers each, and the hidden state size was set to 64. The computations were performed on a Nvidia RTX 2080 Ti GPU, running on a desktop PC with an Intel(R) Core(TM) i7-9800X CPU under Ubuntu 20.04. With this setup one epoch on the dataset including 25 subjects and predicting the activity within one hemisphere including 180 ROIs took approximately 3.4 minutes.

**GWN** The GWN model was also trained incorporating the Adam optimizer [42] to minimize the MAE defined in equation 19. For GWN model it was sufficient to train it 30 epochs with a batch size of 8, thereby initializing the learning rate with \( \eta = 0.0001 \) and decreasing it by a factor of 0.1 at epochs 10 and 20. For the 10 subject dataset the number of epochs was also increased to 60 and the learning rate decay adapted to epochs 20 and 40 correspondingly. The influence of the
hyperparameters of the GWN is evaluated in supplement I (figure S2). A good trade-off between model accuracy and complexity could be found using 32 neurons. The number of layers per block were defined as 2 with a total number of 12 blocks. With this setup one epoch on the 25 subjects’ dataset including 180 ROIs took around 12.2 minutes.

**TAtt** The TAtt model was trained using the Adam optimizer [42] for in total 40 epochs, minimizing the MAE defined in equation 19, using a batch size of 16. The learning rate was initialized with η = 0.1 and decreased by a factor of 0.1 at epochs 10, 20 and 30. The influence of the hyperparameters is evaluated in supplement I (figure S3). The number of neurons in the temporal attentions were set to 32 thereby using 4 attention heads in the 4 TAtt layers. With this setup of hyperparameters one epoch of the TAtt model took around 7.7 minutes.

### 4.6 Vector auto regressive model

Granger causality [30] is currently most often based on linear vector autoregressive (VAR) models for stochastic time series data, and therefore we compare our GNN based approaches with a VAR model, as implemented in the multivariate Granger causality (MVGC) toolbox [8]. An autoregressive process (AR) is based on the idea that a signal $x^{(t)}$ can be described as a linear superposition of the first $T_p$ of its lagged values [50]:

$$x^{(t)} = \beta + \alpha_1 x^{(t-1)} + \alpha_2 x^{(t-2)} + \cdots + \alpha_p x^{(t-T_p)} + u^{(t)}$$  \hspace{1cm} (21)

with coefficients or weights $\alpha_1, \ldots, \alpha_p$, an intercept $\beta$ and the error term $u^{(t)}$. This univariate formulation can be extended to a multivariate VAR model including $N$ time series $x^{(t)} = [x^{(t)}_1, \ldots, x^{(t)}_N]$ like [50]:

$$\mathbf{x}^{(t)} = \mathbf{b} + \mathbf{A}_1 \mathbf{x}^{(t-1)} + \mathbf{A}_2 \mathbf{x}^{(t-2)} + \cdots + \mathbf{A}_p \mathbf{x}^{(t-T_p)} + \mathbf{u}^{(t)}$$  \hspace{1cm} (22)

whereby the coefficients are now collected in matrices $\mathbf{A} \in \mathbb{R}^{N \times N}$, and intercepts and errors are characterized by vectors $\mathbf{b} \in \mathbb{R}^N$ and $\mathbf{u}^{(t)} \in \mathbb{R}^N$. In our study the multivariate time series $\mathbf{x}^{(t)}$ reflect the BOLD signal strength in the $N$ brain regions, sampled at different times $t$.

To estimate parameters of the VAR model we used the ordinary least squares (OLS) fit provided in the MVGC toolbox [8]. As outlined in section 2.2 we used the first 80% of the data from each fMRI session to fit the model. Then for the comparison to the GNN approaches in section 2.4 we tested the model order $p$ in steps of 5 with $p = 5, 10, \ldots, T_p$ and chose the model with highest accuracy on each individual dataset used in section 2.4. To check for stationarity of the signals, an augmented Dickey-Fuller test for unit roots was applied to the BOLD timecourses [33, 51], using a p-value of $p < 0.01$. For the 25 subject dataset, around 10.0% of the BOLD time courses do not fulfill the stationarity criteria of the augmented Dickey-Fuller test ($p > 0.01$) when using such a high lag order of $T_p = 60$. But as the objective criterion of the evaluation in section 2.4 was to assess the capabilities of the models to predict empirically observed neural activity patterns, we chose the VAR model with best prediction accuracy for comparisons with the GNNs.
4.7 Dataset

The MRI data set used in our study is provided by the HCP data repository [35, 76]. We incorporated data of the S1200 release, which provides data from resting state fMRI sessions, each with a duration of 14.4 minutes, whereby 1200 volumes were sampled per session. The data was acquired with customized Siemens Connectome Skyra magnetic resonance imaging scanners with a field strength of $B_0 = 3T$, using multi-band (factor 8) acceleration [55, 24, 62, 83]. A gradient-echo echo-planar imaging (EPI) sequences with a repetition time $TR = 720 \text{ ms}$ and an echo time $TE = 31.1 \text{ ms}$ was used. The field of view of the fMRI sequence was $FOV = 208 \text{ mm} \times 180 \text{ mm}$ and in total $N_s = 72$ slices with a slice thickness of $d_s = 2 \text{ mm}$ were collected, containing voxels with an isotropic size of $2 \text{ mm}$. The fully preprocessed version, which includes motion-correction, structural preprocessing and ICA-FIX denoising was selected [28, 38, 39, 25, 66, 59, 31].

To define our brain network, the multimodal parcellation proposed by Glasser et al. [27] was applied, which divides the cortical surface into 180 segregated areas. The BOLD signal within each area was averaged, to obtain the temporal activity evolution for each node in our brain network. For this study we considered it useful to include global signal regression in our preprocessing, firstly because it showed to effectively reduce movement artifacts in HCP datasets [16]. Furthermore in our study of functional interactions between specific brain regions, the goal was to extract the additional information, which certain regions contain about the activity in other regions, so that local interactions rather than global modulations in the signal were of main interest for us. The BOLD signal time courses were the bandpass filtered by employing a filter with cutoff frequencies $0.04 - 0.07 Hz$ [29, 14, 12, 2].

In the S1200 release, diffusion MRI data was collected in 6 runs, whereby approximately 90 directions were sampled during each run, using three shells of $b = 1000, 2000, \text{ and } 3000 \text{ s/mm}^2$, with additionally 6 $b = 0$ images [67]. A spin-echo EPI sequence was incorporated with repetition time $TR = 5520 \text{ ms}$, echo time $TE = 89.5 \text{ ms}$, using a multi band factor of 3. In total $N_s = 111$ slices were collected, with field of view $FOV = 210 \text{ mm} \times 180 \text{ mm}$ and an isotropic voxel size of $1.25 \text{ mm}$. The HCP diffusion MRI preprocessing included intensity normalization across runs, EPI distortion correction, eddy-current corrections, removing motion artifacts, and gradient non-linearity corrections [28, 68, 4, 6, 5]. To reconstruct the anatomical connection strengths between regions within the multimodal parcellation [27], the MRtrix3 software package was incorporated [73]. Multi-shell multi-tissue constrained spherical deconvolution [40] was applied to obtain response functions for fiber orientation distribution estimation [72, 71]. Then 10 million streamlines were created using anatomical constrained tractography [64]. Finally spherical-deconvolution informed filtering was used [65], reducing the number of streamlines to 1 million. The strength of SC was defined as the number of streamlines connecting two brain regions, normalized by the region volumes. The group structural connectivity matrix $A_{SC}$ was obtained as the average SC across the first 10 subjects, because the variance in the SC strength was relatively low across subjects [88], while probabilistic tractography methods are computationally demanding. For the HCP dataset, including only young healthy subjects, the similarity of the SC across subjects was quite high, and the Pearson correlation coefficient between SC values of the 10 subjects was on average 0.91. But when comparing very different subject cohorts, like healthy and diseased subjects, the anatomical connectivity can differ considerably between those cohorts, and the SC matrix should then be computed for every studied group individually.
Data availability

Preprocessed HCP data is publicly available under: https://db.humanconnectome.org.
A demo version for MRI data preparation and training the DCRNN model is provided under: https://github.com/simonvino/DCRNN_brain_connectivity.
In addition, the demo version for the GWN model is provided under: https://github.com/simonvino/GraphWaveNet_brain_connectivity.

Acknowledgements

The work was supported by the DFG projects GR988/25-1 and ISNT89/393-1 granted to M.W.G.. Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

References

[1] F. Abdelnour, M. Dayan, O. Devinsky, T. Thesen, and A. Raj. Functional brain connectivity is predictable from anatomic network’s laplacian eigen-structure. NeuroImage, 172:728–739, 2018.

[2] S. Achard, R. Salvador, B. Whitcher, J. Suckling, and E. Bullmore. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. J Neurosci, 3:e17, 2006.

[3] E. Amico and J. Goñi. Mapping hybrid functional-structural connectivity traits in the human connectome. Network Neuroscience, 2:306–322, 2018.

[4] J. Andersson, S. Skare, and J. Ashburner. How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. NeuroImage, 20:870–88, 2003.

[5] J. Andersson and S. Sotiropoulos. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. NeuroImage, 125:1063–1078, 2015.

[6] J. Andersson and S. Sotiropoulos. Non-parametric representation and prediction of single- and multi-shell diffusion-weighted MRI data using gaussian processes. NeuroImage, 122:166–76, 2015.

[7] S. Arslan, S. I. Ktena, B. Glocker, and D. Rueckert. Graph saliency maps through spectral convolutional networks: Application to sex classification with brain connectivity. In GRAIL/Beyond-MIC@MICCAI, 2018.
[8] L. Barnett and A. Seth. The MVGC multivariate granger causality toolbox: A new approach to granger-causal inference. *Journal of neuroscience methods*, 223:50–68, 2013.

[9] C. Becker, S. Pequito, G. Pappas, M. Miller, S. T. Grafton, D. S. Bassett, and V. Preciado. Spectral mapping of brain functional connectivity from diffusion imaging. *Scientific Reports*, 8, 12 2018.

[10] S. Bengio, O. Vinyals, N. Jaitly, and N. Shazeer. Scheduled sampling for sequence prediction with recurrent neural networks. In *NIPS*, pages 1171–1179, 2015.

[11] R. G. Bettinardi, G. Deco, V. M. Karlaftis, T. J. V. Hartevelt, H. M. Fernandes, Z. Kourtzi, M. L. Kringelbach, and G. Zamora-López. How structure sculpts function: unveiling the contribution of anatomical connectivity to the brain’s spontaneous correlation structure. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27, 2018.

[12] B. B. Biswal, F. Z. Yetkin, V. Haughton, and J. S. Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic resonance in medicine*, 34 4:537–41, 1995.

[13] M. M. Bronstein, J. Bruna, Y. LeCun, A. D. Szlam, and P. Vandergheynst. Geometric deep learning: Going beyond euclidean data. *IEEE Signal Processing Magazine*, 34:18–42, 2017.

[14] R. Buckner, J. Sepulcre, T. Talukdar, F. Krienen, H. Liu, T. Hedden, J. Andrews-Hanna, R. Sperling, and K. Johnson. Cortical hubs revealed by intrinsic functional connectivity: Mapping, assessment of stability, and relation to alzheimer’s disease. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 29:1860–73, 2009.

[15] E. T. Bullmore and D. S. Bassett. Brain graphs: graphical models of the human brain connectome. *Annu. Rev. Clin. Psychol.*, 7:113–140, 2011.

[16] G. Burgess, S. Kandala, D. Nolan, T. Laumann, J. Power, B. Adeyemo, M. Harms, S. Petersen, and D. Barch. Evaluation of denoising strategies to address motion-correlated artifact in resting state fMRI data from the human connectome project. *Brain Connectivity*, 6, 2016.

[17] X. Chen and Y. Wang. Predicting Resting-state Functional Connectivity with Efficient Structural Connectivity. *EEE/CAA Journal of Automatica Sinica*, 5(6):1079–1088, 2018.

[18] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling, 2014.

[19] P. de Haan, T. S. Cohen, and M. Welling. Natural graph networks. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 3636–3646. Curran Associates, Inc., 2020.

[20] G. Deco, M. L. Kringelbach, V. K. Jirsa, and P. Ritter. The dynamics of resting fluctuations in the brain: metastability and its dynamical cortical core. *Scientific Reports*, 7, 2017.
[21] G. Deco, M. Senden, and V. Jirsa. How anatomy shapes dynamics: a semi-analytical study of the brain at rest by a simple spin model. *Frontiers in computational neuroscience*, 6:68, 2012.

[22] M. Defferrard, X. Bresson, and P. Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *NIPS*, pages 3837–3845, 2016.

[23] F. Deligianni, D. Carmichael, G. H Zhang, C. Clark, and J. Clayden. Noddi and tensor-based microstructural indices as predictors of functional connectivity. *PloS one*, 11:e0153404, 04 2016.

[24] D. Feinberg, S. Moeller, S. M Smith, E. Auerbach, S. Ramanna, M. Günther, M. F Glasser, K. Miller, K. Ugurbil, and E. Yacoub. Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. *PloS one*, 5:e15710, 2010.

[25] B. Fischl. Freesurfer. *NeuroImage*, 62(2):774 – 781, 2012.

[26] K. Friston, R. Moran, and A. K. Seth. Analysing connectivity with granger causality and dynamic causal modelling. *Curr Opin Neurobiol*, 23:172–178, 2013.

[27] M. Glasser, T. Coalson, E. Robinson, C. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Andersson, C. Beckmann, M. Jenkinson, S. Smith, and D. Van Essen. A multi-modal parcellation of human cerebral cortex. *Nature*, 536, 2016.

[28] M. Glasser, S. Sotiropoulos, J. Wilson, T. Coalson, B. Fischl, J. Andersson, J. Xu, S. Jbabdi, M. Webster, J. Polimeni, V. DC, and M. Jenkinson. The minimal preprocessing pipelines for the human connectome project. *NeuroImage*, 80, 2013.

[29] E. Glerean, J. Salmi, J. Lahnakoski, I. Jääskeläinen, and M. Sams. Functional magnetic resonance imaging phase synchronization as a measure of dynamic functional connectivity. *Brain connectivity*, 2:91–101, 2012.

[30] C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:424–438, 1969.

[31] L. Griffanti, G. Salimi-Khorshidi, C. F. Beckmann, E. J. Auerbach, G. Douaud, C. E. Sexton, E. Zsoldos, K. P. Ebmeier, N. Filippini, C. E. Mackay, S. Moeller, J. Xu, E. Yacoub, G. Baselli, K. Ugurbil, K. L. Miller, and S. M. Smith. ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *NeuroImage*, 95:232 – 247, 2014.

[32] A. Grover and J. Leskovec. node2vec: Scalable feature learning for networks. volume 2016, pages 855–864, 07 2016.

[33] J. Hamilton. *Time Series Analysis*. Princeton University Press, Princeton, NJ., 1994.

[34] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.
M. Hodge, W. Horton, T. Brown, R. Herrick, T. Olsen, M. Hileman, M. McKay, K. Archie, E. Cler, M. Harms, G. Burgess, M. Glasser, J. Elam, S. Curtiss, D. Barch, R. Oostenveld, L. Larson-Prior, K. Ugurbil, D. Van Essen, and D. Marcus. ConnectomeDB – sharing human brain connectivity data. *NeuroImage*, 124, 2015.

C. J. Honey, O. Sporns, L. Cammoun, X. Gigandet, J. P. Thiran, R. Meuli, and P. Hagmann. Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America*, 106 6:2035–40, 2009.

S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In F. Bach and D. Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 448–456, Lille, France, 07–09 Jul 2015. PMLR.

M. Jenkinson, P. Bannister, M. Brady, and S. Smith. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17:825 – 841, 2002.

M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, and S. M. Smith. FSL. *NeuroImage*, 62(2):782 – 790, 2012.

B. Jeurissen, J.-D. Tournier, T. Dhollander, A. Connelly, and J. Sijbers. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. *NeuroImage*, 103:411–426, 2014.

B.-H. Kim and J. C. Ye. Understanding graph isomorphism network for rs-fMRI functional connectivity analysis. *Frontiers in Neuroscience*, 14:630, 2020.

D. Kingma and J. Ba. Adam: A method for stochastic optimization. 2014.

S. I. Ktena, S. Parisot, E. Ferrante, M. Rajchl, M. C. H. Lee, B. Glocker, and D. Rueckert. Metric learning with spectral graph convolutions on brain connectivity networks. *NeuroImage*, 169:431–442, 2018.

E. Lang, A. Tomé, I. Keck, J. Gorriz, and C. Puntonet. Brain connectivity analysis: A short survey. *Computational intelligence and neuroscience*, 2012.

X. Li, N. Dvornek, Y. Zhou, J. Zhuang, P. Ventola, and J. Duncan. Graph neural network for interpreting task-fMRI biomarkers. pages 485–493, 2019.

Y. Li, R. Yu, C. Shahabi, and Y. Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting, 2018.

H. Liang and H. Wang. Structure-Function Network Mapping and its Assessment via Persistent Homology. *PLoS Computational Biology*, 2017.
[48] J. Liang, L. Jiang, L. Cao, Y. Kalantidis, L.-J. Li, and A. G. Hauptmann. Focal visual-text attention for memex question answering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(8):1893–1908, 2019.

[49] S. Lim, F. Radicchi, M. P. van den Heuvel, and O. Sporns. Discordant attributes of structural and functional brain connectivity in a two-layer multiplex network. *Scientific Reports*, 9, 12 2019.

[50] H. Luetkepohl. *The New Introduction to Multiple Time Series Analysis*. Springer, 2005.

[51] J. Mackinnon. Approximate asymptotic distribution functions for unit-root and cointegration tests. *Journal of Business and Economic Statistics*, 12:167–76, 1994.

[52] A. Messé, M. T. Hütt, P. König, and C. C. Hilgetag. A closer look at the apparent correlation of structural and functional connectivity in excitable neural networks. *Scientific Reports*, 5:7870, 2015.

[53] A. Messé, D. Rudrauf, H. Benali, and G. Marrelec. Relating Structure and Function in the Human Brain: Relative Contributions of Anatomy, Stationary Dynamics, and Non-stationarities. *PLoS Computational Biology*, 10(3), 2014.

[54] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 26, 10 2013.

[55] S. Moeller, E. Yacoub, C. A. Olman, E. Auerbach, J. Strupp, N. Y. Harel, and K. Ugurbil. Multiband multislice ge-epi at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magnetic resonance in medicine*, 63 5:1144–53, 2010.

[56] C. A. Olman, L. Davachi, and S. J. Inati. Distortion and signal loss in medial temporal lobe. *PLoS ONE*, 4, 2009.

[57] G. Rosenthal, F. Váša, A. Griffa, P. Hagmann, E. Amico, J. Goñi, G. Avidan, and O. Sporns. Mapping higher-order relations between brain structure and function with embedded vector representations of connectomes. *Nature Communications*, 9, 12 2018.

[58] D. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986.

[59] G. Salimi-Khorshidi, G. Douaud, C. F. Beckmann, M. F. Glasser, L. Griffanti, and S. M. Smith. Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage*, 90:449 – 468, 2014.

[60] T. Schnake, O. Eberle, J. Lederer, S. Nakajima, K. T. Schutt, K.-R. Muller, and G. Montavon. Higher-order explanations of graph neural networks via relevant walks. *IEEE transactions on pattern analysis and machine intelligence*, PP, 2021.
[61] Y. Seo, M. Defferrard, P. Vandergheynst, and X. Bresson. Structured sequence modeling with graph convolutional recurrent networks. In L. Cheng, A. C. S. Leung, and S. Ozawa, editors, Neural Information Processing, pages 362–373. Springer International Publishing, 2018.

[62] K. Setsompop, B. Gagoski, J. R. Polimeni, T. Witzel, V. J. Wedeen, and L. L. Wald. Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. Magnetic resonance in medicine, 67 5:1210–24, 2012.

[63] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst. The emerging field of signal processing on graphs: Ex-tending high-dimensional data analysis to networks and other irregular domains. IEEE Signal Processing Magazine, 30(3):83–98, 2013.

[64] R. Smith, J.-D. Tournier, F. Calamante, and A. Connelly. Anatomically-constrained tractography: Improved diffusion mri streamlines tractography through effective use of anatomical information. NeuroImage, 62:1924–38, 2012.

[65] R. Smith, J.-D. Tournier, F. Calamante, and A. Connelly. Sift: Spherical-deconvolution informed filtering of tractograms. NeuroImage, 67:298–312, 2013.

[66] S. M. Smith, C. F. Beckmann, J. Andersson, E. J. Auerbach, J. Bijsterbosch, G. Douaud, E. Duff, D. A. Feinberg, L. Griffanti, M. P. Harms, M. Kelly, T. Laumann, K. L. Miller, S. Moeller, S. Petersen, J. Power, G. Salimi-Khorshidi, A. Z. Snyder, A. T. Vu, M. W. Woolrich, J. Xu, E. Yacoub, K. Uğurbil, D. C. V. Essen, and M. F. Glasser. Resting-state fMRI in the human connectome project. NeuroImage, 80:144 – 168, 2013.

[67] S. Sotiropoulos, S. Jbabdi, J. Xu, J. Andersson, S. Moeller, E. Auerbach, M. Glasser, M. Hernandez Fernandez, G. Sapiro, M. Jenkinson, D. Feinberg, E. Yacoub, C. Lenglet, V. DC, K. Uğurbil, and T. Behrens. Advances in diffusion mri acquisition and processing in the human connectome project. NeuroImage, 80:125, 2013.

[68] S. Sotiropoulos, S. Moeller, S. Jbabdi, J. Xu, J. Andersson, E. Auerbach, E. Yacoub, D. Feinberg, K. Setsompop, L. Wald, T. Behrens, K. Ugurbil, and C. Lenglet. Effects of image reconstruction on fibre orientation mapping from multi-channel diffusion MRI: Reducing the noise floor using SENSE. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine, (70), 2013.

[69] I. Sutskever, O. Vinyals, and Q. V. Le. Sequence to sequence learning with neural networks. CoRR, abs/1409.3215, 2014.

[70] C. Thomas, F. Q. Ye, M. O. Irfanoglu, P. D. Modi, K. S. Saleem, D. A. Leopold, and C. Pierpaoli. Anatomical accuracy of brain connections derived from diffusion MRI tractography is inherently limited. Proceedings of the National Academy of Sciences of the United States of America, 111 46:16574–9, 2014.

[71] J.-D. Tournier, F. Calamante, and A. Connelly. Robust determination of the fibre orientation distribution in diffusion mri: Non-negativity constrained super-resolved spherical deconvolution. NeuroImage, 35:1459–72, 2007.
[72] J.-D. Tournier, F. Calamante, D. Gadian, and A. Connelly. Direct estimation of the fiber orientation density function from diffusion-weighted mri data using spherical deconvolution. *NeuroImage*, 23:1176–85, 2004.

[73] J.-D. Tournier, R. Smith, D. Raffelt, R. Tabbara, T. Dholander, M. Pietsch, D. Christiaens, B. Jeurissen, C.-H. Yeh, and A. Connelly. MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *NeuroImage*, 202, 2019.

[74] K. Uğurbil, J. Xu, E. Auerbach, S. Moeller, A. Vu, J. Duarte-Carvajalino, C. Lenglet, X. Wu, S. Schmitter, P.-F. Van de Moortele, J. Strupp, G. Sapiro, F. De Martino, D. Wang, N. Harel, M. Garwood, L. Chen, D. Feinberg, S. Smith, and E. Yacoub. Pushing spatial and temporal resolution for functional and diffusion MRI in the human connectome project. *NeuroImage*, 80, 2013.

[75] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu. Wavenet: A generative model for raw audio. In *Arxiv*, 2016.

[76] D. Van Essen, S. Smith, D. Barch, T. Behrens, E. Yacoub, and K. Ugurbil. The wu-minn human connectome project: an overview. *NeuroImage*, 80, 2013.

[77] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[78] S. Wein, G. Deco, A. Tomé, M. Goldhacker, W. Malloni, M. Greenlee, and E. Lang. Brain connectivity studies on structure-function relationships: A short survey with an emphasis on machine learning. *Computational Intelligence and Neuroscience*, 2021:1–31, 05 2021.

[79] S. Wein, W. Malloni, A. Tomé, S. Frank, G.-I. Henze, S. Wüst, M. Greenlee, and E. Lang. A graph neural network framework for causal inference in brain networks. *Scientific Reports*, 11, 04 2021.

[80] P. Werbos. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 78(10):1550–1560, 1990.

[81] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–21, 2020.

[82] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang. Graph wavenet for deep spatial-temporal graph modeling. pages 1907–1913, 08 2019.

[83] J. Xu, S. Moeller, J. Strupp, E. Auerbach, L. Chen, D. A. Feinberg, K. Ugurbil, and E. Yacoub. Highly accelerated whole brain imaging using aligned-blipped-controlled-aliasing multiband EPI. *Proceedings of the 20th Annual Meeting of ISMRM*, page 2036, 2012.
[84] K. Xu, J. Ba, R. Kiros, K. Cho, A. C. Courville, R. Salakhutdinov, R. S. Zemel, and Y. Bengio. Show, attend and tell: Neural image caption generation with visual attention. In ICML, 2015.

[85] M. Zeiler and R. Fergus. Visualizing and understanding convolutional neural networks. ECCV 2014, Part I, LNCS 8689, 8689, 2013.

[86] H. Zhang, T. Schneider, C. Gandini Wheeler-Kingshott, and D. Alexander. NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain. NeuroImage, 61:1000–16, 2012.

[87] C. Zheng, X. Fan, C. Wang, and J. Qi. Gman: A graph multi-attention network for traffic prediction. Proceedings of the AAAI Conference on Artificial Intelligence, 34:1234–1241, 04 2020.

[88] J. Zimmermann, J. Griffiths, M. Schirner, P. Ritter, and A. R. McIntosh. Subject-specificity of the correlation between large-scale structural and functional connectivity. Network Neuroscience, pages 1–35, 2019.

[89] R. Řehůrek and P. Sojka. Software framework for topic modelling with large corpora. pages 45–50, 05 2010.
Supplementary Material

Supplement I

In this supplement the influence of the model hyperparameters for the different neural network architectures is discussed. The hyperparameters are chosen like described in section 4.5 ‘Model training’ and held constant, while only the hyperparameter of interest is varied in the following evaluations. Figure S1 and S2 show that the performance of the DCRNN and GWN in general still could slightly improve with a larger number of model parameters. However as the computation time and memory requirements linearly grow with the number of parameters, we chose the model hyperparameters as described in 4.5 to yield a good trade-off between model performance and computational requirements. Also the TAtt model in figure S3 shows some improvement with a larger number of parameters, however the MAE is still considerably higher compared to the RNN and WN based GNN architectures.

Figure S1: Here the influence of the hyperparameters on the prediction accuracy of the DCRNN is depicted. In (A) the test error is shown in dependence of the number of neurons in each layer, and in (B) the error in dependence of the number of DCGRU layers.
Figure S2: In this figure the influence of the GWN hyperparameters on the prediction accuracy is shown. In (A) the test error in dependence of number of neurons (or feature maps) is illustrated. Here (B) shows the influence of the number of DCC blocks used in the GWN architecture and (C) shows the impact of the number of layers per DCC block.

Figure S3: Here the influence of hyperparameters on the TAtt accuracy is illustrated. In (A) the test MAE in dependence of the number of neurons used in each TAtt mechanism is shown. In (B) the influence of the number of TAtt layers is depicted and (C) illustrates the impact of number of attention heads incorporated.
Table 1: List of ROIs involved in visual processing based on the multimodal parcellation proposed by Glasser et al. [27]. The table shows the index of the region in the atlas for the right/left hemisphere including the name of the region.

| Index | Name   |
|-------|--------|
| 1/181 | V1     |
| 2/182 | MST    |
| 3/183 | V6     |
| 4/184 | V2     |
| 5/185 | V3     |
| 6/186 | V4     |
| 7/187 | V8     |
| 13/193| V3A    |
| 16/196| V7     |
| 19/199| V3B    |
| 20/200| LO1    |
| 21/201| LO2    |
| 22/202| PIT    |
| 23/203| MT     |
| 152/332| V6A   |
| 153/333| VMV1  |
| 154/334| VMV3  |
| 156/336| V4t   |
| 158/338| V3CD  |
| 159/339| LO3   |
| 160/340| VMV2  |
| 163/343| VVC   |
Figure S4: This figure illustrates directed spatial relations learned by the DCRNN model. Figure (A) shows the measures of influence $I(n')$, derived from the DCRNN model when using the SC for information propagation and figure (B) depicts the influence when incorporating CEs for the information exchange. The values of the connectivity measures were linearly mapped between 0 and 100 and the default scaling of the color values provided by the connectome workbench (version 1.4.2) was used, adjusting the colormap between the 2th and 98th percentile of the values respectively.