NETWORK NEUROSCIENCE

an open access journal

Citation: Wein, S., Schüller, A., Tomé, A. M., Malloni, W. M., Greenlee, M. W., & Lang, E. W. (2022). Forecasting brain activity based on models of spatiotemporal brain dynamics: A comparison of graph neural network architectures. *Network Neuroscience*, *6*(3), 665–701. https://doi.org/10.1162/netn_a_00252

DOI:
https://doi.org/10.1162/netn_a_00252

Supporting Information:
https://doi.org/10.1162/netn_a_00252

Received: 10 December 2021
Accepted: 2 May 2022

Competing Interests: The authors have declared that no competing interests exist.

Corresponding Author:
S. Wein
simon.wein@ur.de

Handling Editor:
Bratislav Misic

The MIT Press

METHODS

Forecasting brain activity based on models of spatiotemporal brain dynamics: A comparison of graph neural network architectures

S. Wein[1,2], A. Schüller[1], A. M. Tomé[3], W. M. Malloni[2], M. W. Greenlee[2], and E. W. Lang[1]

[1]CIML, Biophysics, University of Regensburg, Regensburg, Germany
[2]Experimental Psychology, University of Regensburg, Regensburg, Germany
[3]IEETA, DETI, Universidade de Aveiro, Aveiro, Portugal

Keywords: Brain connectivity, Graph neural networks, Structure-function relationship, Directed connectivity

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 spatiotemporal GNN architectures and study their ability to model 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 often 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 spatiotemporal dynamics in brain networks.

AUTHOR SUMMARY

In our study we compare different spatial and temporal modeling techniques based on graph neural networks (GNNs) for investigating the spatiotemporal dynamics in brain networks. We show that a convolutional neural network and a recurrent neural network–based approach are both very suitable to capture the temporal characteristics in functional activity distributions. Further, we demonstrate that structural connectome embeddings can effectively reduce the number of parameters in GNN models, by naturally including higher order topological relations between brain areas within the structural network. We compare the prediction accuracy of the GNN-based approaches to a vector autoregressive model, and we show that GNNs remain considerably more accurate when brain networks become large and available data are limited. Finally, we demonstrate how these spatiotemporal GNN models can provide a multimodal perspective on directed connectivity in brain networks.

INTRODUCTION

Distinct concepts of brain connectivity can provide different but complementary aspects of information processing in the brain (Lang, Tomé, Keck, Górriz-Sáez, & Puntonet, 2012). 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 (Bielczyk et al., 2019; Friston, Moran, & Seth, 2013).

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 (Bronstein, Bruna, LeCun, Szlam, & Vandergheynst, 2017; Wu et al., 2021). 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 (Arslan, Ktena, Glocker, & Rueckert, 2018; Kim & Ye, 2020; Ktena et al., 2018; X. Li et al., 2019; Rosenthal et al., 2018; Wein, Malloni, et al., 2021). Given a decomposition of the brain into specified areas, their spatiotemporal 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 reflect the strength of interactions between neural populations (Bullmore & Bassett, 2011). As proposed in our recent study (Wein, Malloni, et al., 2021), these complex signals exhibited in non-Euclidean geometries can be processed with a variant of GNN denoted as spatiotemporal graph neural network (STGNN). Such STGNNs can allow us to simultaneously model spatial and temporal dependencies in such graph structured signals and thereby provide a new possibility to combine DTI with fMRI data. Activity propagation–based approaches made already various interesting contributions to brain connectivity research, and could, for example, explain how resting-state FC (Cole, Ito, Bassett, & Schultz, 2016) or SC (Yan et al., 2021) are related to cognitive task activations observed in task-based fMRI. Moreover, they could give us insights into what way the hierarchical organization of the brain is related to the propagation of information along structural connections (Vézquez-Rodríguez, Liu, Hagmann, & Misic, 2020). In our study, we compare different approaches based on GNNs to study the spatiotemporal propagation of information in brain networks from a multimodal and data-driven perspective.

Graph neural network:
A type of artificial neural network that is used to extract features from data with graph-like geometries.

Recently, several different GNN architectures have been proposed to model the flow of information across time and space in graphical signals (Wu et al., 2021). 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 (Defferrard, Bresson, & Vandergheynst, 2016). They were later combined with recurrent neural networks (RNNs) (Rumelhart, Hinton, & Williams, 1986), which can detect sequential relations in signals. This combined spatiotemporal GNN framework was proposed in the notion of diffusion convolution recurrent neural network (DCRNN) (Y. Li, Yu, Shahabi, & Liu, 2018). However, RNNs can have problems with long time series and, when combined with

Convolution:
Denotes in the context of artificial neural networks a discrete mathematical convolution operation of the input of a neural network layer with a localized filter kernel, whereby the filter kernels are learned during neural network training.

Deep learning:
Learning in an artificial neural network that includes multiple hidden representations of the neural network input.

Graph convolutions:
Denotes a discrete convolution operation generalized to the non-Euclidean geometry of graphs. Similar to classical convolution operations in CNNs, they entail principles like sparse interactions and parameter sharing.

Exploding gradients:
Describes the phenomenon when the gradients of the neural network model weights grow excessively during backpropagation learning.

Vanishing gradients:
Describes the phenomenon when the gradients of the neural network model weights become very small during backpropagation learning.

Receptive field:
Denotes the area in the artificial neural network input a neuron is (possibly via multiple layers) connected to.

Spatial information exchange:
Describes in the context of spatiotemporal graph signal processing the exchange of information along the edges of a graph signal.

graph convolution operations, their gradients are more likely to explode (exploding gradients) or vanish (vanishing gradients) (Y. Li et al., 2018; Seo, Defferrard, Vandergheynst, & Bresson, 2018). This was the motivation for introducing approaches that combine spatial graph convolutions with standard one-dimensional temporal convolutions (Wu et al., 2021). 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 (van den Oord et al., 2016). They have recently been combined with graph convolution operations in an architecture denoted as graph WaveNet (GWN) (Wu, Pan, Long, Jiang, & Zhang, 2019). As an alternative method for the temporal processing, also attention mechanisms have been recently included in STGNN architectures (Zheng, Fan, Wang, & Qi, 2020). 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 (J. Liang et al., 2019; Vaswani et al., 2017; K. Xu et al., 2015), but applications to analyze the 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 activity distributions observed in brain networks. In addition to these distinct 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 (CEs) of the anatomical network to characterize the node relations. In a recent study, Rosenthal et al. (2018) have shown that embeddings of nodes in the anatomical network can inherently capture higher order topological relations between different structurally connected nodes in this network. 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. We demonstrate that by modeling the functional information exchange between regions in STGNNs based on structural connectivity, we can significantly increase the accuracy of predicting future neural signals. This points out that STGNN models are capable of learning informative functional interactions between areas in such brain networks. Based on these different comparisons, we at first try to identify the most effective STGNN architectures to investigate such spatial and temporal dynamics in brain networks.

In a subsequent step we then compare this STGNN-based approach to a currently popular data-driven model for characterizing directed functional information exchange in brain networks. Until now, methods for the inference of directed functional or effective connectivity are often based on Granger causality (Barnett & Seth, 2013) or dynamic causal modeling (Friston et al., 2013) and its recent extensions (Frässle et al., 2018; Prando et al., 2020). In addition various algorithmic and information theory–based methods have been developed meanwhile in this field (Bielczyk et al., 2019; Ramsey, Hanson, & Glymour, 2011). In the following, we compare the performance of the STGNN prediction models to the forecasting model most often used in Granger causality analysis (Barnett & Seth, 2013). 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* (Friston et al., 2013). 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*. For practical applications of this idea, the underlying predictive model in Granger causality is usually based on a vector autoregression (VAR) for multivariate time series inference (Barnett & Seth, 2013; Bielczyk et al., 2019; Friston et al., 2013). 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 prediction accuracies of STGNN models remain significantly more accurate, even when brain networks become very complex and only limited data are available to fit the models. This points out that the STGNN approaches are robust among a large variety of MRI study scenarios, and are therefore also suitable for the analysis of smaller subject cohorts, like in studies of patients with rare neurological diseases.

Finally, we analyze 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. So far, a variety of approaches were proposed to study the structure-function relation in the human brain (Suárez, Markello, Betzel, & Misic, 2020), which are based on computational modeling (Chen & Wang, 2018; Deco et al., 2013; Deco, Senden, & Jirsa, 2012; Honey et al., 2009; Messé, Hütt, König, & Hilgetag, 2015; Messé, Rudrauf, Benali, & Marrelec, 2014), graph theory (Abdelnour, Dayan, Devinsky, Thesen, & Raj, 2018; Becker et al., 2018; H. Liang & Wang, 2017; Lim, Radicchi, van den Heuvel, & Sporns, 2019), and machine learning (Amico & Goñi, 2018; Deligianni, Carmichael, Zhang, Clark, & Clayden, 2016; Rosenthal et al., 2018; Sarwar, Tian, Yeo, Ramamohanarao, & Zalesky, 2021). Biophysically inspired models, for example, could describe how functional connectivity patterns emerge from neural dynamics with static couplings characterized by anatomical connections (Deco et al., 2013; Honey et al., 2009; Messé et al., 2014), and were recently used to also study spatial heterogeneity of local circuit properties across the cortex (Demirtaş et al., 2019; P. Wang et al., 2019). Methods from graph theory can supplement such computational frameworks by specifically pointing out how indirect structural connections contribute to the inference of FC strength (Becker et al., 2018; H. Liang & Wang, 2017). Also, hybrid methods could demonstrate that the typology of structural brain networks supports in neuromorphic networks their memory capacity (Suárez, Richards, Lajoie, & Misic, 2021). Such insights into structural and functional connectivity can then provide a basis to better understand the cognitive information processing in the human brain (Ito, Hearne, Mill, Cocuzza, & Cole, 2020). The vast majority of studies on structure-function relations focuses currently 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 (Wein, Deco, et al., 2021). To the contrary, 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. To study this relation further on the individual brain region level, we demonstrate how a perturbation-based approach can be utilized to reconstruct how dynamic functional interactions are mediated by their structural substrate in STGNN models. By inferring how information is propagated between individual regions in STGNNs, these predictive models have the potential to reveal directed relationships between individual areas in brain networks from a multimodal perspective. In general, due to the low temporal sampling rate

and physiological artifacts, fMRI can have several limitations in detecting directed relations in brain networks (S. M. Smith et al., 2011; Webb, Ferguson, Nielsen, & Anderson, 2013). Still, some recent fMRI studies and computational simulations could demonstrate that also lag-based methods like Granger causality can be useful for detecting such directed dependencies in fMRI data (Duggento et al., 2018; Mill, Bagic, Bostan, Schneider, & Cole, 2017; Seth, Chorley, & Barnett, 2012; H. E. Wang et al., 2014).

The possibility to combine structural and functional imaging data in STGNNs can make these models as well interesting for several practical applications in brain connectivity research. For instance, they can be used to investigate differences in the structure-function relationship between resting-state and task-based fMRI. Furthermore, in clinical applications these models could be employed to study how lesions in the structural connectome affect the functional organization of the brain network. In our current study we compare, therefore, such mechanisms for spatial and temporal modeling in STGNNs, with the objective to establish their methological foundation for brain connectivity research, and thereby providing a basis for future applications of STGNNs in multimodal neuroimaging studies.

RESULTS

Graph Neural Network Models

In our context of MRI, the goal of the spatiotemporal GNNs will be to forecast the observed BOLD signal as accurately as possible in order to precisely capture the spatiotemporal dynamics of the underlying mechanisms in the brain. The learning objective can be formalized by introducing a graph signal $\mathbf{x}^{(t)} \in \mathbb{R}^N$, representing the BOLD signal measured at time step 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 $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A}_w)$, containing vertices (nodes) $\mathcal{V}$, with $|\mathcal{V}| = N$, and edges $\mathcal{E}$. The structure of the graph is characterized by a weighted adjacency matrix $\mathbf{A}_w \in \mathbb{R}^{N \times N}$, where an entry $w_{nn'}$ describes the connection strength between brain region n and n'. An illustration of such a graphical representation of a dynamic brain state is provided in Figure 1. Based on this concept, the task of the

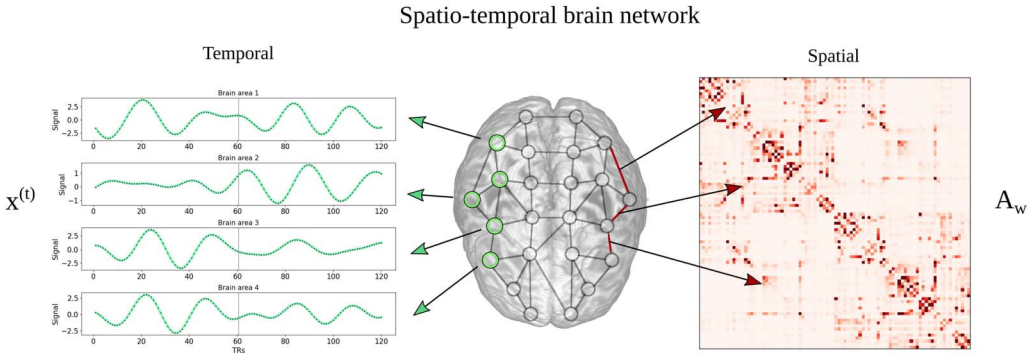

Figure 1. The spatiotemporal representation of a signal in a brain network is illustrated. The temporal component of the graph-like signal is given by the BOLD signal $\mathbf{x}^{(t)} \in \mathbb{R}^N$ in N brain regions sampled at different timesteps t, as shown on the left side. The strength of the edges in the brain network are defined by a weighted adjacency matrix $\mathbf{A}_w \in \mathbb{R}^{N \times N}$, as illustrated on the right side. One entry $w_{nn'}$ of the matrix $\mathbf{A}_w$ characterizes thereby the spatial relation between brain region n and n' in the network.

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:

$$\left[\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(T_p)}; \mathcal{G}\right] \overset{h(\cdot)}{\rightarrow} \left[\mathbf{x}^{(T_p+1)}, \ldots, \mathbf{x}^{(T_p+T_f)}\right] \tag{1}$$

Until now various spatiotemporal GNN architectures have been proposed to account for spatial and temporal dependencies of such graph structured signals (Wu et al., 2021). An overview of different possibilities for the temporal modeling is given in Figure 2A. In time series

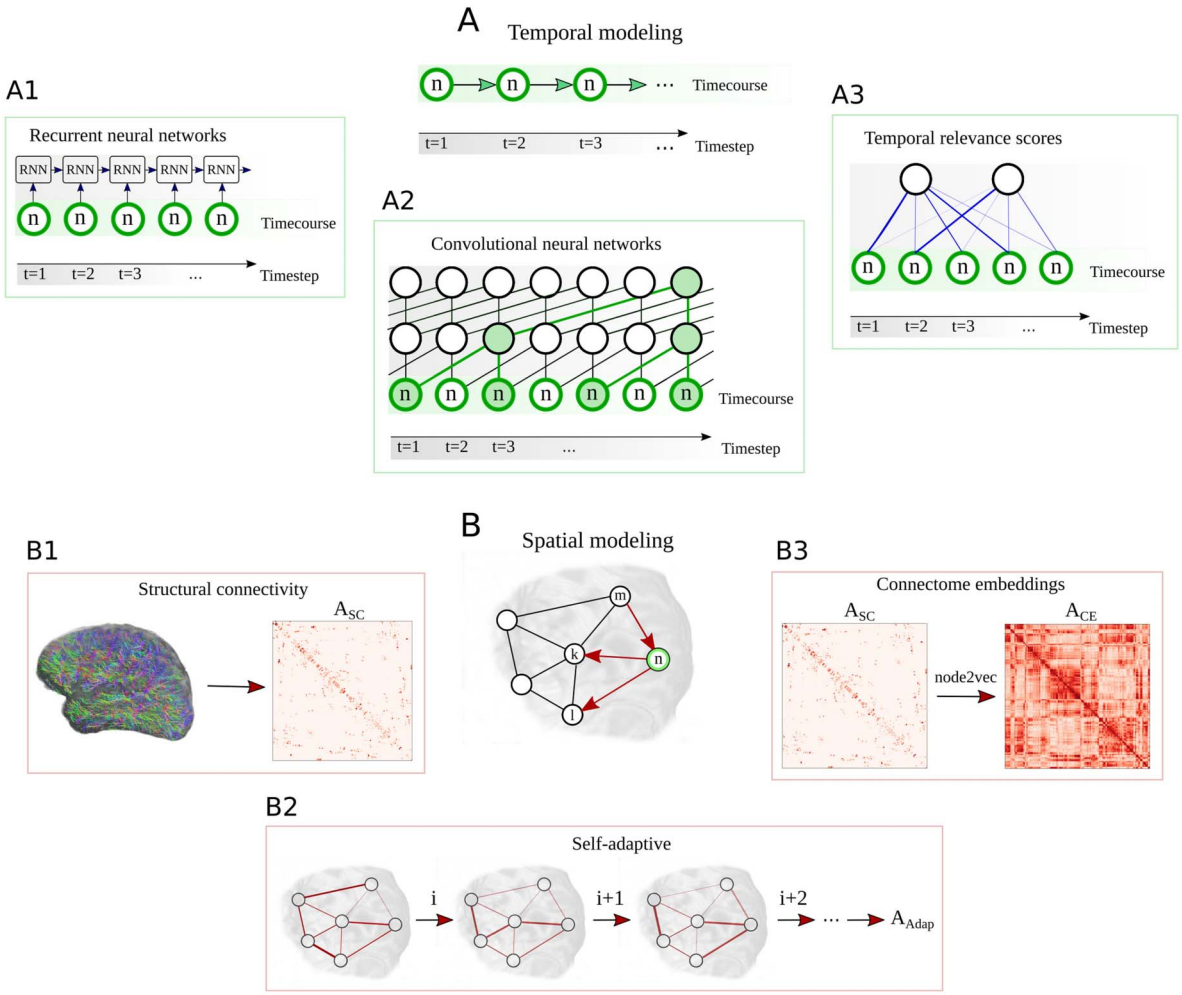

Figure 2. Overview on the different techniques used by spatiotemporal GNNs to learn the spatial and temporal dynamics in brain networks. The temporal component of the STGNN tries to infer dependencies between the activity at different timesteps $t = 1, 2, 3, \ldots$ in a certain brain region n, as illustrated in A. A recurrent neural network (RNN) architecture processes in a recursive manner the activity states at the different time steps t (A1). In our study, this principle will be implemented in an RNN-based sequence-to-sequence architecture for time series forecasting. As an alternative, convolutional neural network (CNN) architectures employ one-dimensional convolutions in the time domain (A2). This idea is picked up in the WaveNet (WN) architecture, which introduces dilated convolutions to obtain exponentially growing receptive fields, as highlighted here in green. Following the idea of an attention mechanism, temporal relevance scores can be dynamically computed to weight the importance of a temporal feature observed at time step t (A3). A temporal attention (TAtt) based architecture is thereby composed of a multitude of stacked attention layers. Based on these temporal modeling approaches, STGNNs additionally detect spatial dependencies between a node n and other interconnected nodes in the brain network, as illustrated in B. The substrate for the spatial interactions can be based on the structural connectivity matrix A_{SC}, as reconstructed from DTI data (B1). Alternatively, the node2vec algorithm can be used to detect higher order relationships between brain regions in the structural connectome, for characterizing the spatial similarity based on connectome embeddings A_{CE} (B3). Finally, spatial connections can be tried to be learned by the model itself, by freely adapting the edges in an adjacency matrix A_{Adap} at different iterations i, $i + 1$, $i + 1$ . . . during the model training (B2).

analysis, recurrent neural networks (RNNs) (Rumelhart et al., 1986) provide one efficient way to detect patterns in sequential data structures, like in our context the BOLD signal, subsequently sampled at different time steps t (Figure 2A1). This approach can be extended to a RNN-based sequence-to-sequence architecture, whereby an encoder recursively processes an input sequence of T_p past neural activity states $\mathbf{x}^{(t)}$ and encodes the temporal information into a hidden state $\mathbf{H}^{(T_p)}$ (Sutskever, Vinyals, & Le, 2014). Next, a decoder network uses the information in $\mathbf{H}^{(T_p)}$ 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 (Chung, Gulcehre, Cho, & Bengio, 2014). An alternative for detecting repetitive patterns in sequential data is provided by convolutional neural networks (CNNs) (Figure 2A2). By employing one-dimensional convolutions in the time domain, they are used in our context to process temporal dynamics of neural activity. To more efficiently capture long-term dependencies in temporal data the WaveNet (WN) architecture has been proposed (van den Oord et al., 2016). 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. The growth of the receptive field of a neuron (marked in green) in a WN layer is illustrated in Figure 2A2. More recently, also attention mechanisms have been proposed to detect underlying hidden correlations in sequential data structures (Vaswani et al., 2017). In time series analysis, the idea of a temporal attention (TAtt) architecture is thereby to adaptively focus on the most important temporal features in a sequence (Figure 2A3).

These different fundamental approaches for temporal dependency modeling have been recently combined with techniques to additionally capture spatial relationships in graph structured signals (Y. Li et al., 2018; Wu et al., 2019; Zheng et al., 2020). Graph convolutional neural networks can be incorporated to model the propagation of information between adjacent nodes in the graphical representation of the signal (Defferrard et al., 2016). The neighborhoods of the vertices/nodes $\mathcal{V}$ in the network are characterized by the adjacency matrix $\mathbf{A}_w$. In our study we additionally investigate different possibilities for defining the spatial layout for the information propagation between brain regions, as illustrated in Figure 2B. As a first choice for the adjacency matrix, we will employ the structural connectivity $\mathbf{A}_{SC}$ between the N brain areas, as it could be reconstructed from DTI data (Figure 2B1). 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. (2018) 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 can thereby allow us to additionally account for long-range and interhemispheric homotopic connections, which are usually only weakly expressed in DTI-based anatomical connectivity (Thomas et al., 2014). 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 $\mathbf{A}_{CE}$ (Figure 2B3). The information is accordingly propagated between brain regions that 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 $\mathbf{A}_{Adap} \in \mathbb{R}^{N \times N}$, which is learned during the model optimization (Figure 2B2). A detailed formal description of the model architectures and the training involved is outlined in the Materials and Methods section. In the following, we will assess the effectiveness of the different spatial and temporal modeling approaches by comparing their predictive performance on an MRI dataset from the Human Connectome Project (HCP) (Van Essen et al., 2013).

Data Description

For the different evaluations in this study, resting-state fMRI data provided by the HCP *S1200 release* was incorporated (S. M. Smith et al., 2013). To define the nodes of the brain network, the multimodal parcellation proposed by Glasser et al. (2016) 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, $T = 1{,}200$ fMRI images were collected, so that the ROI time series can be represented by a data matrix $\mathbf{X} \in \mathbb{R}^{N \times T}$. We filtered the resting-state fMRI time series data with a 0.04–0.07 Hz Butterworth band-pass filter, because this frequency band has shown to be most reliable and functionally relevant for gray matter activity (Achard, Salvador, Whitcher, Suckling, & Bullmore, 2006; B. B. Biswal, Yetkin, Haughton, & Hyde, 1995; Buckner et al., 2009; Deco, Kringelbach, Jirsa, & Ritter, 2017; Glerean, Salmi, Lahnakoski, Jääskeläinen, & Sams, 2012).

For learning the predictions of the BOLD signal, samples of input and output sequences were generated from the time series data in $\mathbf{X}$ (Wein, Malloni, et al., 2021). This was achieved by selecting windows of length T_p to obtain input sequences of neural activity states $[\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(T_p)}]$, and respective target sequences of length T_f denoted as $[\mathbf{x}^{(T_p+1)}, \ldots, \mathbf{x}^{(T_p+T_f)}]$. 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 each fMRI session, the first 80% of those time window samples were used as the training data for the models, the subsequent 10% as a validation set, and the last 10% were employed for testing. 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 $TR = 0.72$ s (Uğurbil et al., 2013). 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 time points is short enough for them to make reasonable nonrandom 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 multishell, multitissue constrained spherical deconvolution model (Jeurissen, Tournier, Dhollander, Connelly, & Sijbers, 2014), made available in the *MRtrix3* software package (Tournier et al., 2019). White matter tractography was performed to estimate the anatomical connection strength between the regions defined by the multimodal parcellation atlas (Glasser et al., 2016). The number of the streamlines that 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 $\mathbf{A}_{SC} \in \mathbb{R}^{N \times N}$. A detailed description of the MRI datasets and their preprocessing is provided in the Dataset section. In addition, the embeddings of the nodes within the structural network $\mathbf{A}_{SC} \in \mathbb{R}^{N \times N}$ were generated using the node2vec algorithm (Grover & Leskovec, 2016). The parameters for this algorithm are outlined in detail in the Spatial Dependencies section, 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 $\mathbf{A}_{CE} \in \mathbb{R}^{N \times N}$.

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, as discussed in the Data Description section, and the generated training, validation, and test samples were then aggregated across the 25 fMRI sessions. The neural signal in regions within the right hemisphere (Glasser et al., 2013), 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 the Model Training section and discussed in detail in Supporting Information I. The BOLD signal data was scaled to zero mean and unit variance for the evaluations, to obtain values of a magnitude that are easier to interpret. Figure 3A shows the test mean absolute error (MAE) between the predicted and the true neural activity. By generating windowed input-output pairs of activity values from the fMRI data, the last 10% of samples from each session correspond to 108 of such input-output pairs per session for testing, each containing 60 output time points (corresponding to roughly 43 s of activity). The overall test errors were

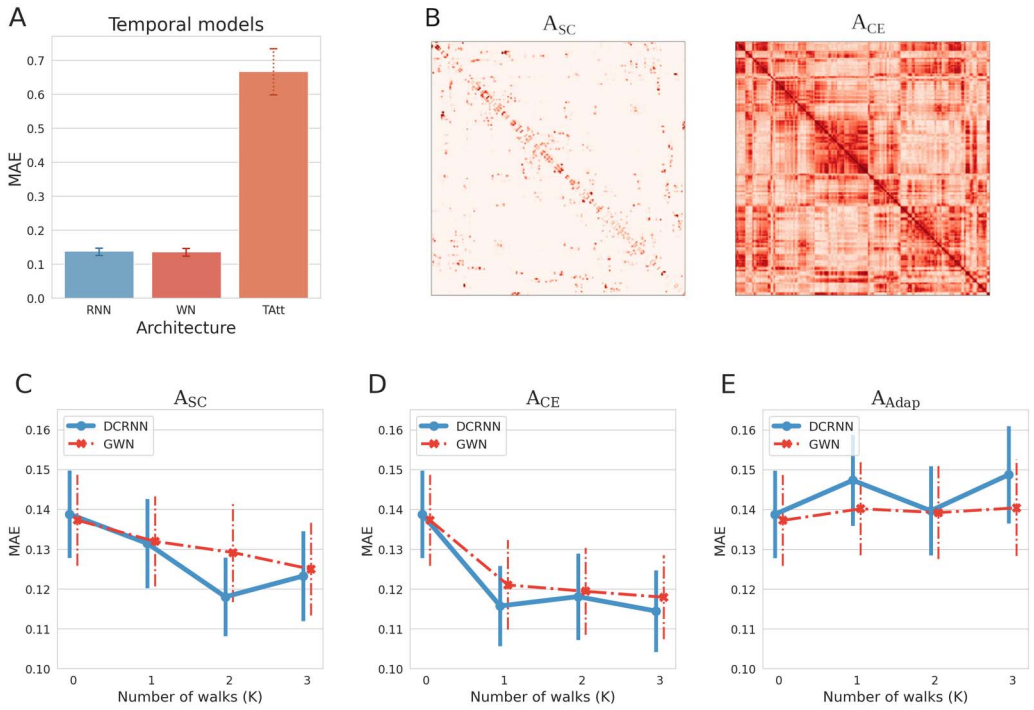

Figure 3. (A) A comparison of different modeling strategies for temporal dynamics in the BOLD signal, comparing the test MAE of the recurrent neural network (RNN), the WaveNet (WN), and the temporal attention (TAtt) architecture. The overall test error was computed as an average across samples, brain regions, and subject sessions. The error bars represent the standard deviation of the test MAE across subjects. Due to their high accuracy in the temporal domain, we focus on RNN- and WN-based approaches for forecasting the spatiotemporal dynamics in the following. Spatial relations are added to the temporal models in form of graph convolutions, and the spatiotemporal extension of the RNN and WN models are respectively denoted as diffusion convolution recurrent neural network (DCRNN) and graph WaveNet (GWN) (Y. Li et al., 2018; Wu et al., 2021). Spatial transitions are based on the relations of network nodes captured in a weighted adjacency matrix, which is either based on structural connectivity ($\mathbf{A}_{SC}$), connectome embedding similiarity ($\mathbf{A}_{CE}$), or adapted during model training ($\mathbf{A}_{Adap}$). In (B) the adjacency matrix $\mathbf{A}_{SC}$ based on structural connectivity within the 180 regions of the right hemisphere is illustrated, together with the adjacency matrix $\mathbf{A}_{CE}$ derived from structural connectome embedding similarities. The regions in this illustration are ordered according to the atlas proposed by Glasser et al. (2016). (C, D, and E) Prediction accuracies of the DCRNN and GWN model in dependence of the walk order K. In (C) the overall test MAE is shown when incorporating the SC as an adjacency matrix $\mathbf{A}_{SC}$, (D) illustrates the test MAE when employing CEs in an adjacency matrix $\mathbf{A}_{CE}$ to define spatial relationships, and (E) displays the case when using a self-adaptive weight matrix $\mathbf{A}_{Adap}$.

computed as the average across all these test samples from the 25 subjects and across all 180 brain regions. The comparison shows that RNN and WN have very similar capabilities in predicting the BOLD signal, while the TAtt model exhibits a worse performance. To test the significance of this difference between the models, we further computed the test MAE of each individual subject as an average across predicted time points, brain regions, and test samples per subject. By applying a paired t test, the differences between the WN and RNN model to the TAtt model were shown to be both highly significant with $p \leq 0.0001$ across subjects (Cohen's $d = 10.68$ and $d = 10.66$, respectively). In addition to the MAE, we evaluated these models in Supporting Information III using scale-free measures like R-squared (R^2) and the similarity of the predicted FC states. Despite their conceptual differences, the results show that the RNN- and WN-based approach can 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 suitable to accurately predict the BOLD signal with this limited amount of data. 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.

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 the Spatial Dependencies section. The definition of an 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 $\mathbf{A}_{SC}$ is illustrated in Figure 3B. 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 (Becker et al., 2018; Bettinardi et al., 2018; H. Liang & Wang, 2017). 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 3C 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 (Y. Li et al., 2018) and the MAE of its predictions, averaged across test samples, brain regions, and predicted time points is depicted in blue. Figure 3C 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 (Wu et al., 2019), and its average test MAE is shown in red in Figure 3C. 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 $\mathbf{A}_{CE}$, as illustrated in Figure 3B. The comparison between $\mathbf{A}_{CE}$ and the structural connectivity matrix $\mathbf{A}_{SC}$ highlights that in the adjacency relation defined by the structural embeddings, long-range connections between brain regions are considerably more pronounced. Figure 3D shows the test MAE of the models when incorporating $\mathbf{A}_{CE}$ in the graph convolution operations. In this case we can observe for both models a sharp drop in the error at walk order $K = 1$. This suggests that the node embeddings already inherently capture higher order relations between nodes in the brain network. Finally, in Figure 3E 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. This observation indicates that it is rather challenging to learn all N^2 connections between brain regions without any 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 whether incorporating information about structural connections significantly increases the prediction accuracy of our models, we at first recomputed the overall test MAE for each subject again. Then by using a paired t test, we find that, for both STGNN models (DCRNN and GWN) and both adjacency types ($\mathbf{A}_{SC}$ and $\mathbf{A}_{CE}$), the impact of structural modeling is positive (Cohen's $d > 1$ for all comparisons) and significant ($p \leq 0.0001$ for all comparisons), compared to the case in which it is not considered. Although the performance differences between the GWN and DCRNN are quite small in general, the DCRNN slightly outperformed with a test error of $MAE = 0.1158$ the GWN with a test error of $MAE = 0.1211$ at $K = 1$ (significant with $p \leq 0.0001$, Cohen's $d = 0.49$). In addition, the distribution of the test error across subjects and ROIs, with and without the structural modeling in STGNNs is illustrated in Supporting Information III in Figure S6. 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 $K = 1$. This observation supports the idea that structural node embeddings can strengthen the relationship between structural data derived from DTI with functional data observed in fMRI (Rosenthal et al., 2018). When applying a paired t test, the improvement of the prediction accuracy when using the CE similarity in comparison to the SC became for both, the DCRNN, and GWN model, significant with $p \leq 0.0001$ at $K = 1$ (Cohen's $d = 1.45$ for the DCRNN and $d = 0.95$ for the GWN). By inherently capturing higher order transitions in $\mathbf{A}_{CE}$, 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.

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 most often used for directed functional connectivity analysis (Barnett & Seth, 2013; Friston et al., 2013). 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 model 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 (Glasser et al., 2016) (a complete list of selected ROIs is provided in the Supporting Information 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 (Glasser et al., 2016). Finally, the whole-brain network of 360 regions in total was incorporated. As discussed in the Data Description section, 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 method as implemented in the multivariate Granger

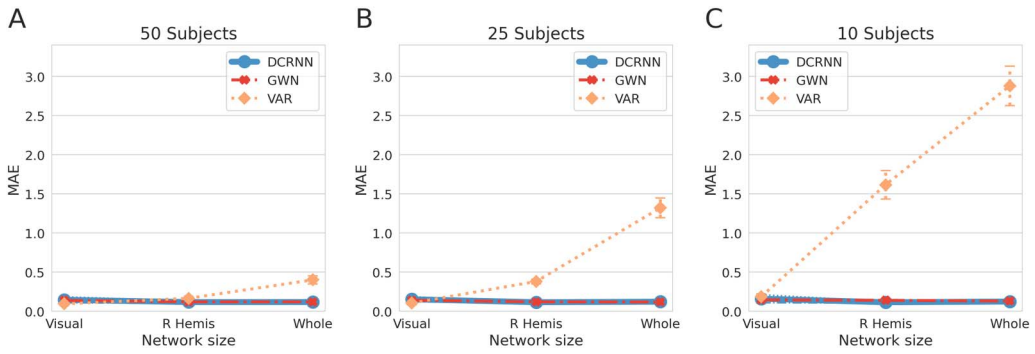

Figure 4. The figure shows a comparison of the model performances when varying the amount of data and the size of the network. The test MAE of the VAR is here depicted in orange, the MAE of the DCRNN in blue, and the error of the GWN in red. The overall error was computed as an average across brain regions, time steps and test samples. The error bars represent the standard deviation of the test error across subjects. (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 (Glasser et al., 2016). (B and C) The test performances in dependence of the network size using the 25 and 10 subject dataset, respectively.

causality toolbox (Barnett & Seth, 2013), 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 the Vector Autoregressive Model section. The hyperparmeters used for the STGNNs are described in the methods part in the Model Training section. Further, for this comparison, the CE similarity $\mathbf{A}_{CE}$ with transition order of $K = 1$ was used to define the structural relations in the STGNN models, which has shown to improve the GNNs forecasting accuracy with low computational cost, as discussed in the section Comparison of GNN Architectures.

Figure 4 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 4A 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 4B 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 4C. The DCRNN and GWN models can still achieve a high accuracy, also when a limited amount of data are available and the network size is relatively large. In addition, this comparison of the models is replicated in Supporting Information IV using additional measures like R^2 and the similarity of predicted FC states. After applying a paired t test, the differences between the DCRNN and GWN to the VAR were shown to be in all cases highly significant with $p \leq 0.0001$ (Cohen's $d \gg 1$), except when the VAR is only fitted to the single visual network, where it still could make reliable forecasts.

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 5. Figure 5A 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 time steps 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

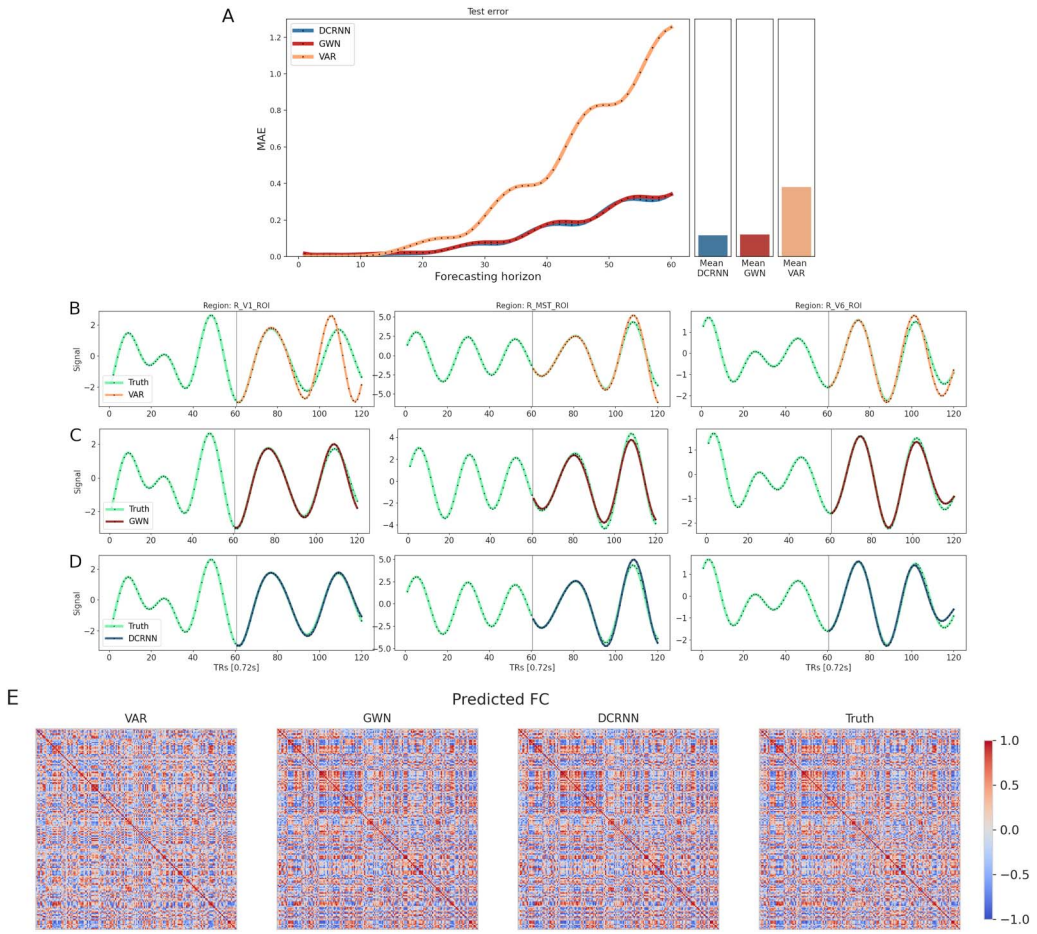

Figure 5. The prediction accuracy of the different models is presented in more detail for the 25 subject dataset and the brain network including the ROIs within the right hemisphere (Glasser et al., 2016). (A) The test MAE in dependence of the forecasting horizon is shown, computed as an average across test samples and brain regions. (B, C, and D) Examples of predictions generated by the VAR, GWN, and DCRNN model, respectively. The examples were chosen to be representative for the whole test set, by selecting only examples in which errors maximally deviate by 0.02 from the corresponding average test MAE of the models. (E) Examples of predicted FC states of the different forecasting models, including the true FC state on the right side. These representative examples deviated maximally by 0.005 from their average correlation to the true FC state.

different models in a few representative samples are shown in Figures 5B, C and D. Additionally, the predicted FC states $\mathbf{A}_{FC} \in \mathbb{R}^{N \times N}$ were computed as the Pearson correlation between predicted BOLD signals of all N brain regions, and a comparison of representative predictions with the true FC state is illustrated in Figure 5E. The average correlation of the predicted FC state to the true FC state was for the VAR model $r_{FC} = 0.635$ on this dataset, while the GWN could achieve a correlation value of $r_{FC} = 0.948$, and the DCRNN a value of $r_{FC} = 0.950$. The overall FC similarity for all different datasets of all prediction models is given in Supporting Information IV Figure S8. Furthermore, in Supporting Information V in Figure S9 we performed the analysis on the same dataset using a more liberal frequency filtering within the 0.01–0.1 Hz frequency range. In this range, the signal dynamic becomes more complex, and we can observe an increase in the prediction error of the different models accordingly.

In addition, we evaluated in more detail how the prediction errors are distributed across different subjects and different ROIs. Figure 6 shows the distribution of the test MAE of the DCRNN, GWN, and VAR model across subjects and in dependence of the brain region within

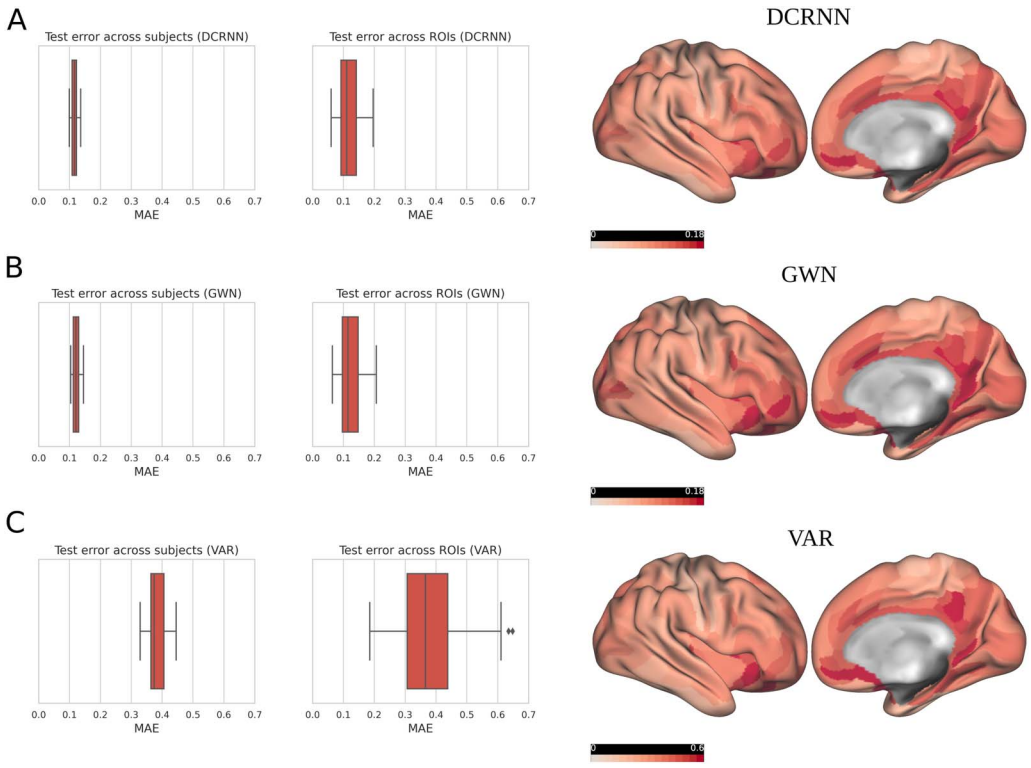

Figure 6. The distribution of the test error across subjects and brain regions is shown. (A) The MAE across subjects and brain regions of the DCRNN is first visualized in a boxplot on the left side. Additionally on the right side of the figure, the MAE values are projected onto the cortical surface within the right hemisphere, where the color map was linearly scaled between 0 and 0.18. (B) The distribution of the test MAE of the GWN is shown. (C) The MAE distribution of the VAR model. For the VAR, the color map was adjusted to account for larger error values by scaling it between 0 and 0.6.

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 toward a more complex BOLD dynamic in those regions. Alternatively, the prediction accuracy might also be affected by a lower signal-to-noise ratio observed in medial brain regions (Olman, Davachi, & Inati, 2009).

Multimodal Directed Connectivity

In the Comparison of GNN Architectures section, 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 GNN models. This points out that the GNNs are able to learn relevant and functional informative transitions of neural activity on their structural spatial layout. Based on the idea of Granger causality (Granger, 1969) that the observation of one event *A* carries information about the occurrence of a future event *B*, this might represent initial evidence for a potentially causal relation between *A* and *B*. Due to the relatively low temporal sampling rate and physiological artifacts in fMRI (S. M. Smith et al., 2011; Webb et al., 2013), it still is a matter of discussion to what extent we can observe a *causal* relationship between brain regions in this imaging modality (Bielczyk et al., 2019; Mill et al., 2017; H. E. Wang et al., 2014). But the observation that activity in one region carries additional information

among activity in another region in the brain can also go beyond simple undirected FC or SC, and we therefore refer to this kind of relation as *directed connectivity* in the following. Propagating the information between ROIs related to their SC or structural CE similarity has the potential to give us in this manner a multimodal perspective of such a directed relationship between different brain areas. For this purpose, we choose a perturbation base approach to reconstruct the amount of information individual ROI carries about other ROIs (Zeiler & Fergus, 2014). By learning a function $h(\cdot)$, the GNN models try to infer from an input sequence of neural activity states $[\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(T_p)}]$ a sequence of future activity states $[\hat{\mathbf{x}}^{(T_p+1)}, \ldots, \hat{\mathbf{x}}^{(T_p+T_f)}]$, whereby $\mathbf{x}^{(t)} \in \mathbb{R}^N$ denotes the activity at timestep t in all regions $n = 1, \ldots, N$. To induce an artificial 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 time series as an input for our trained model $h(\cdot)$ the model generates then a prediction $[\hat{\mathbf{x}}'^{(T_p+1)}, \ldots, \hat{\mathbf{x}}'^{(T_p+T_f)}]$. 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:

$$I_n(n') = \frac{1}{S}\sum_{s=0}^{S}\frac{1}{T_f}\sum_{t=0}^{T_f}\left|\hat{\mathbf{x}}_n^{(t)}(s) - \hat{\mathbf{x}}'^{(t)}_n(s)\right| \tag{2}$$

where $I_n(n')$ denotes the impact of ROI n' on n. Further, $\hat{\mathbf{x}}_n^{(t)}(s)$ and $\hat{\mathbf{x}}'^{(t)}_n(s)$ denote the predictions in ROI n with and without the perturbation in n' of one test sample s at time step t. Note that this artificial perturbation approach has only the goal of making spatial relations between ROIs in STGNN models explainable, and should not be equated with the effect of an experimental perturbation applied to the human brain, like, for example, induced by transcranial magnetic stimulation.

In the following we compare this proposed measure of directed influence $\mathbf{I}(n')$ 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 $\mathbf{I}(n')$, first using the SC as substrate for information propagation, captured in $\mathbf{A}_{SC}$, and then also employing the similarity of CEs, represented by $\mathbf{A}_{CE}$. In the following example, we study the connectivity of V1 within the right hemisphere by incorporating data of 25 subjects. For this purpose, a perturbation was induced into the target region V1, and the impact of this perturbation on all other 179 regions within the right hemisphere was then characterized by computing the measure of directed influence $\mathbf{I}(n')$, as defined in Equation 2. These values for directed connectivity strength can then be visualized by projecting them onto the 179 other areas of the cortical surface. For the following comparison, all connectivity values were rescaled by normalizing them between 0 and 100. At first, in Figure 7A the structural connectivity between V1 and all other 179 regions within the right hemisphere is depicted. The target region V1 is marked here in light blue, and the strength of connectivity to all other regions is encoded in red. Figure 7A shows that we can mainly observe a pronounced structural connectivity between V1 and V2 and some structural connections leading to V3. Figure 7B 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 7C the directed connectivity strength $\mathbf{I}(n')$ is depicted, when using the SC as spatial backbone for the information exchange between brain regions in the GNN. 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

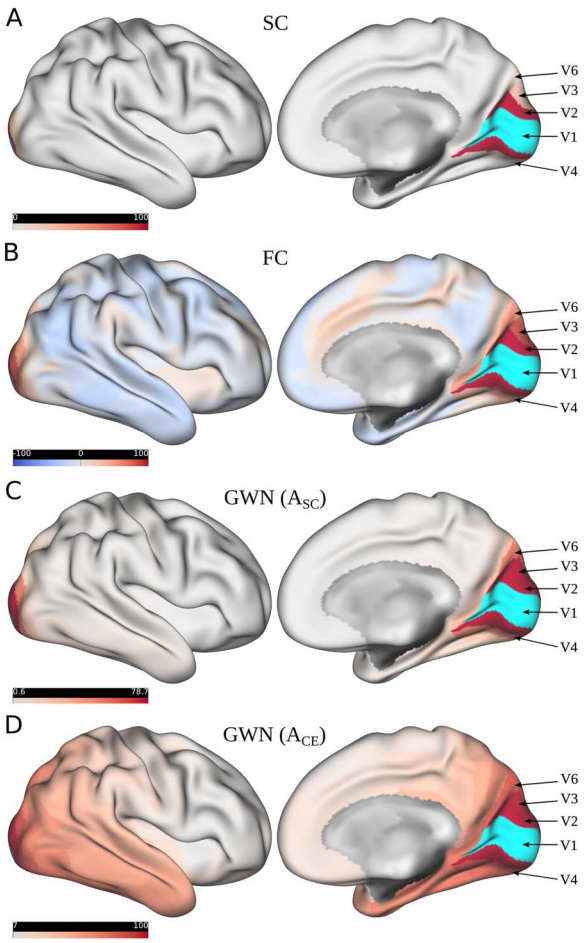

Figure 7. Different types of connectivity are illustrated between V1 and all other regions within the right brain hemisphere. (A) The structural connectivity is shown, whereby the target region V1 is marked in light blue and the connectivity strength is encoded in red. (B) The correlation-based functional connectivity is illustrated, which was computed as an average across subjects. (C) The measures of influence $\mathbf{I}(n')$, derived from the GWN model using the SC for information propagation. (D) 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 color map between the 2*th* and 98*th* percentile of the values respectively.

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 7D 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 (Rosenthal et al., 2018). Figure S10 in Supporting Information VI 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. So far, we have only studied the effect of a single artificial perturbation in V1 directed to all other regions within the right hemisphere. This approach can further be extended to sample a full connectivity network, by systematically inducing perturbations into all regions of interest for the analysis, and then systematically observing the effect on the other network regions. In Figure S11 in Supporting Information VI, we studied the effects of perturbations induced in some different additional areas of the visual network based on this approach. In this manner, this technique can allow us to reconstruct the directed spatial relations between brain areas captured in STGNN models, and could be applied in practical applications by, for example, comparing these connectivity patterns between different conditions in task-based fMRI, or studying the difference between healthy and diseased brain states.

Spatial autocorrelations:
The phenomenon in neuroscience that activity in nearby areas tends to be more strongly correlated than in more distant areas.

Temporal autocorrelations:
Characterizes the smoothness of the memory of neural activity time courses in different brain regions.

A study of Shinn et al. (2021) demonstrated that FC topology in resting-state fMRI is shaped by and can be predicted from spatial autocorrelations and temporal autocorrelations, as typically observed in fMRI data. To also investigate to what extend STGNN-based connectivity patterns are related to such correlations, we computed the temporal autocorrelation as the Pearson correlation between the BOLD signal values and its lagged values with different lag orders j, depicted in Figure S12 in Supporting Information VI. We could observe a relatively high temporal correlation of 0.89 around lag order $j = 14$, and of 0.62 around lag order $j = 27$, which shows that within these first 30 TRs of the signal, such temporal autocorrelations can still be detected. In section Model Accuracy and Network Scaling we could show that STGNNs were also able to make reliable long-term predictions of the BOLD signal up to a horizon of 60 TRs, which demonstrates that STGNNs capture properties of neural activity dynamics that go clearly beyond the range that is shaped by temporal autocorrelations. In addition, spatial autocorrelations can also play a distinctive role in shaping FC network properties (Shinn et al., 2021). The analysis in section Comparison of GNN Architectures showed that by modeling the information exchange between structurally connected brain regions in the spatial domain, the prediction accuracy of the STGNNs could be significantly improved, in comparison to the null models, which incorporated no spatial dynamics. The observation that spatially connected regions contain some *additional* relevant information on functional dynamics points out that spatial interactions captured in STGNNs go beyond simple correlation-based spatial network relations. A more detailed comparison of the individual differences and similarities between correlation-based FC and the STGNN-based connectivity pattern is additionally provided in a bar plot in Supporting Information VI (Figure S13).

CONCLUSION

In this study we have compared different STGNN architectures for learning the spatiotemporal dynamics in brain networks. First, in the section Comparison of GNN Architectures we studied different mechanisms for learning the temporal dynamics in the BOLD signal. We could show that an RNN-based model and a WN-based model exhibit very similar capabilities in learning the temporal characteristics in neural activity time series. 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 (He, Zhang, Ren, & Sun, 2016; Ioffe & Szegedy, 2015), its prediction error was considerably higher in comparison to the RNN- and WN-based approach. This indicates that the geometric

assumptions that 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 ($\mathbf{A}_{SC}$), the CE similarity ($\mathbf{A}_{CE}$), or a self-adaptive adjacency matrix ($\mathbf{A}_{Adap}$) 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 therefore 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. (2018), 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 spatiotemporal dynamics, this method also proved to strengthen the relationship between brain structure and functional dynamics. Learning all N^2 connections of the underlying structural graph during the model training has been shown to be challenging for the STGNN models, in case no prior knowledge is provided to them in the form of the anatomical brain connectivity. While such highly parameterized artificial neural network models can be in theory quite flexible in learning complex relations (Brüel Gabrielsson, 2020; Hornik, Stinchcombe, & White, 1989), often the decisive limitation is the successful optimization of the parameters during model training (Dauphin et al., 2014). In the discussed applications of STGNNs in fMRI, where the amount of training data is most often quite limited, prior knowledge in the form of the anatomical graph structure has been shown to considerably support the learning of spatial relations between brain areas captured in STGNN models.

So far, methods based on biophysical modeling (Deco et al., 2012, 2013; Honey et al., 2009; Messé et al., 2014, 2015; Messé, Rudrauf, Giron, & Marrelec, 2015), graph theory (Abdelnour et al., 2018; Becker et al., 2018; H. Liang & Wang, 2017; Lim et al., 2019), or machine learning (Amico & Goñi, 2018; Deligianni et al., 2016; Rosenthal et al., 2018; Sarwar et al., 2021) have contributed already numerous valuable insights into the structure-function relation in brain networks, and could highlight the role and importance of the structural connectome in shaping functional connectivity patterns. While the majority of approaches studying the structure-function relationship infer brain dynamics by fitting the models to empirically observed FC patterns, STGNNs provide us with a possibility to directly predict the observed neural activity states. Similar to some other recently proposed predictive models (Singh, Braver, Cole, & Ching, 2020; Suárez et al., 2021), this principle can allow us to investigate additional interesting aspects of dynamic brain functions. As discussed in section Comparison of GNN Architectures, this could enable us, for example, to study directly the amount of information on the activity of one ROI that is contained in the activity of other structurally connected ROIs. For a comparison with other currently used approaches investigating SC-FC mappings, the predicted BOLD signal states of STGNN can be used then again to reconstruct predictions for FC states, as shown in Figure 5E. The relatively high accuracy in predicting empirical FC states already points out the potential of STGNNs in this field. Moreover, in comparison to other currently popular approaches used in this area (Messé et al., 2015), by learning localized graph filters in STGNNs, their forecasting accuracy is also robust with regards to the brain network size. While such highly parameterized artificial neural network models appear to be promising for achieving high prediction accuracies of FC states

(Sarwar et al., 2021), they cannot provide us with the same mechanistic insights into physiological processes as biophysically inspired models. Still, they can be used to supplement current biophysically inspired models, for studying different aspects of the structure-function relationship from a novel data-driven perspective. A more comprehensive comparison of these different new approaches, evaluating in detail their interrelations like in the study of Messé et al. (2015), could be thereby interesting for future studies in this area.

In the Model Accuracy and Network Scaling section, we have compared the STGNN models to a VAR model, which is currently most often used in Granger causality analysis for inference of directed relationships between brain regions (Barnett & Seth, 2013). We evaluated the accuracy of the different approaches on a variety of brain network sizes and dataset 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, also 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. A classical VAR model fits a parameter for every possible connection between the N regions in a network, so that the number of parameters in a VAR-based approach grow strongly with an order of N^2. In contrast thereto, STGNNs utilize prior information in the form of the anatomical connectivity, and then model the functional information exchange based on this underlying structural substrate. By incorporating graph convolution operations in STGNNs, the amount of parameters only linearly scale with walk order K, which can even be chosen to be $K = 1$, if higher order structural relations are already expressed in an adjacency matrix derived from connectome embeddings ($\mathbf{A}_{CE}$). This property allowed STGNNs to make very robust inferences also on large networks and when only limited data are available, thereby providing a flexible method for various connectivity analysis scenarios.

Finally in the Multimodal Directed Connectivity section, we studied the individual spatial interactions within the brain network that 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 relationships between brain regions. When comparing this measure of influence to the original structural connectivity, we can observe that STGNNs have learned to include transitions along higher order structural connections in the network. The models could infer links between $V1$ and $V2$, but additionally strong connections to $V3$ and $V4$. Especially when incorporating the CE-based similarity $\mathbf{A}_{CE}$ to define spatial node relations in the STGNN models, we can observe a high integration of $V1$ within the visual system. However, due to the relatively low temporal sampling rate in fMRI (Friston et al., 2013), and the indirect measurement of neural signals based on their hemodynamic response (Webb et al., 2013), one should also be aware of these limitations in the inference of directed and potentially causal connections in fMRI studies (S. M. Smith et al., 2011). Our lag-based predictive approach based on STGNN models might therefore also be affected by the same limitations as classical Granger causality in fMRI. On the other hand, a combined fMRI-MEG study by Mill et al. (2017) and different computational simulations of fMRI data

(Duggento et al., 2018; Seth et al., 2012; H. E. Wang et al., 2014; Wen, Rangarajan, & Ding, 2013) could establish evidence that Granger causality is still able to identify meaningful directed relationships between brain areas in fMRI, despite the indirect measurements based on the hemodynamic response. As an alternative, deconvolution-based approaches can have the potential to infer from the measured BOLD signals the underlying neural time series (Bush et al., 2015; Mill et al., 2017) for assessing *effective* brain connectivity, rather than only estimating *directed* functional connectivity. But the estimation of the underlying hemodynamic response from the data might come with the cost of introducing additional assumptions and uncertainties (Bielczyk et al., 2019; Roebroeck, Formisano, & Goebel, 2011). A more detailed discussion on considerations concerning Granger causality, and, in general, causal inference in fMRI, is provided in the comprehensive review of Bielczyk et al. (2019), as well as in the perspective on FC and its variants by Reid et al. (2019). Despite these current limitations in fMRI, a multimodal GNN-based approach can allow us to join structural and functional imaging data in a new manner, and reveals thereby potential for supplementing current analysis methods in brain connectivity research by studying such directed relations under a novel perspective (Reid et al., 2019).

In conclusion, in our study we found that 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 STGNNs in the high-frequency range. While in this presented approach we only incorporated a single temporal feature (the BOLD signal) into the STGNNs, in general, such a flexible data-driven approach could be expanded to account for different types of data and annotations. For example, the activity measured in a combined EEG-fMRI experiment (Abreu, Leal, & Figueiredo, 2018; Mele et al., 2019) could be also simultaneously integrated in STGNNs as different temporal features, or adding the temporal response of a subject could be helpful to better predict activity patterns in task-based fMRI. Also, alternative structural imaging techniques like neurite orientation dispersion and density imaging (NODDI) (Zhang, Schneider, Wheeler-Kingshott, & Alexander, 2012) might capture additional aspects of the brain structure, which could be included as structural information in STGNN-based models. In clinical applications multimodal STGNNs could be interesting for studying how the relationship between structure and function is affected in the diseased brain (Panda et al., 2021), or which impact a structural lesion might have on the functional organization of the brain network (Alstott, Breakspear, Hagmann, Cammoun, & Sporns, 2009). 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 (de Haan, Cohen, & Welling, 2020; Schnake et al., in press).

MATERIALS AND METHODS

Dataset

The MRI dataset used in our study is provided by the HCP data repository (Hodge et al., 2015; Van Essen et al., 2013). As part of the HCP protocol, the study participants gave written informed consent to the HCP consortium. The MRI scanning protocols were approved by the Institutional Review Board at Washington University in St. Louis. We incorporated data of the *S1200 release*, which provides data from resting-state fMRI sessions, each with a

duration of 14.4 minutes, whereby 1,200 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 multiband (factor 8) acceleration (Feinberg et al., 2010; Moeller et al., 2010; Setsompop et al., 2012; J. Xu et al., 2012). A gradient-echo echo-planar imaging (EPI) sequence with a repetition time $TR = 720$ ms and an echo time $TE = 31.1$ ms was used. The field of view of the fMRI sequence was FOV = 208 mm × 180 mm and in total $N_s = 72$ slices with a slice thickness of $d_s = 2$ mm were collected, containing voxels with an isotropic size of 2 mm. The preprocessing of the HCP fMRI data includes corrections of gradient-nonlinearity-induced distortions, registration to a single-band reference image to account for subject motion and registration to the structural T1w image (Fischl, 2012; Glasser et al., 2013; Jenkinson, Bannister, Brady, & Smith, 2002; Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012). Further, ICA-FIX was applied to automatically classify and remove artifactual components in the resting-state fMRI data (Griffanti et al., 2014; Salimi-Khorshidi et al., 2014; S. M. Smith et al., 2013). Finally, the volumetric fMRI images are mapped into the CIFTI grayordinate space and Gaussian surface smoothing with a FWHM of 2 mm is performed. A detailed description of the standard minimal preprocessing pipelines of the HCP can be found in Glasser et al. (2013). In a next step to define our brain network, the multimodal parcellation proposed by Glasser et al. (2016) was applied, which divides the cortical surface into 180 segregated areas per hemisphere. 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 apply global signal regression in our preprocessing (Power, Plitt, Laumann, & Martin, 2017). Firstly, in a systematic comparison of different preprocessing methods to address motion artifacts Ciric et al. (2017) could show that an ICA-based denosing in combination with global signal regression is among the most effective methods to reduce movement artifacts. This result is in line with the study of Burgess et al. (2016), investigating the effect of ICA-FIX in combination with global/grayordinate signal regression on resting-state fMRI data provided by the HCP. Furthermore, in our study of functional interactions between specific brain regions, the objective was to extract the additional information, which certain regions contain about the activity in other regions. Therefore, local interactions rather than global modulations in the signal were the main interest for our analysis (Power et al., 2017). The time courses were then band pass filtered in the 0.04–0.07 Hz frequency range. In a summary of several different studies that account for different artifacts in the BOLD signal related to MRI scanner drift in the frequency range below 0.015 Hz (A. M. Smith et al., 1999), respiratory and cardiac frequencies around 0.3 Hz and 1–2 Hz respectively (B. Biswal, DeYoe, & Hyde, 1996), and fluctuations in arterial carbon dioxide level around 0.0–0.05 Hz (Wise, Ide, Poulin, & Tracey, 2004), the study of Glerean et al. (2012) identified the 0.04–0.07 Hz frequency band to be most reliable and relevant for gray matter activity in resting-state fMRI (Achard et al., 2006; Buckner et al., 2009; Zou et al., 2008). To additionally ensure that the low-frequency signals are not mainly related to respiratory artifacts, we studied the respiratory signals recorded with a Siemens respiratory belt during resting-state fMRI, as provided by the HCP (S. M. Smith et al., 2013). The average respiratory frequency spectrum is depicted in Supporting Information VII in Figure S14, and we can observe that respiratory fluctuations are mainly present in the higher frequency range around 0.28 Hz in this resting-state fMRI dataset. In addition, the different models were tested on data incorporating a more liberal frequency filtering within the 0.01–0.1 Hz range, as presented in Supporting Information V.

In the S1200 release, diffusion MRI data was collected in six runs, whereby approximately 90 directions were sampled during each run, using three shells of $b = 1{,}000$, 2,000, and 3,000 s/mm^2, with additionally 6 $b = 0$ images (Sotiropoulos, Jbabdi, et al., 2013). A spin-echo EPI

sequence was incorporated with repetition time TR = 5,520 ms, echo time TE = 89.5 ms, using a multiband factor of 3. In total N_s = 111 slices were collected, with field of view FOV = 210 mm × 180 mm and an isotropic voxel size of 1.25 mm. The minimal preprocessing pipeline of the HCP includes intensity normalization across runs, EPI distortion correction using the FSL5 "topup" tool, correction of eddy current–induced field inhomogeneities and head motion artifacts using the FSL5 "eddy" tool, and finally includes gradient nonlinearity corrections and registration to the structural T1w image (Andersson, Skare, & Ashburner, 2003; Andersson & Sotiropoulos, 2015a, 2015b; Glasser et al., 2013; Sotiropoulos, Moeller, et al., 2013). More details on the minimal preprocessing of the HCP diffusion MRI are described in Glasser et al. (2013). To reconstruct the anatomical connection strengths between regions within the multimodal parecellation (Glasser et al., 2016), the MRtrix3 software package was incorporated (Tournier et al., 2019). Multishell multitissue-constrained spherical deconvolution (Jeurissen et al., 2014) was applied to obtain response functions for fiber orientation distribution estimation (Tournier, Calamante, & Connelly, 2007; Tournier, Calamante, Gadian, & Connelly, 2004). Then 10 million streamlines were created using anatomical constrained tractography (R. Smith, Tournier, Calamante, & Connelly, 2012). Finally, spherical deconvolution–informed filtering was used (R. Smith, Tournier, Calamante, & Connelly, 2013), 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 $\mathbf{A}_{SC}$ was obtained as the average SC across the first 10 subjects, because the variance in the SC strength was relatively low across subjects (Zimmermann, Griffiths, Schirner, Ritter, & McIntosh, 2018), 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.

Graph Neural Networks

Different brain areas communicate via bioelectrical signals transmitted along neuronal axons and collected by neuronal dendrites. Spatiotemporal GNNs provide a novel possibility to incorporate such a structural scaffold into a graph-based prediction model (Wein, Malloni, et al., 2021). Due to cognitive information processing in the brain, the spatial interactions of the activity distribution changes dynamically. Spatiotemporal 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 spatiotemporal brain dynamics in resting-state fMRI (Wein, Malloni, et al., 2021). 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 (Y. Li et al., 2018). 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. (2019). These authors built upon WaveNets (van den Oord et al., 2016) 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 (van den Oord et al., 2016). 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 (Vaswani et al., 2017; Zheng et al., 2020).

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 spatiotemporal 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 spatiotemporal modeling of functional MRI.

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 a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A}_w)$ encompassing N vertices, that is, 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 $\mathbf{A}_w \in \mathbb{R}^{N\times N}$, whose entries $w_{nn'}$ provide the connection strengths between vertices 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 $\mathbf{A}_{SC}$, we next employed CE to additionally capture higher order topological features in SC represented by $\mathbf{A}_{CE}$, and finally we treated the spatial relations as adaptive learnable parameters $\mathbf{A}_{Adap}$ 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 $\mathbf{X} = (\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(T)}) \in \mathbb{R}^{N\times T}$ with $\mathbf{x}^{(t)} \in \mathbb{R}^N$. 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:

$$\mathcal{X}_{::m} \equiv \mathbf{X}^{(m)} = \left(\mathbf{x}_1^{(m)} \ldots \mathbf{x}_T^{(m)}\right) = \begin{pmatrix} x_{11}^{(m)} & \cdots & x_{1T}^{(m)} \\ \vdots & x_{nt}^{(m)} & \vdots \\ x_{N1}^{(m)} & \cdots & x_{NT}^{(m)} \end{pmatrix} \in \mathbb{R}^{N\times T} \tag{3}$$

Note that the columns $\mathbf{x}_t^{(m)} \in \mathbb{R}^N$ of the data matrix describe the activation of all ROIs at any given time point $1 \le t \le T$, while its rows $\widetilde{\mathbf{x}}_n^{(m)}$ represent the meta-voxel time course of every single ROI $1 \le n \le N$. 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 $\mathcal{X}_{::m} \in \mathbb{R}^{N\times T}$, $\mathcal{X}_{:t:} \in \mathbb{R}^{N\times M}$, and $\mathcal{X}_{n::} \in \mathbb{R}^{T\times M}$.

In addition to above frontal slices $\mathcal{X}_{::m} \equiv \mathbf{X}^{(m)}$ of the data tensor $\mathcal{X}$, we thus have the lateral tensor slices:

$$\mathcal{X}_{:t:} \equiv \mathbf{X}^{(t)} = \left(\mathbf{x}_1^{(t)} \ldots \mathbf{x}_M^{(t)}\right) = \begin{pmatrix} x_{11}^{(t)} & \cdots & x_{1M}^{(t)} \\ \vdots & x_{nm}^{(t)} & \vdots \\ x_{N1}^{(t)} & \cdots & x_{NM}^{(t)} \end{pmatrix} \in \mathbb{R}^{N\times M} \tag{4}$$

and the horizontal tensor slices:

$$\mathcal{X}_{n::} \equiv \mathbf{X}^{(n)} = \left(\mathbf{x}_1^{(n)} \dots \mathbf{x}_M^{(n)}\right) = \begin{pmatrix} x_{11}^{(n)} & \cdots & x_{1M}^{(n)} \\ \vdots & x_{mt}^{(n)} & \vdots \\ x_{T1}^{(n)} & \cdots & x_{TM}^{(n)} \end{pmatrix} \in \mathbb{R}^{T \times M} \tag{5}$$

Note that the column fibers of the data tensor $\mathcal{X}_{:tm:}$ denoted as $\mathbf{x}_t^{(m)}$ 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 $\mathbf{x}_n^{(m)}$, represent the time course of every feature m at node n. Finally, the tube fibers of the tensor $\mathcal{X}_{nt:}$ denoted as $\mathbf{x}_t^{(n)}$, 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.

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 (Y. Li et al., 2018; Wu et al., 2021). The information flow in the underlying graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A}_w)$ is considered as a stochastic random walk process modeled by a state transition matrix $\mathbf{T} = \mathbf{D}^{-1}\mathbf{A}_w = (\hat{\mathbf{w}}_1 \dots \hat{\mathbf{w}}_N)$ where $\mathbf{A}_w$ represents a weighted adjacency matrix. The diagonal node degree matrix is given by:

$$\mathbf{D} = diag(\mathbf{A}_w \mathbf{1}) \tag{6}$$

where $\hat{\mathbf{w}}_n = (\hat{w}_{1n} \dots \hat{w}_{Nn})^T \in \mathbb{R}^N \; \forall \; n = 1, \dots, N$ with $\hat{w}_{nn'} = w_{nn'}/\Sigma_{n'}$, $w_{nn'}$ 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:

$$\mathbf{L}_{rw} = \mathbf{I} - \mathbf{T} = \mathbf{U}\hat{\mathbf{\Lambda}}\mathbf{U}^T = \mathbf{U}(\mathbf{I} - \mathbf{\Lambda})\mathbf{U}^T = \mathbf{I} - \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \tag{7}$$

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 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 $\mathcal{G}_C$ (Shuman, Narang, Frossard, Ortega, & Vandergheynst, 2013), which served to compute the spatial convolution of the input signal and a spatial filter kernel to yield the output of the ℓ-th convolution layer as:

$$\begin{aligned} \mathbf{y}_t^{(q)} &= \mathbf{U}\mathbf{\Theta}_\omega^{(q)}\mathbf{x}_\omega^{(m)} = \mathbf{U}\mathbf{\Theta}_\omega^{(q)}\mathbf{U}^T\mathbf{x}_t^{(m)} \\ &\approx \sum_{k=0}^{K} \theta_k^{(q)}(\boldsymbol{\omega})\mathbf{T}^k\mathbf{x}_t^{(m)} \end{aligned} \tag{8}$$

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 $\mathbf{T}$ (Defferrard et al., 2016; Wein, Malloni, et al., 2021). Finally, considering a CNN architecture and applying the graph convolution operator $\mathcal{G}_C$, the filtered input signal $\mathbf{y}_t^{(q)}$ was transformed with an activation function $\sigma(\cdot)$ to yield the output $\mathbf{h}_t^{(q)}$ of each of the $q \in \{1, \dots, Q\}$ graph convolution layers as follows:

$$\mathbf{h}_t^{(q)} = \sigma\left(\mathbf{y}_t^{(q)}\right) = \sigma\left(\sum_{k=0}^{K} \theta_k^{(q)}(\boldsymbol{\omega})\mathbf{T}^k\mathbf{x}_t^{(m)}\right) \tag{9}$$

Hereby $\mathbf{x}_t^{(m)} \in \mathbb{R}^N$ denotes the m-th input feature component at time t, $\mathbf{h}_t^{(q)} \in \mathbb{R}^N$ the corresponding output component of the q-th convolution channel, $\mathbf{\Theta}_k^{(q)} \in \mathbb{R}^N$ 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 ℓ 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 incorporate 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 (Tournier et al., 2004). 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 the Dataset section.

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 (Rosenthal et al., 2018). 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, Sutskever, Chen, Corrado, and Dean (2013) 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 (Grover & Leskovec, 2016; Rosenthal et al., 2018). 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 (Řehůřek & Sojka, 2010) using the skip-gram model to learn the node representations (Mikolov et al., 2013). 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 (Grover & Leskovec, 2016). 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 Pearson correlation coefficient, yielding a connectivity matrix denoted with $\mathbf{A}_{CE} \in \mathbb{R}^{N \times N}$. As illustrated in Figure 3B, the embeddings could yield meaningful representations that revealed long-range connections between regions that were not present in the original SC (Rosenthal et al., 2018).

Adaptive adjacency matrix. So far the spatial layout of the brain graph has been represented with the help of the orthogonal eigenbasis system $\mathbf{U}$ 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 $\mathbf{A}_{Adap} \in \mathbb{R}^{N \times N}$ (Wu et al., 2019). The latter has been constructed as a matrix of trainable weights $\mathbf{V}_{Adap} \in \mathbb{R}^{N \times N}$, which were at first initialized as zero and then again optimized via

gradient descent (Kingma & Ba, 2014). Inspired by the study of Wu et al. (2019), a normalized self-adaptive adjacency matrix was computed as:

$$\mathbf{A}_{Adap} = \frac{\sigma\left(\mathbf{V}_{Adap}\right)}{N} \tag{10}$$

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:

$$\mathbf{h}_t^{(q)} = \sigma\left(\mathbf{y}_t^{(q)}\right) = \sum_{k=0}^{K}\left[\theta_k^{(q)}\mathbf{T}^k + \theta_k^{(q),Adap}\left(\mathbf{A}_{Adap}\right)^k\right]\mathbf{x}_t^{(m)} \tag{11}$$

Note that the normalized self-adaptive adjacency matrix $\mathbf{A}_{Adap}$ 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 can in this case be obtained with:

$$\mathbf{h}_t^{(q)} = \sigma\left(\mathbf{y}_t^{(q)}\right) = \sigma\left(\sum_{k=0}^{K}\theta_k^{(q),Adap}\left(\mathbf{A}_{Adap}\right)^k\mathbf{x}_t^{(m)}\right) \tag{12}$$

Temporal Dependencies

Recurrent neural networks. In the DCRNN model, the temporal variations of the signal $\mathbf{x}_t^{(m)} \in \mathbb{R}^N$ in N brain regions at T_p past time points were explored with sequence-to-sequence learning in RNNs (Sutskever et al., 2014), 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 8A.

Given that the graph convolution operation accounts for the spatial layout of graph structure at any time point t, temporal dynamics on the graph can be modeled in the DCRNN via GRUs (Chung et al., 2014). 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) (Y. Li et al., 2018):

$$\begin{aligned}
\mathbf{r}_t^{(q)} &= \sigma\left(\mathcal{G}_C\left(\mathbf{\Theta}_r^{(q)}, \left[\mathbf{x}_t^{(m)} \,||\, \mathbf{h}_{t-1}^{(q)}\right]\right) + \mathbf{b}_r\right)\\
\mathbf{u}_t^{(q)} &= \sigma\left(\mathcal{G}_C\left(\mathbf{\Theta}_u^{(q)}, \left[\mathbf{x}_t^{(m)} \,||\, \mathbf{h}_{t-1}^{(q)}\right]\right) + \mathbf{b}_u\right)\\
\mathbf{c}_t^{(q)} &= \tanh\left(\mathcal{G}_C\left(\mathbf{\Theta}_c^{(q)}, \left[\mathbf{x}_t^{(m)} \,||\, \left(\mathbf{r}_t^{(q)} \odot \mathbf{h}_{t-1}^{(q)}\right)\right]\right) + \mathbf{b}_c\right)\\
\mathbf{h}_t^{(q)} &= \mathbf{u}_t^{(q)} \odot \mathbf{h}_{t-1}^{(q)} + \left(1 - \mathbf{u}_t^{(q)}\right) \odot \mathbf{c}_t^{(q)}
\end{aligned} \tag{13}$$

where $\mathbf{x}_t^{(m)}$, $\mathbf{h}_t^{(q)}$ denote the m-th input and q-th output graph signal feature component of the GRU, respectively, at time t, and $[\mathbf{x}_t^{(m)} \,||\, \mathbf{h}_{t-1}^{(q)}]$ denotes their concatenation. Also $\mathbf{r}_t^{(q)}$, $\mathbf{u}_t^{(q)}$ represent reset and update gates at time t, and $\mathbf{b}_r$, $\mathbf{b}_u$, $\mathbf{b}_c$, respectively, denote bias terms.

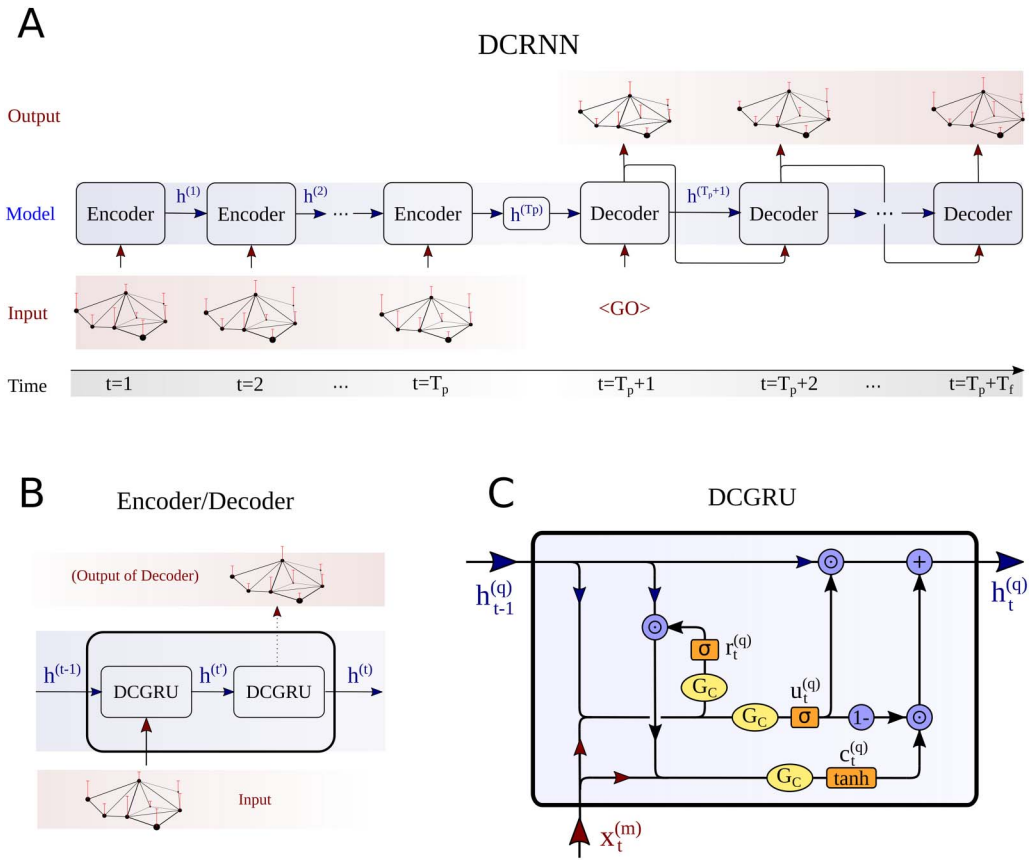

Figure 8. (A) Overview of the complete DCRNN model. The RNN architecture consists of an encoder and decoder, which recursively process the graph structured signals. The encoder receives a sequence of inputs $[\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(T_p)}]$ and iteratively updates the hidden state $\mathbf{h}^{(t)}$. The final state of the encoder $\mathbf{h}^{(T_p)}$ is passed to the decoder branch, which then recursively predicts the output sequence of future signals $[\mathbf{x}^{(T_p+1)}, \ldots, \mathbf{x}^{(T_p+T_f)}]$. 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 (< *GO* > symbol) is simply a vector of zeros. The processing steps in an individual DCGRU cell are shown in (C). The input $\mathbf{x}_t^{(m)}$, as well as the previous hidden state $\mathbf{h}_{t-1}^{(q)}$ are concatenated and passed to the reset gate $\mathbf{r}_t^{(q)}$, as well as to the update gate $\mathbf{u}_t^{(q)}$. The reset gate $\mathbf{r}_t^{(q)}$ determines the proportion of $\mathbf{h}_{t-1}^{(q)}$, which enters $\mathbf{c}_t^{(q)}$, together with input $\mathbf{x}_t^{(m)}$. Then the hidden state $\mathbf{h}_{t-1}^{(q)}$ is updated by $\mathbf{c}_t^{(q)}$, whereby the amount of new information is controlled by $\mathbf{u}_t^{(q)}$.

Furthermore, $\mathbf{\Theta}_r^{(q)}$, $\mathbf{\Theta}_u^{(q)}$, $\mathbf{\Theta}_c^{(q)}$ denote the parameter sets of the corresponding filters. An illustration of the complete sequence-to-sequence architecture incorporating DCGRU cells is provided in Figure 8.

WaveNets. Rather than incorporating diffusion convolution layers into RNNs, dilated causal convolution (DCC) layers (van den Oord et al., 2016) have been instead employed in the GWN architecture (Wu et al., 2019). The full GWN model is illustrated in Figure 9. The DCC was defined through a dilated causal convolution operator $\mathcal{D}_C$:

$$\mathcal{D}_C\left(\mathbf{\Theta}_t^{(q)}, \mathbf{x}_t^{(m)}\right) = \sum_{r=0}^{R-1} \mathbf{\Theta}_r^{(q)} \mathbf{x}_{t-d\cdot r}^{(m)} \tag{14}$$

whereby d denoted the dilation factor and $\mathbf{\Theta}_t^{(q)}$ represented the filter kernel. DCC could be implemented by sliding over the input time series $\mathbf{x}_t^{(m)}$ while skipping input values while, from

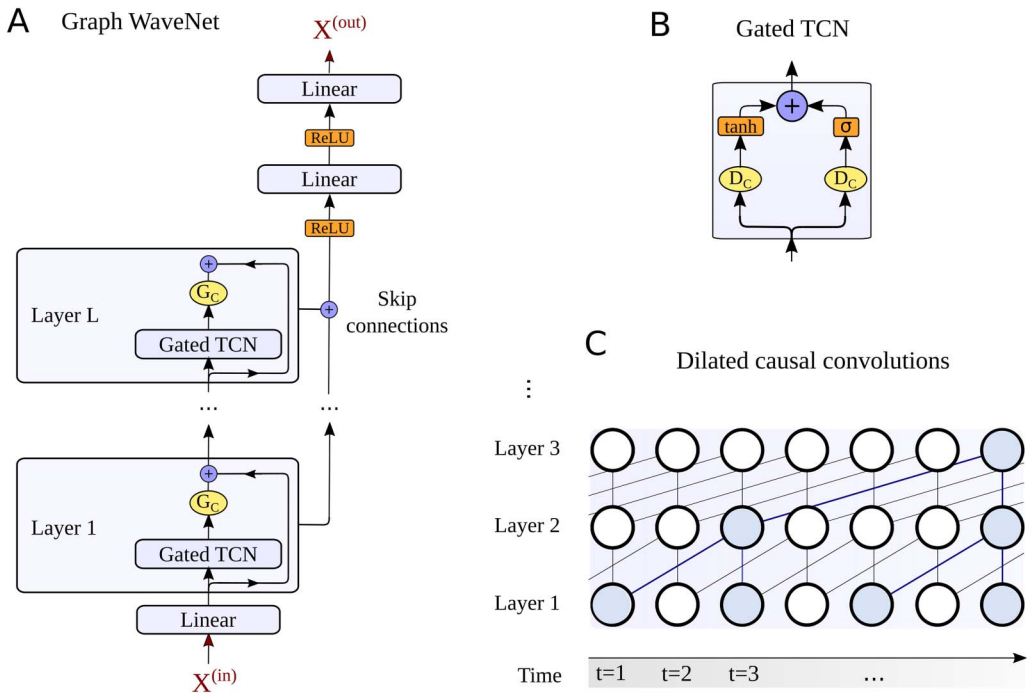

Figure 9. (A) An overview of the complete GWN model. The GWN model consists of L layers. For the temporal modeling, the GWN applies first the gated TCN mechanism and then for the spatial aspects utilizes graph convolution operations ($\mathcal{G}_C$) in each layer. Each layer additionally incorporates residual connections to stabilize the gradient during learning (He et al., 2016). 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 ($\mathcal{D}_C$) in combination with a tanh(·) and a σ(·) activation function to control the information flow. (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. These dilations lead to exponentially growing receptive fields for neurons in higher layers. The receptive field of a neuron in layer is highlighted in blue.

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 9C. The information flow was controlled by a gated temporal convolution network (TCN) as shown in Figure 9B, which is obtained as:

$$\mathbf{h}_t^{(q)} = \tanh\left(\mathcal{D}_C\left(\mathbf{\Theta}_1^{(q)}, \mathbf{x}_t^{(m)} + \mathbf{b}_1\right)\right) \odot \sigma\left(\mathcal{D}_C\left(\mathbf{\Theta}_2^{(q)}, \mathbf{x}_t^{(m)} + \mathbf{b}_2\right)\right) \tag{15}$$

Here tanh(·) denotes the output activation function, and $\mathbf{\Theta}_1^{(q)}$, $\mathbf{\Theta}_2^{(q)}$, and $\mathbf{b}_1$, $\mathbf{b}_2$ represent the convolution filers and bias terms, respectively. Further, $\mathcal{D}_C$ represents the causal convolution operator, $\odot$ the Hadamard product and σ(·) 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 (van den Oord et al., 2016). After each such dilated convolution layer, a diffusion convolution layer $\mathcal{G}_C$ (Equation 9) is subsequently applied to account for the spatial dependencies, as illustrated in Figure 9A.

Temporal relevance. Yet another approach to solve spatiotemporal time series prediction problems considers attention mechanisms in spatial and temporal domains to capture dynamic correlations (Vaswani et al., 2017; Zheng et al., 2020). In this study, we therefore additionally explore nonlinear 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 $\mathcal{X}_{n::} \equiv \mathbf{X}^{(n)} \in \mathbb{R}^{T\times M}$. The activity at any node n and at any time t was then represented by the tube fibers $\widetilde{\mathbf{x}}_t^{(n)} \in \mathbb{R}^M$, 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, as illustrated in Figure 10. The queries and keys are computed from the input in the ℓ-th block at time point t with a simple nonlinear transformation $g_r(\widetilde{\mathbf{x}}_t^{(n)}) = \mathrm{ReLU}(\mathbf{W}_r\widetilde{\mathbf{x}}_t^{(n)} + \mathbf{b}_r)$ with parameters $\mathbf{W}_r \in \mathbb{R}^{D\times M}$ and $\mathbf{b}_r \in \mathbb{R}^D$. For any node n and any time point t_i the relevance of its states $\widetilde{\mathbf{x}}_{t_i}^{(n)}$ at time points $t_j < t_i$ with respect to the considered state $\widetilde{\mathbf{x}}_{t_j}^{(n)}$ could then be assessed by computing the inner product between the queries and keys:

$$\delta_{t_i,t_j}^{(n)} = \frac{\left(g_r\left(\widetilde{\mathbf{x}}_{t_i}^{(n)}\right)\right)\cdot\left(g_r\left(\widetilde{\mathbf{x}}_{t_j}^{(n)}\right)\right)^T}{\sqrt{D}}. \tag{16}$$

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

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

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 derived as:

$$\widetilde{\mathbf{h}}_{t_i} = g_r\left(\sum_{t_j<t_i}\hat{\delta}_{t_i,t_j}^{(n)}\cdot g_r\left(\widetilde{\mathbf{x}}_{t_j}^{(n)}\right)\right) \tag{18}$$

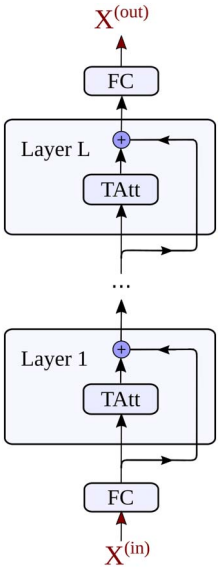

Figure 10. An overview over the temporal relevance or attention model. The single feature input $X^{(in)}$, representing the BOLD signal, is first projected by a fully connected layer onto M output features. The temporal relevance scores are computed in each of the L attention layers, and to further account for vanishing gradients, additional residual connects are incorporated (He et al., 2016). The output of the final layer L is then projected back onto a single feature, representing the predicted neural signal.

whereby $g_r(\cdot)$ denotes a nonlinear projection again. Note that all parameters $\mathbf{W}_r$ and $\mathbf{b}_r$ 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 (He et al., 2016).

Model Training

In this section we outline the training procedures that 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 linearly scaled between 0 and 1, to get gradients of a small magnitude during the backpropagation learning, which facilitates the fine-tuning of the learning rate. For all models, the mean absolute error (MAE) was used as an objective function to quantify the overall difference between the true BOLD signal $\mathbf{x}^{(t)}$ and predicted signal $\hat{\mathbf{x}}^{(t)}$ in all N brain regions:

$$\mathbf{MAE}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{T_f}\sum_{t=1}^{T_f}\left|x_n^{(t)} - \hat{x}_n^{(t)}\right| \tag{19}$$

DCRNN. The DCRNN model, based on an RNN architecture, was trained with backpropagation through time (Werbos, 1990), with the objective to maximize the likelihood of generating the target time series. To additionally account for a mismatch between training and testing distributions of stimuli, a scheduled sampling strategy was used (Bengio, Vinyals, Jaitly, & Shazeer, 2015). The probability of using a true label as a decoder input decayed according to:

$$\epsilon(i) = \frac{\tau}{\tau + \exp(i/\tau)} \in (0, 1) \tag{20}$$

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 (Kingma & Ba, 2014) was employed, and the model was trained for 70 epochs on minibatches of 16 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 dataset including only 10 subjects used in the Model Accuracy and Network Scaling section, 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 Supporting Information I (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 Linux 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. To have error values of a magnitudes that are easier to interpret, for the evaluations in the Results section the whole dataset was rescaled to zero mean and unit variance after the training of all STGNN models was finished.

GWN. The GWN model was also trained incorporating the Adam optimizer (Kingma & Ba, 2014) to minimize the MAE defined in Equation 19. For the GWN model, it was sufficient to train it 30 epochs with a batch size of eight, 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 Supporting Information 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 (Kingma & Ba, 2014) for in total 40 epochs, minimizing the MAE defined in Equation 19, using a batch size of 16. The learning rate was initialized with $\eta = 0.1$ and decreased by a factor of 0.1 at epochs 10, 20, and 30. The influence of the hyperparmeters is evaluated in Supporting Information I (Figure S3). The number of neurons in the temporal attentions were set to 32 thereby using four attention heads in the four TAtt layers. With this setup of hyperparameters one epoch of the TAtt model took around 7.7 minutes.

Vector Autoregressive Model

Granger causality (Granger, 1969) is currently most often based on linear vector autoregressive (VAR) models for stochastic time series data. Therefore we compare our GNN-based approach with a VAR model, as implemented in the multivariate Granger causality (MVGC) toolbox (Barnett & Seth, 2013). An autoregressive process 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 (Luetkepohl, 2005):

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

with coefficients or weights $\alpha_1, \ldots, \alpha_p$, an intercept β and the error term $u^{(t)}$. This univariate formulation can be then extended to a multivariate VAR model including N time series $\mathbf{x}^{(t)} = [x_1^{(t)}, \ldots, x_N^{(t)}]$ (Luetkepohl, 2005):

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

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 $\mathbf{u}^{(t)} \in \mathbb{R}^N$, respectively. In our study, the multivariate time series $\mathbf{x}^{(t)}$ reflect the BOLD signal strength in the N brain regions, sampled at different timesteps t.

To estimate parameters of the VAR model, we used the ordinary least squares fit provided in the MVGC toolbox (Barnett & Seth, 2013). As outlined in the Data Description section, 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 the Model Accuracy and Network Scaling section, we tested the model order p in steps of five with $p = 5, 10, \ldots, T_p$ and chose the model with highest accuracy individually for each dataset. To check for stationarity of the signals, an augmented Dickey-Fuller test for unit roots was applied to the BOLD time courses (Hamilton, 1994; MacKinnon, 1994), 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 the Model Accuracy and Network Scaling section 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.

ACKNOWLEDGMENTS

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.

SUPPORTING INFORMATION

Supporting information for this article is available at https://doi.org/10.1162/netn_a_00252. A demo version for MRI data preparation and training the DCRNN model is provided at https://github.com/simonvino/DCRNN_brain_connectivity. In addition, a demo version for the GWN model is provided at https://github.com/simonvino/GraphWaveNet_brain_connectivity. Preprocessed HCP data is publicly available under: https://db.humanconnectome.org.

AUTHOR CONTRIBUTIONS

Simon Wein: Conceptualization; Investigation; Methodology; Writing – original draft. Alina Schüller: Investigation; Writing – review & editing. Ana Maria Tomé: Validation; Writing – review & editing. Wilhelm M. Malloni: Validation; Writing – review & editing. Mark W. Greenlee: Supervision; Writing – review & editing. Elmar W. Lang: Methodology; Supervision; Writing – original draft; Writing – review & editing.

FUNDING INFORMATION

Mark W. Greenlee, Deutsche Forschungsgemeinschaft (https://dx.doi.org/10.13039/501100001659), Award ID: GR988/25-1. Mark W. Greenlee, Deutsche Forschungsgemeinschaft (https://dx.doi.org/10.13039/501100001659), Award ID: ISNT89/393-1.

REFERENCES

Abdelnour, F., Dayan, M., Devinsky, O., Thesen, T., & Raj, A. (2018). Functional brain connectivity is predictable from anatomic network's Laplacian eigen-structure. *NeuroImage*, *172*, 728–739. https://doi.org/10.1016/j.neuroimage.2018.02.016, PubMed: 29454104

Abreu, R., Leal, A., & Figueiredo, P. (2018). EEG-informed fMRI: A review of data analysis methods. *Frontiers in Human Neuroscience*, *12*, 29. https://doi.org/10.3389/fnhum.2018.00029, PubMed: 29467634

Achard, S., Salvador, R., Whitcher, B., Suckling, J., & Bullmore, E. (2006). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *Journal of Neuroscience*, *26*, 63–72. https://doi.org/10.1523/JNEUROSCI.3874-05.2006, PubMed: 16399673

Alstott, J., Breakspear, M., Hagmann, P., Cammoun, L., & Sporns, O. (2009). Modeling the impact of lesions in the human brain. *PLoS Computational Biology*, *5*, e1000408. https://doi.org/10.1371/journal.pcbi.1000408, PubMed: 19521503

Amico, E., & Goñi, J. (2018). Mapping hybrid functional-structural connectivity traits in the human connectome. *Network Neuroscience*, *2*, 306–322. https://doi.org/10.1162/netn_a_00049, PubMed: 30259007

Andersson, J., Skare, S., & Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *NeuroImage*, *20*, 870–888. https://doi.org/10.1016/S1053-8119(03)00336-7, PubMed: 14568458

Andersson, J., & Sotiropoulos, S. (2015a). An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage*, *125*, 1063–1078. https://doi.org/10.1016/j.neuroimage.2015.10.019, PubMed: 26481672

Andersson, J., & Sotiropoulos, S. (2015b). Non-parametric representation and prediction of single- and multi-shell diffusion-weighted MRI data using gaussian processes. *NeuroImage*, *122*, 166–176. https://doi.org/10.1016/j.neuroimage.2015.07.067, PubMed: 26236030

Arslan, S., Ktena, S. I., Glocker, B., & Rueckert, D. (2018). Graph saliency maps through spectral convolutional networks: Application to sex classification with brain connectivity. *arXiv*, arXiv:1806.01764. https://doi.org/10.1007/978-3-030-00689-1_1

Barnett, L., & Seth, A. (2013). The MVGC multivariate granger causality toolbox: A new approach to granger-causal inference. *Journal of Neuroscience Methods*, *223*, 50–68. https://doi.org/10.1016/j.jneumeth.2013.10.018, PubMed: 24200508

Becker, C. O., Pequito, S., Pappas, G. J., Miller, M. B., Grafton, S. T., Bassett, D. S., & Preciado, V. M. (2018). Spectral mapping of brain functional connectivity from diffusion imaging. *Scientific Reports*, *8*, 1411. https://doi.org/10.1038/s41598-017-18769-x, PubMed: 29362436

Bengio, S., Vinyals, O., Jaitly, N., & Shazeer, N. (2015). Scheduled sampling for sequence prediction with recurrent neural networks. *arXiv*, arXiv:1506.03099. https://doi.org/10.48550/arXiv.1506.03099

Bettinardi, R. G., Deco, G., Karlaftis, V. M., Hartevelt, T. J. V, Fernandes, H. M., Kourtzi, Z., … Zamora-López, G. (2018). 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*, 047409. https://doi.org/10.1063/1.4980099, PubMed: 28456160

Bielczyk, N., Uithol, S., van Mourik, T., Anderson, P., Glennon, J., & Buitelaar, J. (2019). Disentangling causal webs in the brain using functional magnetic resonance imaging: A review of current approaches. *Network Neuroscience*, *3*, 237–273. https://doi.org/10.1162/netn_a_00062, PubMed: 30793082

Biswal, B., DeYoe, E. A., & Hyde, J. S. (1996). Reduction of physiological fluctuations in fMRI using digital filters. *Magnetic Resonance in Medicine*, *35*(1), 107–113. https://doi.org/10.1002/mrm.1910350114, PubMed: 8771028

Biswal, B. B., Yetkin, F. Z., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, *34*(4), 537–541. https://doi.org/10.1002/mrm.1910340409, PubMed: 8524021

Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A. D., & Vandergheynst, P. (2017). Geometric deep learning: Going beyond Euclidean data. *IEEE Signal Processing Magazine*, *34*, 18–42. https://doi.org/10.1109/MSP.2017.2693418

Brüel Gabrielsson, R. (2020). Universal function approximation on graphs. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, & H. Lin (Eds.), *Advances in neural information processing systems* (Vol. 33, pp. 19762–19772). Red Hook, NY: Curran Associates.

Buckner, R. L., Sepulcre, J., Talukdar, T., Krienen, F. M., Liu, H., Hedden, T., … Johnson, K. A. (2009). Cortical hubs revealed by intrinsic functional connectivity: Mapping, assessment of stability, and relation to Alzheimer's disease. *Journal of Neuroscience*, *29*, 1860–1873. https://doi.org/10.1523/JNEUROSCI.5062-08.2009, PubMed: 19211893

Bullmore, E. T., & Bassett, D. S. (2011). Brain graphs: Graphical models of the human brain connectome. *Annual Review of Clinical Psychology*, *7*, 113–140. https://doi.org/10.1146/annurev-clinpsy-040510-143934, PubMed: 21128784

Burgess, G. C., Kandala, S., Nolan, D., Laumann, T. O., Power, J. D., Adeyemo, B., … Barch, D. M. (2016). Evaluation of denoising strategies to address motion-correlated artifacts in resting-state fMRI data from the Human Connectome Project. *Brain Connectivity*, *6*, 669–680. https://doi.org/10.1089/brain.2016.0435, PubMed: 27571276

Bush, K., Cisler, J., Bian, J., Hazaroglu, G., Hazaroglu, O., & Kilts, C. (2015). Improving the precision of fMRI BOLD signal deconvolution with implications for connectivity analysis. *Magnetic Resonance Imaging*, *33*, 1314–1323. https://doi.org/10.1016/j.mri.2015.07.007, PubMed: 26226647

Chen, X., & Wang, Y. (2018). Predicting resting-state functional connectivity with efficient structural connectivity. *EEE/CAA Journal of Automatica Sinica*, *5*(6), 1079–1088. https://doi.org/10.1109/JAS.2017.7510880

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv*, arXiv:1412.3555. https://doi.org/10.48550/arXiv.1412.3555

Ciric, R., Wolf, D. H., Power, J. D., Roalf, D. R., Baum, G. L., Ruparel, K., … Satterthwaite, T. D. (2017). Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *NeuroImage*, *154*, 174–187. https://doi.org/10.1016/j.neuroimage.2017.03.020, PubMed: 28302591

Cole, M. W., Ito, T., Bassett, D. S., & Schultz, D. H. (2016). Activity flow over resting-state networks shapes cognitive task activations. *Nature Neuroscience*, *19*, 1718–1726. https://doi.org/10.1038/nn.4406, PubMed: 27723746

Dauphin, Y., Pascanu, R., Gulcehre, C., Cho, K., Ganguli, S., & Bengio, Y. (2014). Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. *arXiv*, arXiv:1406.2572. https://doi.org/10.48550/arXiv.1406.2572

Deco, G., Kringelbach, M. L., Jirsa, V. K., & Ritter, P. (2017). The dynamics of resting fluctuations in the brain: Metastability and its dynamical cortical core. *Scientific Reports*, *7*, 3095. https://doi.org/10.1038/s41598-017-03073-5, PubMed: 28596608

Deco, G., Ponce-Alvarez, A., Mantini, D., Romani, G.-L., Hagmann, P., & Corbetta, M. (2013). Resting-state functional connectivity emerges from structurally and dynamically shaped slow linear fluctuations. *Journal of Neuroscience*, *33*, 11239–11252. https://doi.org/10.1523/JNEUROSCI.1091-13.2013, PubMed: 23825427

Deco, G., Senden, M., & Jirsa, V. (2012). How anatomy shapes dynamics: A semi-analytical study of the brain at rest by a simple spin model. *Frontiers in Computational Neuroscience*, *6*, 68. https://doi.org/10.3389/fncom.2012.00068, PubMed: 23024632

Defferrard, M., Bresson, X., & Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. *arXiv*, arXiv:1606.09375. https://doi.org/10.48550/arXiv.1606.09375

de Haan, P., Cohen, T. S., & Welling, M. (2020). Natural graph networks. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), *Advances in neural information processing systems* (Vol. 33, pp. 3636–3646). Red Hook, NY: Curran Associates.

Deligianni, F., Carmichael, D. W., Zhang, G. H., Clark, C. A., & Clayden, J. D. (2016). NODDI and tensor-based microstructural indices as predictors of functional connectivity. *PLoS One*, *11*, e0153404. https://doi.org/10.1371/journal.pone.0153404, PubMed: 27078862

Demirtaş, M., Burt, J. B., Helmer, M., Ji, J. L., Adkinson, B. D., Glasser, M. F., … Murray, J. D. (2019). Hierarchical heterogeneity

across human cortex shapes large-scale neural dynamics. *Neuron*, *101*, 1181–1194. https://doi.org/10.1016/j.neuron.2019.01.017, PubMed: 30744986

Duggento, A., Passamonti, L., Valenza, G., Barbieri, R., Guerrisi, M., & Toschi, N. (2018). Multivariate Granger causality unveils directed parietal to prefrontal cortex connectivity during task-free MRI. *Scientific Reports*, *8*, 5571. https://doi.org/10.1038/s41598-018-23996-x, PubMed: 29615790

Feinberg, D. A., Moeller, S., Smith, S. M., Auerbach, E., Ramanna, S., Gunther, M., ... Yacoub, E. (2010). Multiplexed echo planar imaging for sub-second whole brain fMRI and fast diffusion imaging. *PLoS One*, *5*, e15710. https://doi.org/10.1371/journal.pone.0015710, PubMed: 21187930

Fischl, B. (2012). FreeSurfer. *NeuroImage*, *62*(2), 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021, PubMed: 22248573

Friston, K., Moran, R., & Seth, A. K. (2013). Analysing connectivity with Granger causality and dynamic causal modelling. *Current Opinion in Neurobiology*, *23*, 172–178. https://doi.org/10.1016/j.conb.2012.11.010, PubMed: 23265964

Frässle, S., Lomakina, E. I., Kasper, L., Manjaly, Z. M., Leff, A., Pruessmann, K. P., ... Stephan, K. E. (2018). A generative model of whole-brain effective connectivity. *NeuroImage*, *179*, 505–529. https://doi.org/10.1016/j.neuroimage.2018.05.058, PubMed: 29807151

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., ... Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*, 171–178. https://doi.org/10.1038/nature18933, PubMed: 27437579

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., ... Jenkinson, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, *80*, 105–124. https://doi.org/10.1016/j.neuroimage.2013.04.127, PubMed: 23668970

Glerean, E., Salmi, J., Lahnakoski, J. M., Jääskeläinen, I. P., & Sams, M. (2012). Functional magnetic resonance imaging phase synchronization as a measure of dynamic functional connectivity. *Brain Connectivity*, *2*, 91–101. https://doi.org/10.1089/brain.2011.0068, PubMed: 22559794

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*, 424–438. https://doi.org/10.2307/1912791

Griffanti, L., Salimi-Khorshidi, G., Beckmann, C. F., Auerbach, E. J., Douaud, G., Sexton, C. E., ... Smith, S. M. (2014). ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *NeuroImage*, *95*, 232–247. https://doi.org/10.1016/j.neuroimage.2014.03.034, PubMed: 24657355

Grover, A., & Leskovec, J. (2016). node2vec: Scalable feature learning for networks. In *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 855–864). https://doi.org/10.1145/2939672.2939754, PubMed: 27853626

Hamilton, J. (1994). *Time series analysis*. Princeton, NJ.: Princeton University Press. https://doi.org/10.1515/9780691218632

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). Las Vegas, NV: IEEE. https://doi.org/10.1109/CVPR.2016.90

Hodge, M. R., Horton, W., Brown, T., Herrick, R., Olsen, T., Hileman, M. E., ... Marcus, D. S. (2015). ConnectomeDB—Sharing human brain connectivity data. *NeuroImage*, *124*, 1102–1107. https://doi.org/10.1016/j.neuroimage.2015.04.046, PubMed: 25934470

Honey, C. J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J. P., Meuli, R., & Hagmann, P. (2009). Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(6), 2035–2040. https://doi.org/10.1073/pnas.0811168106, PubMed: 19188601

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359–366. https://doi.org/10.1016/0893-6080(89)90020-8

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In F. Bach & D. Blei (Eds.), *Proceedings of the 32nd International Conference on Machine Learning* (Vol. 37, pp. 448–456). Lille, France: PMLR.

Ito, T., Hearne, L., Mill, R., Cocuzza, C., & Cole, M. W. (2020). Discovering the computational relevance of brain network organization. *Trends in Cognitive Sciences*, *24*(1), 25–38. https://doi.org/10.1016/j.tics.2019.10.005, PubMed: 31727507

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*, 825–841. https://doi.org/10.1006/nimg.2002.1132, PubMed: 12377157

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, *62*(2), 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015, PubMed: 21979382

Jeurissen, B., Tournier, J.-D., Dhollander, T., Connelly, A., & Sijbers, J. (2014). Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. *NeuroImage*, *103*, 411–426. https://doi.org/10.1016/j.neuroimage.2014.07.061, PubMed: 25109526

Kim, B.-H., & Ye, J. C. (2020). Understanding graph isomorphism network for rs-fMRI functional connectivity analysis. *Frontiers in Neuroscience*, *14*, 630. https://doi.org/10.3389/fnins.2020.00630, PubMed: 32714130

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv*, arXiv:1412.6980. https://doi.org/10.48550/arXiv.1412.6980

Ktena, S. I., Parisot, S., Ferrante, E., Rajchl, M., Lee, M., Glocker, B., & Rueckert, D. (2018). Metric learning with spectral graph convolutions on brain connectivity networks. *NeuroImage*, *169*, 431–442. https://doi.org/10.1016/j.neuroimage.2017.12.052, PubMed: 29278772

Lang, E. W., Tomé, A. M., Keck, I. R., Górriz-Sáez, J. M., & Puntonet, C. G. (2012). Brain connectivity analysis: A short survey. *Computational Intelligence and Neuroscience*, *2012*, 412512. https://doi.org/10.1155/2012/412512, PubMed: 23097663

Li, X., Dvornek, N. C., Zhou, Y., Zhuang, J., Ventola, P., & Duncan, J. S. (2019). Graph neural network for interpreting task-fMRI biomarkers. *arXiv*, arXiv:1907.01661. https://doi.org/10.48550/arXiv.1907.01661

Li, Y., Yu, R., Shahabi, C., & Liu, Y. (2018). Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv*, arXiv:1707.01926. https://doi.org/10.48550/arXiv.1707.01926

Liang, H., & Wang, H. (2017). Structure-function network mapping and its assessment via persistent homology. *PLoS Computational Biology*, *13*, e1005325. https://doi.org/10.1371/journal.pcbi.1005325, PubMed: 28046127

Liang, J., Jiang, L., Cao, L., Kalantidis, Y., Li, L.-J., & Hauptmann, A. G. (2019). Focal visual-text attention for Memex question answering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(8), 1893–1908. https://doi.org/10.1109/TPAMI.2018.2890628, PubMed: 30624212

Lim, S., Radicchi, F., van den Heuvel, M. P., & Sporns, O. (2019). Discordant attributes of structural and functional brain connectivity in a two-layer multiplex network. *Scientific Reports*, *9*, 2885. https://doi.org/10.1038/s41598-019-39243-w, PubMed: 30814615

Luetkepohl, H. (2005). *The new introduction to multiple time series analysis*. Springer. https://doi.org/10.1007/978-3-540-27752-1

MacKinnon, J. G. (1994). Approximate asymptotic distribution functions for unit-root and cointegration tests. *Journal of Business and Economic Statistics*, *12*, 167–176. https://doi.org/10.1080/07350015.1994.10510005

Mele, G., Cavaliere, C., Alfano, V., Orsini, M., Salvatore, M., & Aiello, M. (2019). Simultaneous EEG-fMRI for functional neurological assessment. *Frontiers in Neurology*, *10*, 848. https://doi.org/10.3389/fneur.2019.00848, PubMed: 31456735

Messé, A., Hütt, M.-T., König, P., & Hilgetag, C. C. (2015). A closer look at the apparent correlation of structural and functional connectivity in excitable neural networks. *Scientific Reports*, *5*, 7870. https://doi.org/10.1038/srep07870, PubMed: 25598302

Messé, A., Rudrauf, D., Benali, H., & Marrelec, G. (2014). Relating structure and function in the human brain: Relative contributions of anatomy, stationary dynamics, and non-stationarities. *PLoS Computational Biology*, *10*(3), e1003530. https://doi.org/10.1371/journal.pcbi.1003530, PubMed: 24651524

Messé, A., Rudrauf, D., Giron, A., & Marrelec, G. (2015). Predicting functional connectivity from structural connectivity via computational models using MRI: An extensive comparison study. *NeuroImage*, *111*, 65–75. https://doi.org/10.1016/j.neuroimage.2015.02.001, PubMed: 25682944

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, *26*, https://doi.org/10.48550/arXiv.1310.4546

Mill, R. D., Bagic, A., Bostan, A., Schneider, W., & Cole, M. W. (2017). Empirical validation of directed functional connectivity. *NeuroImage*, *146*, 275–287. https://doi.org/10.1016/j.neuroimage.2016.11.037, PubMed: 27856312

Moeller, S., Yacoub, E., Olman, C. A., Auerbach, E., Strupp, J., Harel, N., & Uğurbil, K. (2010). 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), 1144–1153. https://doi.org/10.1002/mrm.22361, PubMed: 20432285

Olman, C. A., Davachi, L., & Inati, S. (2009). Distortion and signal loss in medial temporal lobe. *PLoS One*, *4*, e8160. https://doi.org/10.1371/journal.pone.0008160, PubMed: 19997633

Panda, R., Thibaut, A., Lopez-Gonzalez, A., Escrichs, A., Bahri, M. A., Hillebrand, A., ... Tewarie, P. (2021). Disruption in structural-functional network repertoire and time-resolved subcortical-frontoparietal connectivity in disorders of consciousness. *bioRxiv*. https://doi.org/10.1101/2021.12.10.472068

Power, J. D., Plitt, M., Laumann, T. O., & Martin, A. (2017). Sources and implications of whole-brain fMRI signals in humans. *NeuroImage*, *146*, 609–625. https://doi.org/10.1016/j.neuroimage.2016.09.038, PubMed: 27751941

Prando, G., Zorzi, M., Bertoldo, A., Corbetta, M., Zorzi, M., & Chiuso, A. (2020). Sparse DCM for whole-brain effective connectivity from resting-state fMRI data. *NeuroImage*, *208*, 116367. https://doi.org/10.1016/j.neuroimage.2019.116367, PubMed: 31812714

Ramsey, J. D., Hanson, S. J., & Glymour, C. (2011). Multi-subject search correctly identifies causal connections and most causal directions in the DCM models of the Smith et al. simulation study. *NeuroImage*, *58*(3), 838–848. https://doi.org/10.1016/j.neuroimage.2011.06.068, PubMed: 21745580

Řehůřek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* (pp. 46–50). Valletta, Malta: University of Malta. https://doi.org/10.13140/2.1.2393.184

Reid, A. T., Headley, D. B., Mill, R. D., Sanchez-Romero, R., Uddin, L. Q., Marinazzo, D., ... Cole, M. W. (2019). Advancing functional connectivity research from association to causation. *Nature Neuroscience*, *22*(11), 1751–1760. https://doi.org/10.1038/s41593-019-0510-4, PubMed: 31611705

Roebroeck, A., Formisano, E., & Goebel, R. (2011). The identification of interacting networks in the brain using fMRI: Model selection, causality and deconvolution. *NeuroImage*, *58*, 296–302. https://doi.org/10.1016/j.neuroimage.2009.09.036, PubMed: 19786106

Rosenthal, G., Váša, F., Griffa, A., Hagmann, P., Amico, E., Goñi, J., ... Sporns, O. (2018). Mapping higher-order relations between brain structure and function with embedded vector representations of connectomes. *Nature Communications*, *9*(1), 2178. https://doi.org/10.1038/s41467-018-04614-w, PubMed: 29872218

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*, 533–536. https://doi.org/10.1038/323533a0

Salimi-Khorshidi, G., Douaud, G., Beckmann, C. F., Glasser, M. F., Griffanti, L., & Smith, S. M. (2014). Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage*, *90*, 449–468. https://doi.org/10.1016/j.neuroimage.2013.11.046, PubMed: 24389422

Sarwar, T., Tian, Y., Yeo, B. T. T., Ramamohanarao, K., & Zalesky, A. (2021). Structure-function coupling in the human connectome: A machine learning approach. *NeuroImage*, *226*, 117609. https://doi.org/10.1016/j.neuroimage.2020.117609, PubMed: 33271268

Schnake, T., Eberle, O., Lederer, J., Nakajima, S., Schütt, K. T., Müller, K.-R., & Montavon, G. (in press). Higher-order explanations of graph neural networks via relevant walks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Seo, Y., Defferrard, M., Vandergheynst, P., & Bresson, X. (2018). Structured sequence modeling with graph convolution recurrent

networks. In L. Cheng, A. C. S Leung, & S. Ozawa (Eds.), *Neural information processing* (pp. 362–373). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-030-04167-0_33

Seth, A. K., Chorley, P., & Barnett, L. C. (2012). Granger causality analysis of fMRI BOLD signals is invariant to hemodynamic convolution but not downsampling. *NeuroImage*, *65*, 540–555. https://doi.org/10.1016/j.neuroimage.2012.09.049, PubMed: 23036449

Setsompop, K., Gagoski, B. A., Polimeni, J. R., Witzel, T., Wedeen, V. J., & Wald, L. L. (2012). Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magnetic Resonance in Medicine*, *67*(5), 1210–1224. https://doi.org/10.1002/mrm.23097, PubMed: 21858868

Shinn, M., Hu, A., Turner, L., Noble, S., Achard, S., Anticevic, A., … Murray, J. D. (2021). Spatial and temporal autocorrelation weave human brain networks. *bioRxiv*. https://doi.org/10.1101/2021.06.01.446561

Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A., & Vandergheynst, P. (2013). The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, *30*(3), 83–98. https://doi.org/10.1109/MSP.2012.2235192

Singh, M. F., Braver, T. S., Cole, M. W., & Ching, S. (2020). Estimation and validation of individualized dynamic brain models with resting state fMRI. *NeuroImage*, *221*, 117046. https://doi.org/10.1016/j.neuroimage.2020.117046, PubMed: 32603858

Smith, A. M., Lewis, B. K., Ruttimann, U. E., Ye, F. Q., Sinnwell, T. M., Yang, Y., … Frank, J. A. (1999). Investigation of low frequency drift in fMRI signal. *NeuroImage*, *9*, 526–533. https://doi.org/10.1006/nimg.1999.0435, PubMed: 10329292

Smith, R. E., Tournier, J.-D., Calamante, F., & Connelly, A. (2012). Anatomically-constrained tractography: Improved diffusion MRI streamlines tractography through effective use of anatomical information. *NeuroImage*, *62*(3), 1924–1938. https://doi.org/10.1016/j.neuroimage.2012.06.005, PubMed: 22705374

Smith, R. E., Tournier, J.-D., Calamante, F., & Connelly, A. (2013). Sift: Spherical-deconvolution informed filtering of tractograms. *NeuroImage*, *67*, 298–312. https://doi.org/10.1016/j.neuroimage.2012.11.049, PubMed: 23238430

Smith, S. M., Beckmann, C. F., Andersson, J., Auerbach, E. J., Bijsterbosch, J., Douaud, G., … Glasser, M. F. (2013). Resting-state fMRI in the Human Connectome Project. *NeuroImage*, *80*, 144–168. https://doi.org/10.1016/j.neuroimage.2013.05.039, PubMed: 23702415

Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., … Woolrich, M. W. (2011). Network modelling methods for fMRI. *NeuroImage*, *54*(2), 875–891. https://doi.org/10.1016/j.neuroimage.2010.08.063, PubMed: 20817103

Sotiropoulos, S. N., Jbabdi, S., Xu, J., Andersson, J. L., Moeller, S., Auerbach, E. J., … Behrens, T. E. J. (2013). Advances in diffusion MRI acquisition and processing in the Human Connectome Project. *NeuroImage*, *80*, 125–143. https://doi.org/10.1016/j.neuroimage.2013.05.057, PubMed: 23702418

Sotiropoulos, S. N., Moeller, S., Jbabdi, S., Xu, J., Andersson, J. L., Auerbach, E. J., … Lenglet, C. (2013). Effects of image reconstruction on fibre orientation mapping from multi-channel diffusion MRI: Reducing the noise floor using SENSE. *Magnetic Resonance in Medicine*, *70*, 1682–1689. https://doi.org/10.1002/mrm.24623, PubMed: 23401137

Suárez, L. E., Richards, B. A., Lajoie, G., & Misic, B. (2021). Learning function from structure in neuromorphic networks. *Nature Machine Intelligence*, *3*, 771–786. https://doi.org/10.1038/s42256-021-00376-1

Suárez, L. E., Markello, R. D., Betzel, R. F., & Misic, B. (2020). Linking structure and function in macroscale brain networks. *Trends in Cognitive Sciences*, *24*, 302–315. https://doi.org/10.1016/j.tics.2020.01.008, PubMed: 32160567

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *CoRR. abs/1409.3215*. https://doi.org/10.48550/arXiv.1409.3215

Thomas, C., Ye, F. Q., Irfanoglu, M. O., Modi, P., Saleem, K. S., Leopold, D. A., & Pierpaoli, C. (2014). 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), 16574–16579. https://doi.org/10.1073/pnas.1405672111, PubMed: 25368179

Tournier, J.-D., Calamante, F., & Connelly, A. (2007). Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution. *NeuroImage*, *35*(4), 1459–1472. https://doi.org/10.1016/j.neuroimage.2007.02.016, PubMed: 17379540

Tournier, J.-D., Calamante, F., Gadian, D. G., & Connelly, A. (2004). Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage*, *23*, 1176–1185. https://doi.org/10.1016/j.neuroimage.2004.07.037, PubMed: 15528117

Tournier, J.-D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M., … Connelly, A. (2019). MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *NeuroImage*, *202*, 116137. https://doi.org/10.1016/j.neuroimage.2019.116137, PubMed: 31473352

Uğurbil, K., Xu, J., Auerbach, E. J., Moeller, S., Vu, A. T., Duarte-Carvajalino, J. M., … Yacoub, E. (2013). Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. *NeuroImage*, *80*, 80–104. https://doi.org/10.1016/j.neuroimage.2013.05.012, PubMed: 23702417

van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., … Kavukcuoglu, K. (2016). WaveNet: A generative model for raw audio. *arXiv*, arXiv:1609.03499. https://doi.org/10.48550/arXiv.1609.03499

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., & Ugurbil, K. (2013). The WU-Minn Human Connectome Project: An overview. *NeuroImage*, *80*, 62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041, PubMed: 23684880

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). Attention is all you need. In I. Guyon et al. (Eds.), *Advances in neural information processing systems* (Vol. 30). Red Hook, NY: Curran Associates.

Vézquez-Rodríguez, B., Liu, Z.-Q., Hagmann, P., & Misic, B. (2020). Signal propagation via cortical hierarchies. *Network Neuroscience*, *4*(4), 1072–1090. https://doi.org/10.1162/netn_a_00153, PubMed: 33195949

Wang, H. E., Bénar, C. G., Quilichini, P. P., Friston, K. J., Jirsa, V. K., & Bernard, C. (2014). A systematic framework for functional

connectivity measures. *Frontiers in Neuroscience*, *8*, 405. https://doi.org/10.3389/fnins.2014.00405, PubMed: 25538556

Wang, P., Kong, R., Kong, X., Liégeois, R., Orban, C., Deco, G., ... Yeo, B. T. T. (2019). Inversion of a large-scale circuit model reveals a cortical hierarchy in the dynamic resting human brain. *Science Advances*, *5*(1), eaat7854. https://doi.org/10.1126/sciadv.aat7854, PubMed: 30662942

Webb, J. T., Ferguson, M. A., Nielsen, J. A., & Anderson, J. S. (2013). BOLD Granger causality reflects vascular anatomy. *PLoS One*, *8*(12), e84279. https://doi.org/10.1371/journal.pone.0084279, PubMed: 24349569

Wein, S., Deco, G., Tomé, A. M., Goldhacker, M., Malloni, W. M., Greenlee, M. W., & Lang, E. W. (2021). Brain connectivity studies on structure-function relationships: A short survey with an emphasis on machine learning. *Computational Intelligence and Neuroscience*, *2021*, 5573740. https://doi.org/10.1155/2021/5573740, PubMed: 34135951

Wein, S., Malloni, W. M., Tomé, A. M., Frank, S. M., Henze, G.-I., Wüst, S., ... Lang, E. W. (2021). A graph neural network framework for causal inference in brain networks. *Scientific Reports*, *11*, 8061. https://doi.org/10.1038/s41598-021-87411-8, PubMed: 33850173

Wen, X., Rangarajan, G., & Ding, M. (2013). Is Granger causality a viable technique for analyzing fMRI data? *PLoS One*, *8*(7), e67428. https://doi.org/10.1371/journal.pone.0067428, PubMed: 23861763

Werbos, P. J. (1990). Backpropagation through time: What it does and how to do it. *Proceedings of the IEEE*, *78*(10), 1550–1560. https://doi.org/10.1109/5.58337

Wise, R. G., Ide, K., Poulin, M. J., & Tracey, I. (2004). Resting fluctuations in arterial carbon dioxide induce significant low frequency variations in BOLD signal. *NeuroImage*, *21*, 1652–1664. https://doi.org/10.1016/j.neuroimage.2003.11.025, PubMed: 15050588

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2021). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, *32*, 4–24. https://doi.org/10.1109/TNNLS.2020.2978386, PubMed: 32217482

Wu, Z., Pan, S., Long, G., Jiang, J., & Zhang, C. (2019). Graph WaveNet for deep spatial-temporal graph modeling. *Proceedings of the 28th International Joint Conference on Artificial Intelligence* (pp. 1907–1913). AAAI Press. https://doi.org/10.24963/ijcai.2019/264

Xu, J., Moeller, S., Strupp, J., Auerbach, E. J., Chen, L., Feinberg, D. A., ... Yacoub, E. (2012). Highly accelerated whole brain imaging using aligned-blipped-controlled-aliasing multiband EPI. *Proceedings of the 20th Annual Meeting of ISMRM*, *2306*, 1907–1913.

Xu, K., Ba, J. L., Kiros, R., Cho, K., Courville, A., Salakhutdinov, R., ... Bengio, Y. (2015). Show, attend and tell: Neural image caption generation with visual attention. In *Proceedings of the 32nd International Conference on Machine Learning* (Vol. 37, pp. 2048–2057). Lille, France: PMLR.

Yan, T., Liu, T., Ai, J., Shi, Z., Zhang, J., Pei, G., & Wu, J. (2021). Task-induced activation transmitted by structural connectivity is associated with behavioral performance. *Brain Structure and Function*, *226*(5), 1437–1452. https://doi.org/10.1007/s00429-021-02249-0, PubMed: 33743076

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional neural networks. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *European Conference on Computer Vision 2014: Computer Vision – ECCV 2014* (Vol. 8689, pp. 818–833). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-10590-1_53

Zhang, H., Schneider, T., Wheeler-Kingshott, C. A., & Alexander, D. C. (2012). NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain. *NeuroImage*, *61*(4), 1000–1016. https://doi.org/10.1016/j.neuroimage.2012.03.072, PubMed: 22484410

Zheng, C., Fan, X., Wang, C., & Qi, J. (2020). GMAN: A graph multi-attention network for traffic prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*, 1234–1241. https://doi.org/10.1609/aaai.v34i01.5477

Zimmermann, J., Griffiths, J., Schirner, M., Ritter, P., & McIntosh, A. R. (2018). Subject specificity of the correlation between large-scale structural and functional connectivity. *Network Neuroscience*, *3*(1), 90–106. https://doi.org/10.1162/netn_a_00055, PubMed: 30793075

Zou, Q.-H., Zhu, C.-Z., Yang, Y., Zuo, X.-N., Long, X.-Y., Cao, Q.-J., ... Zang, Y.-F. (2008). An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: Fractional ALFF. *Journal of Neuroscience Methods*, *172*(1), 137–141. https://doi.org/10.1016/j.jneumeth.2008.04.012, PubMed: 18501969