Title: Deep learning models of cognitive processes constrained by human brain connectomes

Authors: Yu Zhang [1,2,3,*], Nicolas Farrugia[4] and Pierre Bellec [2,3,*]

[1] Research Center for Healthcare Data Science, Zhejiang Lab, Hangzhou, China

[2] Centre de recherche de l'Institut universitaire de gériatrie de Montréal, Montréal, QC H3W 1W6, Canada

[3] Department of Psychology, Université de Montréal, Montréal, QC H3C 3J7, Canada

[4] Department of Mathematical and Electrical Engineering, IMT Atlantique, Brest, France

*** Corresponding Authors:**

Yu Zhang

Research Center for Healthcare Data Science, Zhejiang Lab

Zhongtai Street, Yuhang District, Hangzhou 311100, Zhejiang, China

yuzhang2bic@gmail.com

Pierre Bellec

Département de Psychologie, Université de Montréal

4565, Chemin Queen-Mary, Montréal (Québec) H3W 1W5

pierre.bellec@gmail.com

Abstract

Brain decoding aims to infer cognitive states from recordings of brain activity. Current literature has mainly focused on isolated brain regions engaged in specific experimental conditions, but ignored the integrative nature of cognitive processes recruiting distributed brain networks. To tackle this issue, we propose a connectome-based graph neural network to integrate distributed patterns of brain activity in a multiscale manner, ranging from localized brain regions, to a specific brain circuit/network and towards the full brain. We evaluate the decoding model using a large task-fMRI database from the human connectome project. By implementing connectomic constraints and multiscale interactions in deep graph convolutions, the model achieves high accuracy of decoding 21 cognitive states (93%, chancel level: 4.8%) and shows high robustness against adversarial attacks on the graph architecture. Our study bridges human connectomes with deep learning techniques and provides new avenues to study the underlying neural substrates of human cognition at scale.

Keywords: fMRI, brain decoding, cognitive functions, graph convolutional networks, human connectome, brain atlas

Highlights

- A full-brain integrative model is critical for cognitive decoding
- Human connectomes and long-range connections accelerate the propagation of activity
- ChebNet decoding is robust to random attacks on brain connectomes and regions

Introduction

Modern imaging techniques, such as functional magnetic resonance imaging (fMRI), provide an opportunity to map the neural substrates of cognition in-vivo, and to decode cognitive processes non-invasively. Brain decoding has been an active topic since Haxby and colleagues first proposed the idea of using fMRI brain responses to predict the category of visual stimuli presented to a subject [1]. Nowadays, a variety of computational models have been proposed in the field, including multi-voxel pattern recognition [2,3], linear regression models [4,5], as well as nonlinear models such as deep artificial neural networks [6,7]. These decoding studies have been mainly focused on the functional specialization of isolated brain regions under a small number of experimental tasks. Such brain decoders have gained many successes when tackling unimodal cognitive processes, in particular vision [8], and focusing on specific brain regions, for example in the ventral visual stream network [9]. However, it is still unclear how this approach generalizes towards high-order cognitive functions, featuring in distributed patterns of brain activity and integrated cognitive processes across large-scale brain networks [10–12]. In order to model such cognitive process, several key components are required in the decoding model include 1) a full-brain model that embeds both local and global information of brain activity [13,14], and 2) an integrative model that encodes complex forms of functional interactions during cognitive processes [15,16].

We aimed to tackle this issue by implementing a connectome-constrained graph neural network (GNN) architecture, which characterizes human connectomes as a static graph, embeds task-evoked brain activity as signals on the graph, and then models the spatiotemporal dynamics underlying cognitive processes through deep graph convolutions. Graph convolution provides a promising tool to propagate activity flow through brain connectomes and capture long-range interactions during cognitive processes. We have shown that GNN successfully decoded over a variety of cognitive domains and on a large population of healthy subjects, and achieved high-performing decoding on this benchmark [17]. However, the previous study used a simplified version of graph convolution that integrates neural activity within a limited portion of human connectomes, mostly within a specific brain network. It

remains unclear how the choices of connectome priors and the interactions at different scales impact large-scale cognition decoding.

Here, we investigate a more sophisticated form of graph convolution, namely ChebNet, which has the ability to propagate information over a relatively larger neighborhood on the graph and integrate neural activity in a multiscale manner, ranging from segregated brain activity from local areas (K=0), to information integration within the same brain circuit/network (K=1) as well as between multiple networks (K>1), and eventually towards the full brain. We propose a multi-domain decoding model using ChebNet graph convolutions and investigate how to integrate connectomic constraints in the decoding pipeline, i.e. the implementation of multiscale functional integration and the construction of a proper graph architecture. The connectomic constraints start with a brain parcellation, which divides the whole brain into hundreds of brain regions, and a brain graph that captures hierarchical and modular structures in brain organization. A variety of parcellation schemes have been proposed in the literature, see the review paper by [18], which subdivide the entire cortex into non-overlapping regions based on connectivity profiles derived from diffusion tractography [19], functional organization [20], or multimodal imaging features [21]. As the abstract representation of human connectome, a brain graph captures the network organization of the brain structure and function, by using anatomical and functional connectivity, in both healthy and diseased populations [22–24]. These are the two key components in the graph architecture, which define the nature of interactions in graph convolutions. On the other hand, the scale of interactions is specified by the path length of information propagation, i.e. K-order in ChebNet. As a result, the model learns a new representation of cognitive processes after integrating task-evoked brain activity at different levels. This representation is then used to predict cognitive states.

We evaluate the ChebNet decoding model on a large population of 1200 participants, under 21 different experimental conditions, acquired from the task-fMRI database from the Human Connectome Project (HCP). We explore the optimal choices of functional integration and graph architectures on this decoding benchmark, including the resolution or homogeneity of brain parcels (nodes), and the type of interactions (edges), the inclusion of multiple path lengths in the graph (graph convolutions). Moreover, we assess the robustness of brain decoding by introducing perturbations on

the graph architecture, for instance network misspecifications due to random rewiring and node attacks. Lastly, we visualize the salient features of brain decoding that highly contribute to the prediction of cognitive states and compare them to the findings in the neuroscience literature.

Results

Decoding cognitive states with fine temporal resolution and high accuracy

We proposed a multi-domain decoding pipeline based on ChebNet graph convolution (Figure **1**). The ChebNet decoding model was evaluated using the cognitive battery of HCP task-fMRI dataset acquired from 1200 healthy subjects. Using a 10-second window, the 21 cognitive states were identified with an average test accuracy of 93% (mean=93.43%, STD=0.44% by using 10 fold cross-validation with shuffle splits). The temporal resolution of the decoding model can go down to a single fMRI volume (720 ms of duration), with a prediction accuracy well above chance level (60%, chance level=4.8%). The accuracy of single-volume state annotation was highly improved (reaching 76%) after taking into account the delay effect of hemodynamic response function in BOLD signals, i.e. excluding fMRI volumes within 6s after the onset of each task trial from both training and test sets. We have also evaluated other baseline approaches including multi-class support vector machine classification (SVC) with linear and nonlinear kernels, random forest and a multilayer perceptron (MLP, consisting of two fully connected layers) on the same dataset. The results with a 10-second time window demonstrated that high-order ChebNet model outperformed all the other linear and nonlinear decoding models by a wide margin (see Table S2). ChebNet also showed a significant improvement over GCN which relies on first-order graph convolution.

Low misclassification rate within and between cognitive domains

In order to clarify the nature of errors made by the ChebNet brain decoder, we examined the confusion matrix on the test set, which indicates the proportion of true and false predictions given a cognitive task state or domain. When using a 10-second time window, the confusion matrix showed a clear block diagonal architecture (see Figure **2a**), where the majority of experimental tasks were accurately identified for all task conditions (e.g. cross-domain misclassification rate<2%). Mistakes were found in tasks with similar cognitive processes, for instance, between relational processing and pattern matching conditions, as well as 0back vs 2back WM tasks. After summarizing the confusion matrix according to the six cognitive domains (Figure **2b**), each cognitive domain was identified with

an accuracy greater than 95%. Among the six cognitive domains, the language tasks (story vs math) and motor tasks (left/right hand, left/right foot and tongue) were the most easily recognizable conditions by showing the highest precision and recall scores (average F1-score = 98% and 97%, respectively for two language conditions and five motor conditions). Even higher scores were achieved when decoding a smaller number of experimental conditions restricted to a specific cognitive domain (Figure 2-S1), i.e. using task-fMRI data exclusively from a single cognitive domain during model training and evaluation. For instance, the model achieved near-perfect decoding on other high-order cognitive functions, including working-memory (94.51%, classifying 8 conditions using 25s) and social cognition (96.58%, classifying 2 conditions using 23s), as opposed to 92.6% and 92.9% for the two domains when decoding on 30s of fMRI data [6].

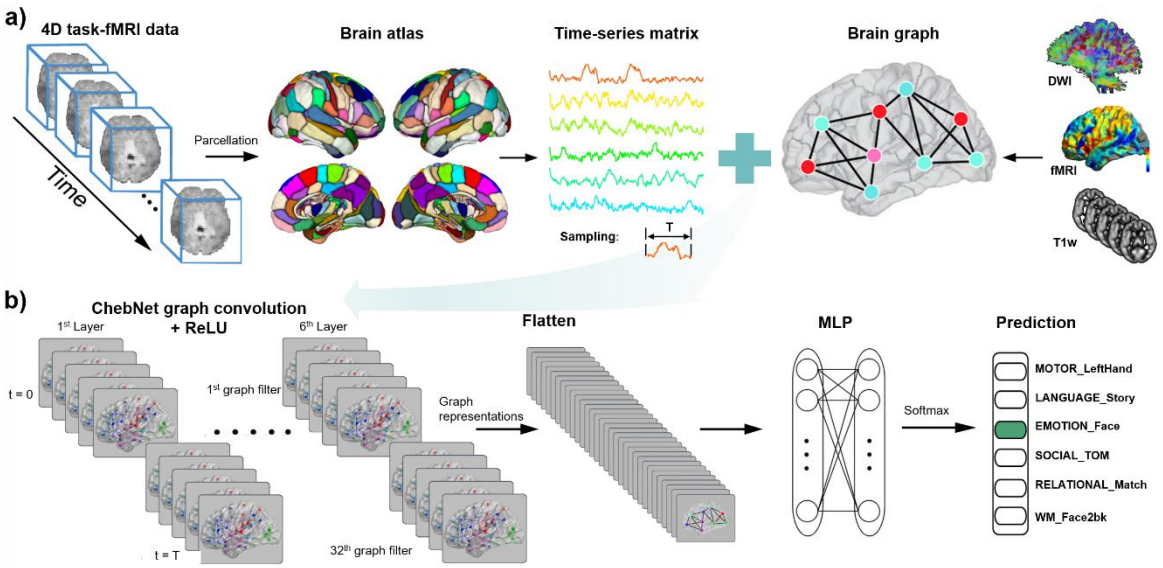

Figure 1. Pipeline of brain decoding using graph convolution network.

We used a similar network architecture as proposed in our previous paper [17] except for a more sophisticated form of graph convolution, namely ChebNet graph convolution, along with different brain graph architectures derived from various resolutions of brain atlas (nodes), and different types of brain connectivity (edges). The decoding model consists of six ChebNet graph convolutional layers with 32 graph filters at each layer, followed by a flatten layer and a two-layer Multilayer Perceptron (MLP, consisting of 256-64 units). Specifically, for a short series of fMRI volumes, the measured brain activity is first mapped onto a predefined brain atlas consisting of hundreds of brain regions (e.g.

246 regions from Brainnetome atlas [19]) and resulted in a 2D time-series matrix. Then, a brain graph indicating the edges between each pair of brain regions is constructed via either tractography of fiber projections using diffusion-weighted images (DWI), or functional correlation of low-frequency fluctuations in fMRI, or structural covariance of cortical thickness or gray matter density across a large population. Next, both the time-series matrix and brain graph are imported into the graph convolutional network (**b**). The model learns a new representation of task-evoked neural activity by stacking multiple graph convolutional layers, which takes into account both brain activity of each region (functional segregation) and the interactions within and between brain networks (functional integration). Finally, the learned graph representations are passing through a two-layer MLP and softmax function in order to predict the cognitive state associated with each input time window. The implementation of the ChebNet graph convolution was based on PyTorch 1.1.0, and was made publicly available in the following repository: https://github.com/zhangyu2ustc/gcn_tutorial_test.git .

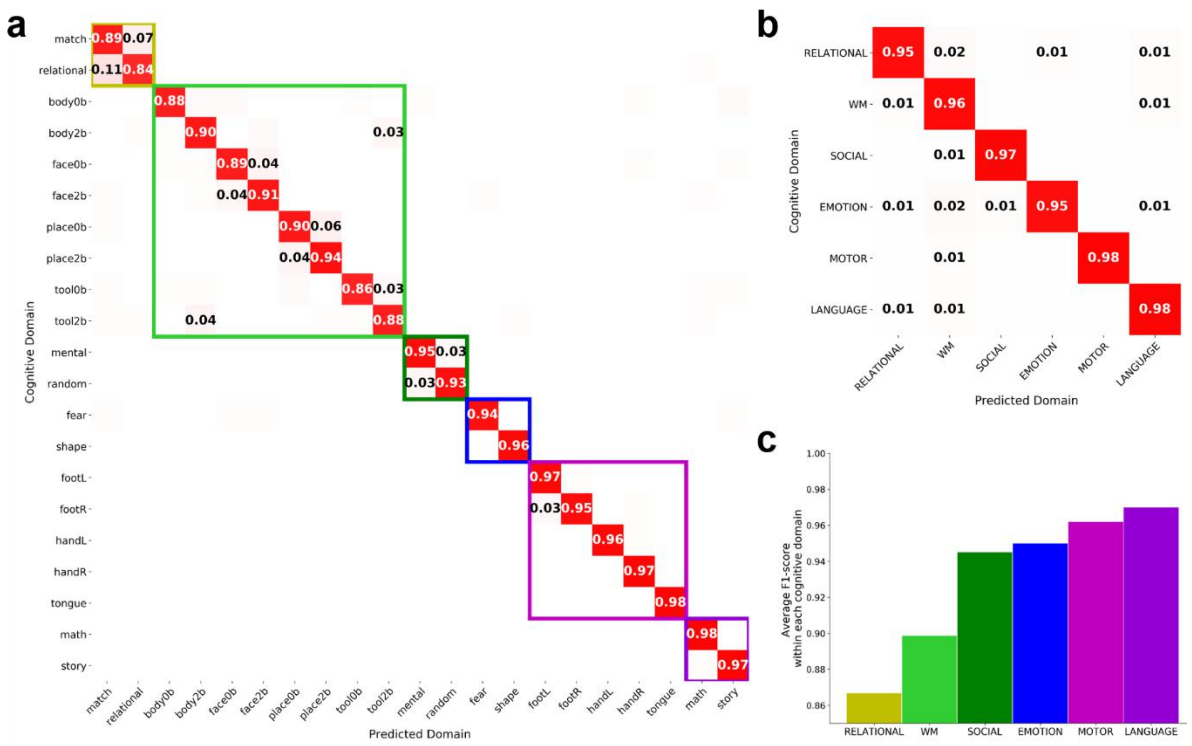

Figure 2. Confusion matrices of decoding 21 cognitive states and 6 cognitive domains.

We calculated the confusion matrix of cognitive decoding for the predicting cognitive states using every 10s of fMRI signals. We used the ChebNet-K5 model on a functional graph derived from Glasser's atlas in this analysis. The majority of misclassifications were found within the same

cognitive domain rather than between domains. A clear block diagonal architecture of the confusion matrix, which indicates the majority of the cognitive tasks or domains were accurately identified, was shown for both 21 task conditions (**a**) and 6 cognitive domains (**b**). For the visualization purpose, a threshold of 0.02 was applied to the confusion matrix in (**a**), which indicates a low misclassification rate between domains (<2%). The averaged F1-score for each domain was shown in (**c**), which was calculated by averaging the scores within each cognitive domain based on the number of samples. Among the six cognitive domains, the Language tasks (in dark violet) and Motor tasks (in magenta) were the most easily recognizable conditions. Note: the six cognitive domains include relational processing (RELATIONAL), working memory (WM), social cognition (SOCIAL), facial emotional processing (EMOTION), body movements (MOTOR), and language comprehension (LANGUAGE).

Order K in ChebNet controls the propagation rate on the graph

At each ChebNet graph convolutional layer, the context of brain activity is propagated within a *K*-step neighborhood on the graph. As illustrated in Figure 2-S1, both the choices of K-order and graph architectures significantly impact the propagation rate of information flow in the brain, estimated by the expected first arrival time. We first investigated the impact of *K*-order on the decoding model by spanning over the list of [0,1,2,5,10], where K=0 indicates a global scaling factor on the input features by treating each node independently; K=1 indicates the integration of brain activity for each node and its direct neighbors, while K>1 indicates the integration not limited to a region of interest or within a specific brain network, but instead expanding among inter-connected brain networks.

Our results (see Figure **3**) showed that the decoding performance gradually improved by increasing the K-order and reached a plateau when $K \geq 5$. For example, when using a functional graph (Figure **3a**), the ChebNet- $K1$ model (i.e. ChebNet with $K = 1$) showed lower decoding accuracy than high-order models (91.67% vs 93.22%, respectively for $K = 1$ and $K > 1$), but was substantially higher than the $K = 0$ model (83.72%). When $K \geq 5$, the training curves (indicated by the validation accuracy at the end of each training epoch) generally followed the same behaviors and the overall prediction accuracy plateaued approximately at the same level. We evaluated the sensitivity to the *K*-order for each cognitive domain, which demonstrated a strong domain-specific effect. For instance,

189 when using functional-graph (Figure **3e**), the decoding on the Motor and Language tasks showed little
190 impact by the *K*-order, which indicates that decoding on the two tasks was driven by functional
191 interactions within the same functional network. On other hand, the decoding on the Working-
192 memory and Relational-processing tasks showed high sensitivity to the choice of *K*-order and were
193 continuously improved as increasing *K*, implying that the inter-network functional interactions might
194 play an important role in the high-order cognitive functions. Similar findings were observed were
195 observed using the spatial-graph as well ((Figure **3f**), except that all cognitive tasks benefitted from
196 high-order graph convolutions including Language tasks, which involve spatially distributed brain
197 networks for language comprehension and numerical processing.

198 Communities and brain networks accelerate information propagation on brain graph

199 There are multiple ways of capturing the network organization of human brain, using either spatial
200 constraints (e.g. spatial adjacency in the geometry of brain surface (spatial-graph)), brain morphology
201 (structural covariance of cortical thickness across a population of subjects (structural-graph)), or
202 connectivity information (e.g. functional correlation of low-frequency fluctuations in BOLD signals
203 (functional-graph) and tractography of fiber projections using diffusion-weighted images (diffusion-
204 graph)), and others. Here, we mainly investigated these four types of brain graphs (Figure **4**). Among
205 which, the spatial graph featured local geometric structures by connecting each parcel to its spatial
206 neighbors on the cortical surface mesh. By contrast, modular and community structures have been
207 widely demonstrated in the diffusion and functional brain graphs [23,25]. The structural graph (i.e.
208 structural covariance of cortical thickness) somewhat resembles the functional graphs [26,27], but was
209 contaminated by random noise effects (as shown in Figure **4c**).
210 The training curves on all brain graphs followed a similar trend as increasing the *K*-order (see Figure
211 **3 a-d**) and showed similar behaviors when using a high propagation rate (i.e. K=5, as shown in Figure
212 **4a**), except for the structural-graph which showed slightly lower performance than other graphs.
213 Besides, we observed a significant interaction between the *K*-order and choices of brain graphs, such
214 that the best decoding accuracy, evaluated on the test set after the entire training process, was
215 achieved on the functional graph with a high-order model (Figure **4b**). When using spatial-graph (also

shown in Figure **3b**), a gap in the decoding performance was detected between the low and high propagation groups (90.36% vs 92.4% for K≤2 and K>2, respectively) and reached a stable range after K>5 (93%). A much smaller impact of the K-order was shown in the functional-graph (Figure **3a**), with a smaller gap in brain decoding appearing at K=1 (92.52% vs 93.22%, respectively for K=1 and K>1). Diffusion-graph showed very similar behaviors as the functional-graph at all propagation rates (92.19% vs 93.04%, respectively for K=1 and K>1) and eventually converged to the same level of decoding performance when using a sufficiently high propagation rate (Figure **3c**). This is probably because the spatial-graph is only composed of short-distance connections, which restricts the integration of brain activity to a small local area. In order to reach out for the spatially distributed functional networks and even multiple brain systems, the model requires a relatively high *K*-order or a high path length, i.e. taking multiple walks on the graph. However, even with a high-order model, the spatial-graph still showed lower decoding performance compared to the diffusion and functional graphs (Figure **4b**). Our results indicated that the connectome-constrained graph architecture accelerates the propagation of information flow in the brain by integrating the neural dynamics through long-range connections. These findings coincided with the concept of global efficiency in the graph theory. Numerous studies have illustrated that the diffusion and functional brain networks, rather than regular spatial networks, followed the small-world configuration with information segregation and integration at low wiring and energy costs [23,24,28].

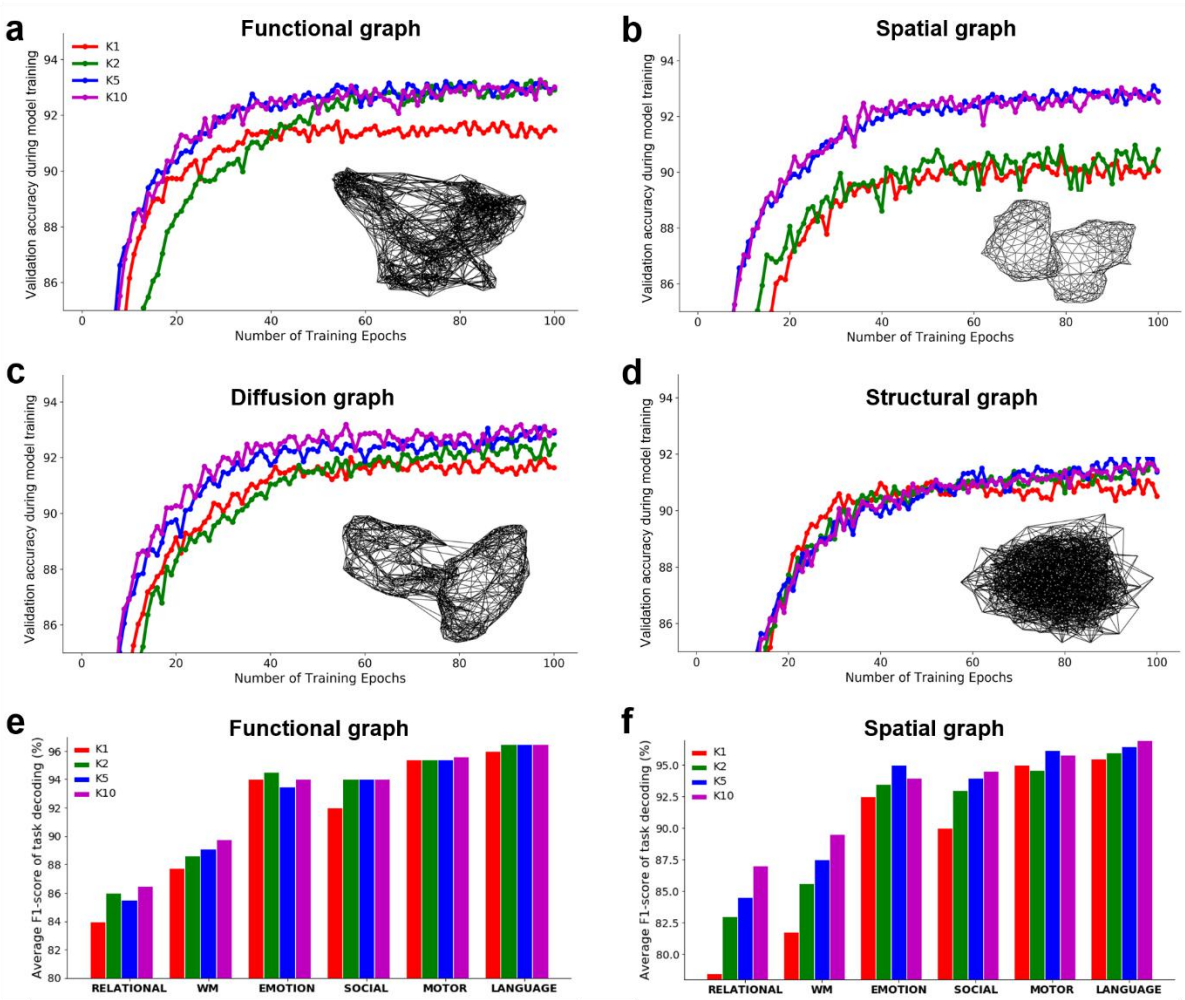

Figure 3. Effect of the *K*-order in ChebNet on four brain graphs and six cognitive domains. We evaluated four different types of brain graphs in the construction of ChebNet, including functional-graph computed using resting-state functional connectivity (**a**), spatial-graph representing spatial adjacency in the cortical surface (**b**), diffusion-graph estimated from whole-brain diffusion tractography (**c**), and structural-graph derived from structural covariance of cortical thickness across subjects (**d**). We first evaluated the effect of *K*-order on each brain graph in terms of the training curve, i.e. evaluating the model performance on the validation set at the end of each training epoch. We found a significant boost in decoding performance by using high-order graph convolutions (K>1) instead of integrating within the network (K=1) on all brain graphs. We also evaluated the impact of *K*-order on each cognitive domain by calculating the averaged F1-score on the test set and compared the effects across six cognitive domains and different brain graphs, e.g. the functional graph (**e**) and spatial graph (**f**). These results indicate that the decoding of high-order cognitive functions relies more

on high-level interactions in the brain and this effect is more dramatic when using a graph architecture purely based on the topological organization of human brain.

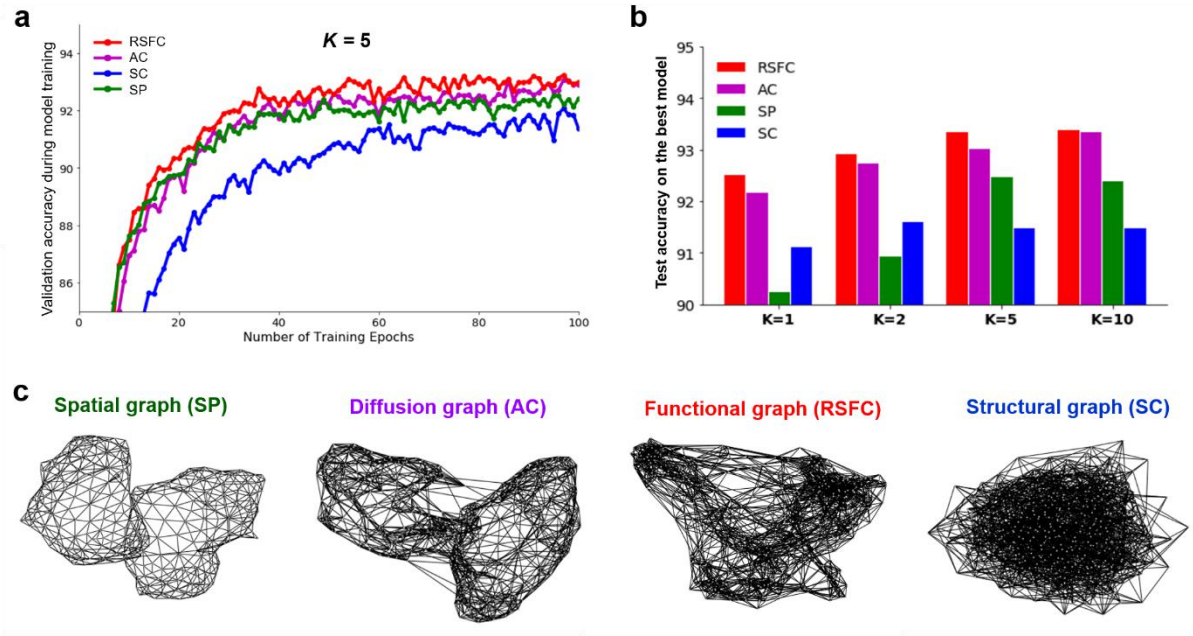

Figure 4. ChebNet decoding at different *K*-orders and using different brain graphs.

We evaluated four different types of brain graphs for the multi-domain brain decoder, with the corresponding brain graph architecture shown in (**c**). We observed a strong interaction effect between brain graphs and *K*-orders (**b**) such that the decoding performance gradually improved by increasing the *K*-order for all brain graphs, but the graph architecture also had a big impact on the decoding performance. When fixing the *K*-order, for example using the ChebNet-*K*5 model (**a**), the functional (RSFC, in red) and diffusion graphs (AC, in purple) showed similar performance during model training, followed by the spatial graph (SP, in green) showing slightly lower decoding accuracy. The structural graph (SC, in blue) showed the lowest performance during the entire training process. Overall, the best decoding model was using ChebNet-*K*5 on the functional-graph (test accuracy=93.43%). To be noted that the baseline *K=0* model, which was invariant to the structure of brain graphs by treating each node independently, achieved much lower decoding performance on the same dataset (test accuracy=83.72%). Moreover, if we removed the graph convolutional layers from the decoding model, i.e. using a two-layer multilayer perceptron (MLP) with 256 and 64 neurons respectively, the decoding performance was further reduced (test accuracy=76.1%). These results

265 indicate that high-order functional integration as well as the proper graph architectures are critical
266 steps towards high-performing cognitive decoding.

267 Functional homogeneity and small parcel size promotes local information processing

268 Another factor that significantly impacts the decoding model is the scale of the graph, i.e. number of
269 brain parcels. Generally speaking, finer-scale atlases (smaller parcel size) will have higher internal
270 homogeneity (Figure 5-S2 d) and result in less information loss after data projecting and averaging.
271 Consequently, the decoding model using smaller parcels achieves better prediction of cognitive states
272 and eventually reaches the plateau with a balance of model complexity and local homogeneity (Figure
273 5-S1 and S2). In order to investigate this effect, we have tested four different brain atlases derived
274 from different modalities or human connectomes, including Yeo's 7 and 17 functional networks [20] (50
275 and 112 spatially continuous brain parcels), Brainnetome atlas [19] derived from diffusion tractography
276 (246 brain parcels), and Glasser's multi-model atlas [21] (360 brain parcels), as well as brain atlases at
277 multiple resolutions, e.g. Schaefer's multiresolution brain parcellation (consisting of 100, 200, 400
278 and 1000 parcels respectively) [30]. For each of the chosen atlases, we constructed the functional graph
279 by calculating the resting-state functional connectivity based on 1080 subjects from the HCP database
280 and evaluated the functional homogeneity of each brain parcel by calculating pairwise correlations of
281 the connectivity patterns between all vertex within a parcel [30,31]. Coinciding with the literature that
282 smaller parcels have higher functional homogeneity, we further demonstrated that finer-scale atlas
283 results in higher decoding performance across different parcellation schemes (Figure 4-S2). Our
284 results indicated that smaller parcel size resulted in higher functional homogeneity and better
285 decoding performance (Figure 5-S2). As shown in Figure **5a**, the decoding performance started with a
286 relatively low accuracy, for instance, 80.78% when using 50 spatially continuous regions derived
287 from Yeo's 7 network, and quickly improved by simply increasing the resolution of brain atlas, e.g.
288 85.23% when using 112 regions derived from the 17 network. Small improvement was detected when
289 using around 300 regions or more (92.16% vs 93.43%, respectively for the Brainnetome atlas and
290 Glasser's atlas). Moreover, we observed that, on all brain atlases, the decoding performance gradually
291 improved as increasing the *K*-order and reached the stable performance at *K*=5 (Figure **5b**).

We also evaluated Schaefer's multiresolution brain parcellation [30], consisting of 100, 200, 400 and 1000 parcels respectively. As shown in Figure 5-S1, we found that higher resolution (smaller parcel size) improved the decoding performance at all *K*-orders and more dramatic improvement was detected at a small *K*-order than high-order models (e.g. K=1 vs K=5 in A and B, respectively). For instance, when using the ChebNet-K5 model, the decoding performance plateaued at 400 parcels with no further improvement by using higher resolutions, e.g. 1000 parcels. It's worth noting that, compared to the Glasser's atlas, the Schaefer's atlas showed slightly lower performance at all resolutions and *K*-orders (e.g. 91.35% vs 93.43%, respectively for Schaefer's (400 parcels) and Glasser's atlas (360 parcels)). This is probably due to imperfect matching of cortical surface templates between different populations and consequently having much lower functional homogeneity when projecting Schaefer atlas onto the surface of HCP subjects (Figure 5-S2). The relationship of parcel size, functional homogeneity, and decoding accuracy was systematically investigated in Figure 5-S2. Our results indicate that the size and functional homogeneity of brain parcels significantly impact the decoding model. By contrast, how the brain atlas was constructed, for instance, using different imaging modalities or various types of connectome information, show little impact.

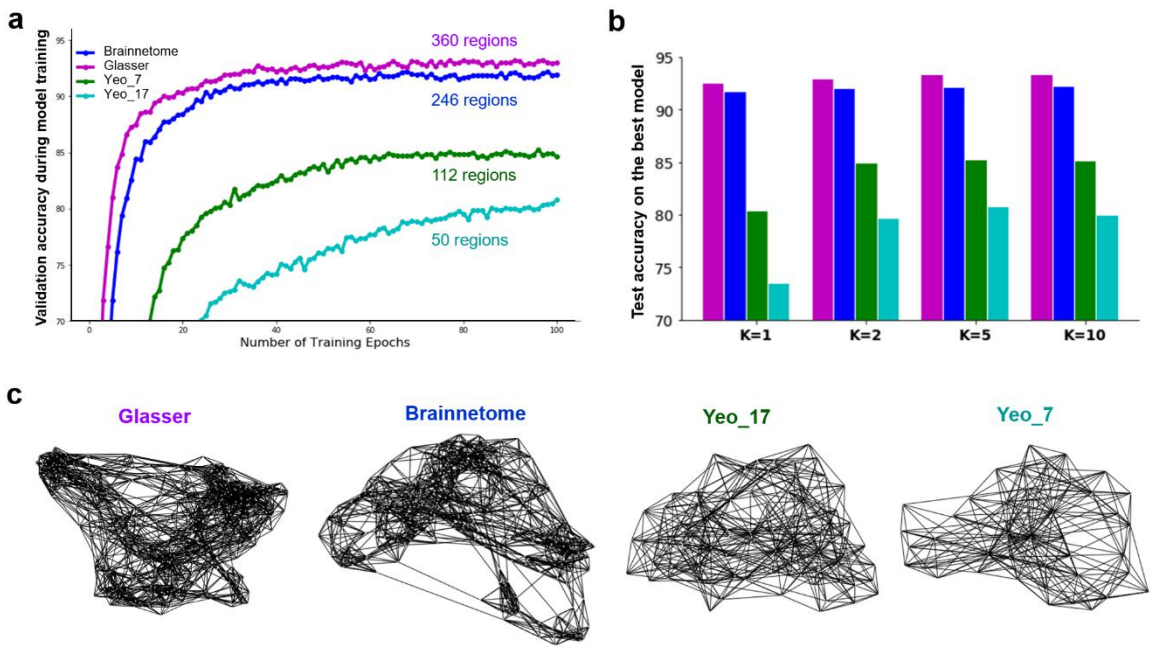

Figure 5. ChebNet decoding at *K*-orders and using different brain atlases.

We evaluated different brain atlases at various resolutions for the construction of brain graph, including Yeo's 7 and 17 functional networks (50 and 112 spatially continuous brain parcels,

311 respectively), Brainnetome atlas derived from diffusion tractography (246 brain parcels), and
312 Glasser's multi-model atlas (360 brain parcels). For each brain atlas, we constructed the functional
313 graph by evaluating the region-wise functional connectivity based on resting-state fMRI data (**c**). We
314 trained separated decoding models at different *K*-orders and found that the Glasser's atlas achieved
315 the best decoding performance, followed by the Brainnetome atlas, and Yeo's 17 and 7 functional
316 atlases (**b**). When fixing the *K*-order, for example using the ChebNet-*K*5 model (**a**), we found
317 significant advantages in model training by using finer-scale brain atlases (i.e. more brain regions and
318 smaller parcel size) regardless of the information used for brain parcellation, e.g. functional, diffusion
319 or multi-modal human connectomes.

320 ## High-order ChebNet adapts to network misspecification

321 The robustness of the decoding model was first evaluated by using randomized brain graphs (as
322 shown in Figure **6c**), which were generated by randomly swapping a proportion of edges in the
323 functional graph while keeping the node degrees unchanged. This procedure has been widely used as
324 null models in network neuroscience literature, for instance [25,32]. Here, we found a significant
325 interaction between graph randomization and the *K*-orders in the decoding model (Figure **6b**) such
326 that all randomized graphs showed lower prediction accuracy than the functional graph at small
327 propagation rates (e.g. $K = 1$) and gradually converged to a similar level at high propagation rates
328 ($K > 5$). Specifically, when using a low random ratio (e.g. 0.1), the model achieved very similar
329 decoding performance as the original functional-graph when $K \geq 2$ (93.15% vs 93.22%, respectively
330 for the randomized and functional graphs). When using a relatively high random ratio (e.g. 0.5), the
331 decoding performance started with a low prediction accuracy and gradually improved by increasing
332 the *K*-order (86.95% vs 91.52%, respectively for $K < 2$ and $K \geq 2$), consistently lower than the
333 original functional-graph at all *K*-orders. Besides, during model training, the randomized graphs
334 showed much lower convergence speed than the original functional-graph and were prone to getting
335 stuck in local minima, especially when only a small set of subjects were available (as shown in Figure
336 6-S1). Moreover, as stated before, structural-graph also exhibits some random effects (Figure **3d**),

which resembles the behaviors of random graphs with the random ratio between 0.1 to 0.2 (91.13% vs 91.53%, respectively for $K < 2$ and $K \geq 2$).

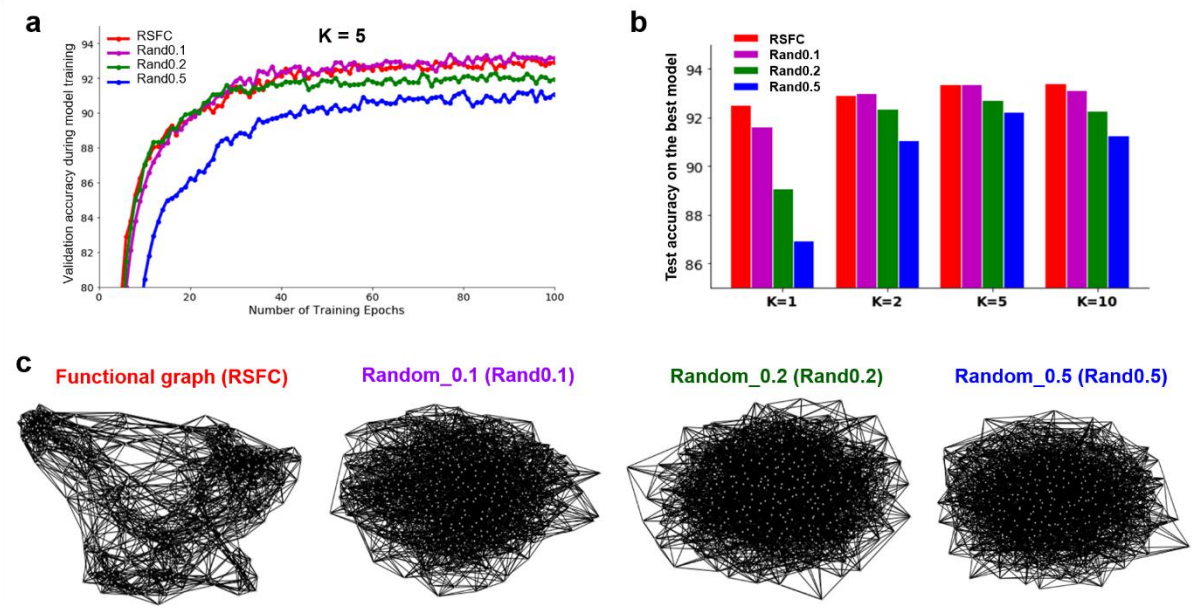

Figure 6. ChebNet decoding using random graphs derived from the functional graph. We generated three sets of randomized brain graphs by randomly swapping a portion of edges in the functional graph while keeping the node degree unchanged. The graph structures at different random ratios (e.g. 0.1, 0.2, 0.5) were shown in (**c**). When fixing the *K*-order, for example using the ChebNet-*K*5 model (**a**), the functional graph (RSFC, in red) and its low randomness copy (Rand0.1, in purple) showed very similar performance during model training, followed by the moderately (Rand0.2, in green) and highly randomized graph (Rand0.5, in blue). We observed a strong interaction effect between randomness ratios and the *K*-orders (**b**) such that the decoding performance on randomized graphs started with a low prediction accuracy and gradually improved when increasing the *K*-orders, consistently lower than the original functional-graph at all *K*-orders. The impact of graph randomization was more dramatic at a small *K*-order (e.g. *K*=1) and smaller impact was detected when using high-order models. The best performance was achieved by using a high-order model on the functional graph or its low random copies.

Robustness to random attacks in brain regions and networks

The robustness of the decoding model was also evaluated under random attacks, by removing brain responses from a small set of regions in the brain. This procedure has been commonly used to simulate brain lesions due to brain injury or neurological disorders [33,34]. Here, we mainly focused on two types of attacks, either using randomly chosen regions or targeted regions from a specific brain network. Our results demonstrated that the proposed ChebNet graph convolution was resilient to random attacks on regions but not networks. Specifically, lesions on a small set of randomly chosen brain regions, e.g. 40 parcels or less, did not affect the decoding performance (median decay in the decoding accuracy: 1.2%). More severe decays were detected as more regions were affected by lesions (median decay: 2.8%, 15.3% and 40.4% respectively for 72, 180 and 288 affected regions). This pattern of performance decay was not related to the betweenness or centrality of targeted regions (r=0.091, p=0.08), but rather associated with their involvement in task-related brain activations. For instance, the decay in the decoding of WM tasks was mostly driven by areas in the ventral visual stream, including PH, VVC and V8, as well as frontoparietal network regions, including PFm, MFG and area 6r, as revealed by a correlation analysis between the decay of decoding performance and the infection of each brain region (as shown in Figure **7a**). These regions have been reported to be engaged in the process of WM tasks [35–37] with a similar pattern revealed by the saliency map analysis (as shown in Figure **8**). On the other hand, much larger decays were observed when switching to network attacks (around 10%), with most severe cases found in the limbic, frontoparietal network (FPN) and attention networks (consisting of 60, 82, 104 affected parcels, respectively). A similar pattern of decays was observed when using the 17-network parcellation, with the largest decays found in the networks related to cognitive control and attention (as shown in Figure **7b** and **7c**). Interestingly, the distribution of performance decays in cognitive decoding was not related to the size of affected intrinsic networks (r=0.097, p=0.332). To conclude, the ChebNet decoding model was robust to random attacks but was more vulnerable to targeted attacks on task-related brain regions and networks. Moreover, compared to the low-order model (e.g. K=1), the high-order ChebNet model was more

resilient to random or targeted attacks by showing smaller decays in cognitive decoding due to lesions on regions and networks (Figure 7-S1).

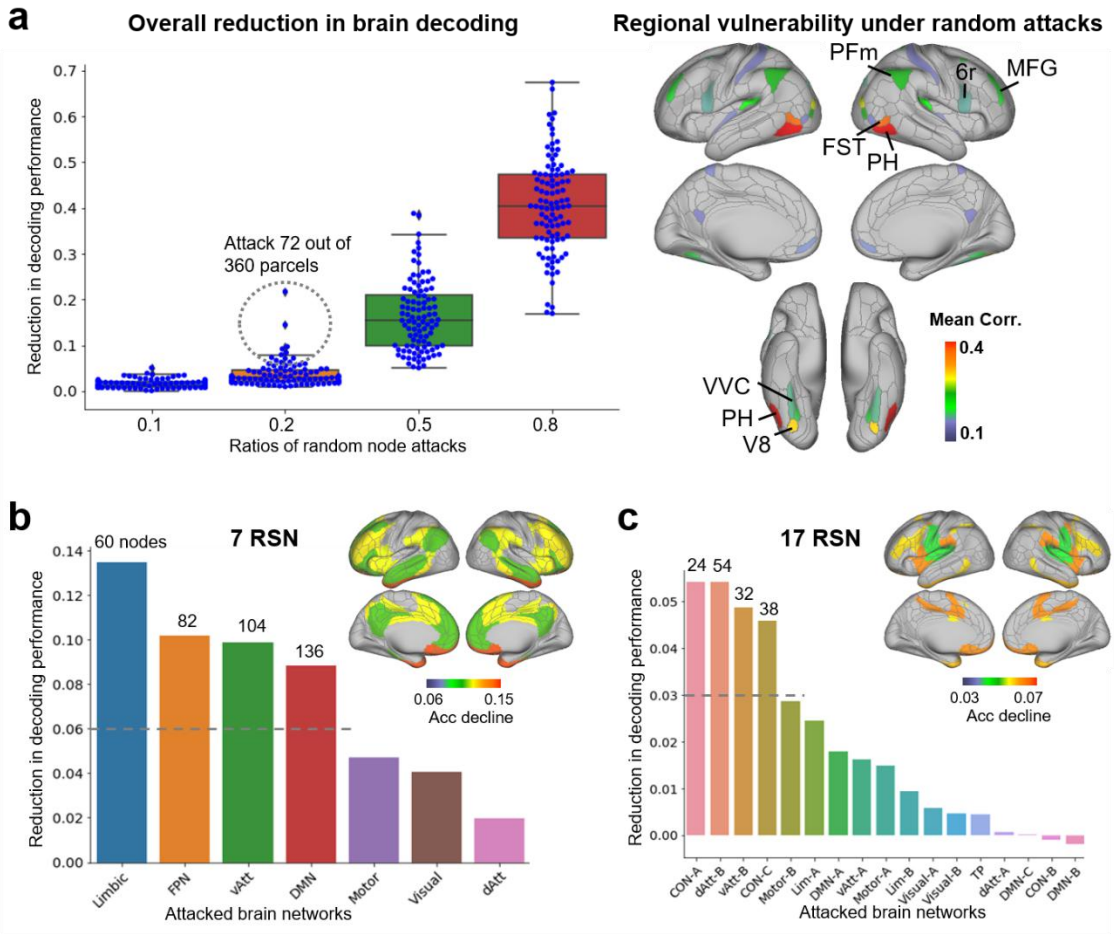

Figure 7. Robustness of ChebNet decoding to random attacks on nodes and networks.

We simulated two types of brain lesions, either attacking randomly chosen regions or targeting for regions from a specific brain network. For random node attacks (**a**), we silenced brain responses from randomly chosen nodes, ranging from 10% to 80% of brain parcels, and repeated the process for 100 times. We found that the decoding model was highly resilient to random node attacks by showing a small decay in the decoding accuracy (<3% when attacking around 80 regions). The most vulnerable regions include areas in the ventral and dorsal visual streams (top right panel). For network attacks, we removed one brain network at a time, identified by Yeo's 7 (**b**) and 17 (**c**) resting-state functional networks, and silenced the associated brain parcels with at least 30% areas affected. We found severe decays in the cognitive decoding when attacking specific brain networks (around 10% and 5% for 7 and 17 network parcellation, respectively). Among which, the most vulnerable networks were those

related to cognitive control and attention, for instance the frontoparietal network (FPN) and ventral attention networks (vAtt).

Node contributions revealed by saliency maps

In order to illustrate the biological basis of the decoding model, we generated the saliency maps on the trained model by propagating the non-negative gradients backwards to the input layer[38]. An input feature is *salient* or *important* only if its little variation causes big changes in the decoding output. As shown in Figure **8**, category-specific salient features were detected for all WM tasks and separately for 0back and 2back conditions. For 0back tasks, regions in the visual cortex and ventral visual stream were identified, including V1-V3, PH, FFC and PHA, coinciding with the literature that the ventral visual stream is involved in the recognition of visual stimuli [2,9]. By contrast, the frontoparietal network regions, including IPS and MFG, and the temporal regions, e.g. superior temporal gyrus (STG) and Perirhinal Cortex (PRC), were contributed to the decoding of 2back tasks. These regions have been shown to be engaged in object recognition and visual working memory [39,40]. Our findings revealed that the ChebNet decoding model captured task-related brain regions and category-specific patterns of brain activity, and consequently extracted biologically meaningful features for the classification of different task conditions.

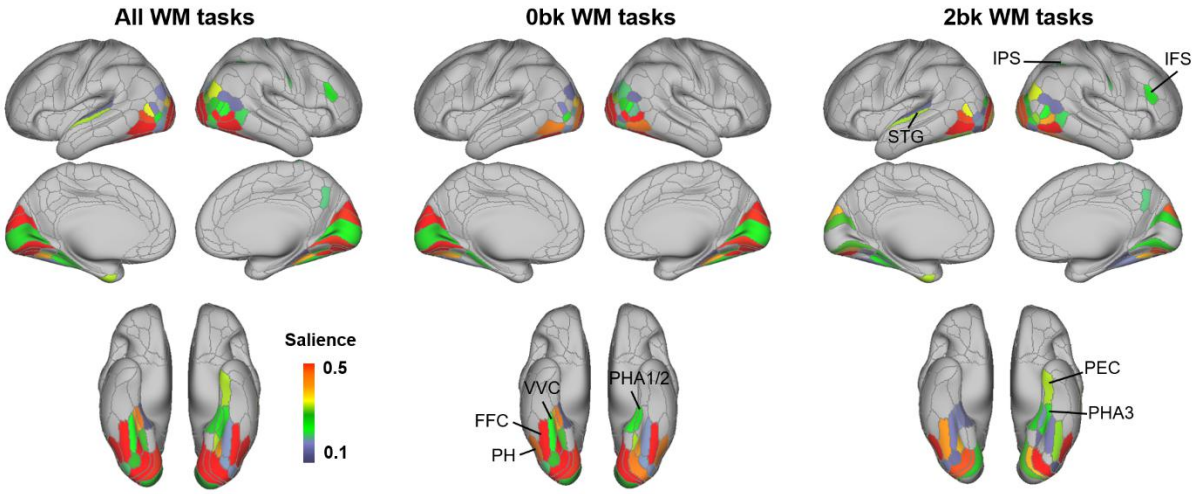

Figure 8. Saliency maps for the decoding of WM tasks.

We generated the salient features that showed high contributions to the classification of WM tasks by using the guided backpropagation approach. The model detected biologically meaningful and

category-specific salient features on 0back and 2back WM tasks. For 0back tasks, the model identified salient regions in the visual cortex and ventral visual stream, including V1-V3, PH, FFC and PHA. For 2back tasks, the model detected salient features in the frontoparietal network regions, including IPS and MFG, and the temporal regions, e.g. superior temporal gyrus (STG) and Perirhinal Cortex (PRC). Note that only brain regions with a high saliency (saliency values>0.1, full range of saliency is (0,1)) and a significant 'task condition' effect ($p<0.001$) were shown in the brain maps.

Discussion

We proposed a generalized framework for brain decoding based on ChebNet graph convolutions. The model takes in a short window of fMRI time series and a brain graph (with nodes representing brain parcels and edges representing brain connectivity), and integrates brain activity in a multiscale manner, ranging from local regions, to brain circuits/networks, and across multiple brain systems. Using a 10s window of fMRI signals, our model identified 21 different task conditions across multiple cognitive domains with a test accuracy of 93% (chance level of 4.8%). We systematically investigated the impacts of using high-order graph convolutions and connectome-constrained graph architectures in the decoding model. Our findings revealed that 1) connectomic constraints accelerate the propagation and integration of brain activity by adding shortcuts through long-range connections; 2) smaller parcel size or finer-scale atlas improves the internal homogeneity of brain parcels and boosts the overall decoding performance; 3) high-order graph convolution encodes functional integration of distributed brain activity at multiple scales and boots the decoding performance on all brain graphs. These results support an important role of functional integration in cognitive decoding, which leads to higher decoding accuracy as well as better adaptation to network misspecifications due to random rewiring and node attacks.

Functional integration plays an important role in cognitive decoding

A variety of computational models have been proposed in the field of brain decoding in the last decades, with the aim of learning a linear discriminative function on the spatial patterns of brain activation under specific experimental conditions. For instance, researchers have successfully attempted to use brain activity to reconstruct the frames of movies [5], or to decode the semantic context from words [41] and visual scenes [4] by using linear regression models. Recently, with the fast development of deep learning, artificial neural networks has also been applied in the field of decoding human cognition from recorded brain activity, for instance using classical convolutional [7] and recurrent neural networks [6], or a generalized form of convolutions in the graph domain [17]. However, the majority of current brain decoding studies so far focused on segregating neural substrates of

445 different cognitive tasks by treating each brain area independently in the decoding model [2,3,42,43]. The
446 interactions between different brain areas and networks have been largely ignored, but starts to play
447 an important part in the field. Cole and colleagues [15] demonstrated the possibility of predicting
448 activations of unseen brain regions or new subjects by transmitting information flow of brain activity
449 within a functional network. Using a similar idea, Li and Fan [6] have successfully inferred cognitive
450 states from network-level neural activity by integrating signals from large-scale brain networks.
451 Accordingly, we recently proposed a GNN architecture that propagates brain activity through
452 functional connectomes and integrates the context of brain activity in both local and global extent [17].
453 The model has achieved high decoding accuracy over a variety of cognitive domains -90% in [17]-
454 outperforming other baseline approaches on the benchmark, including non-integrative full-brain
455 models, e.g. 64% in linear SVM and 76% in MLP.
456 In this study, we further extend the GNN framework by enlarging the scale of functional integration in
457 graph convolutions and exploring the impacts of different connecome priors in cognitive decoding.
458 Firstly, we implement high-level integration of task-evoked brain activity in the decoding model,
459 extending from a single brain network (K=1, similar to [17]), to multiple brain systems (K>1) and
460 towards the full brain (K>5). Secondly, we explore different choices of graph architectures that
461 restrict the propagation of brain activity through the topology or connectome of the human brain. Our
462 results demonstrate a significant improvement in large-scale cognitive decoding by implementing
463 high-order interactions on functional connectomes (93% using the 5-order ChebNet model). We
464 observed this effect on all types of brain graphs (Figure **3**). In addition, the high-order integrative
465 model resulted in a similar level of decoding performance when using different connectome priors, for
466 instance the diffusion-graph derived from whole-brain tractography of fiber projections (Figure **4**), but
467 higher than topological and morphological priors.
468 The sensitivity to functional integration was not uniformly distributed among all cognitive domains
469 but rather depended on the cognitive demands of tasks (Figure **3e**). For instance, WM and Relational-
470 processing required a high-order model to integrate brain activity across multiple brain systems, while
471 Language and Motor tasks already achieved the optimal performance by taking into account the
472 interactions within targeted brain networks. These findings may partially explain the excellent

473 decoding performance in previous decoding studies when tackling unimodal cognitive functions, for
474 instance, distinguishing body movements using patterns of brain activity from in the motor cortex [44]
475 or restricting the context of brain activity within a local area [7], but showed poor decoding on high-
476 order cognitive functions, for instance, classifying 0-back and 2-back WM tasks [7]. By contrast, when
477 taking into account the integration within the functional networks, the model achieved high accuracy
478 on both unimodal and high-order cognitive functions [6,17], and showed a further improvement by
479 implementing high-order interactions among brain networks (Figure 3e). Our results indicate that
480 high-performing cognitive decoding relies on the encoding of high-level functional integration in
481 brain activity, not limited to local brain activity or low-level interactions. In addition to higher
482 decoding accuracy, the integrative model also improved the robustness to perturbations on the graph
483 architecture, for instance, by using different types of interactions (Figure **4**), network rewiring (Figure
484 **6**), as well as random attacks on brain regions and networks (Figure **7**). These findings greatly expand
485 current main perspectives on brain decoding that aims to segregate neural substrates of different
486 cognitive tasks, and suggest an important role of functional integration during cognitive processes,
487 especially for high-order cognitive functions.

488 Connectome-constrained graph architectures accelerate information propagation and 489 functional integration

490 A series of studies in the literature have illustrated a great potential of using resting-state functional
491 connectivity in predicting cognitive functions, not only in behavior [45,46] but also in task-evoked brain
492 activity [15,47]. Recent decoding studies have shown that the connectivity patterns [48,49] as well as the
493 activity flow on functional networks [6,17] are also predictive of cognitive states. In the current study, we
494 further demonstrated the potentials of using other types of human connectomes beyond functional
495 networks. As one of the basic graph architectures, brain topology and its derived spatial constraints
496 has been widely used in fMRI analysis, for instance generating brain parcels on individual brains [50–52].
497 When implementing GNN with such constraints, the model generalizes the classical convolutional
498 operations defined in the 3D volume space, e.g., [7], onto the geometry of cortical surfaces.
499 Interestingly, the topologically constrained decoding model achieved a decent performance on

unimodal cognitive functions, similar to previous studies [7], but showed lower performance on high-order cognitions compared to functional connectomes (Figure **3f**). The morphological brain networks, e.g. structural covariance of cortical thickness, on the other hand, resembled randomized functional connectomes by showing slightly lower prediction accuracy than the original graph in all cases (Figures **4** and **6**). Compared to them, the biologically constrained graph convolutions using either functional or diffusion brain connectomes achieved the highest decoding performance along with high robustness to network misspecifications due to random rewiring and node attacks. Our results revealed a clear gradient in the contribution of graph architectures for cognitive decoding, ranging from the lowest decoding performance by restricting activity flow within local areas (spatial-graph), to moderate performance by integrating information within neural circuits (diffusion-graph), and to the highest performance when using functional networks (functional-graph). Moreover, this gradient was only dominant at a low propagation rate and mostly diminished when using a high propagation rate, by converging to a similar level on all brain graphs (Figure **4**). We observed a similar interaction effect on the randomized functional graphs as well (Figure **6**). One possible explanation is that the spatial graph only consists of short-distance connections, which restricts information integration to a small local area on the cortical surface. When using a high-order propagation rate (K>1), the model expands the receptive field or neighborhood size at each graph convolutional layer by taking multiple steps at once, and eventually reaches a large neighborhood that includes distributed brain regions from a task-related functional network when K is sufficiently high (K>5). On the other hand, by adding modular and community structures to the graph, extracted from diffusion and functional connectomes, the model further accelerates the propagation of brain activity in the brain by adding shortcuts through long-range connections. Numerous studies have illustrated that the diffusion and functional brain networks, rather than regular spatial networks, followed the small-world configuration with information segregation and integration at low wiring and energy costs [23,24,28].

Our results revealed that, in addition to the K-order in ChebNet graph convolution, which controls the propagation rate of information flow on the graph by taking multiple walks at once, the architecture of brain graphs restricts the scale of the information flow at each step, ranging from local areas (spatial-graph), to neural circuits (diffusion-graph), and to functional networks (functional-graph). Among

which, the biologically constrained graph structures derived from human brain connectomes highly boost the decoding of cognitive processes.

Limitations and Future directions

In this study, we only investigated the choices of group-wise brain graphs in the decoding model, i.e. all subjects taking the same definition of nodes and edges in the graph architecture. The benefit of using such a brain graph is that the decoding model can easily generalize across large populations (more than 1000 subjects in our case), transfer onto new populations and datasets, as well as better adaptation to network misspecifications due to random rewiring and node attacks. However, our findings did not rule out the possibility of using an optimal individual brain graph in cognitive decoding, for instance constructing an individualized brain parcellation and subject-level graph architecture. This might be beneficial when handling massive fMRI data from a single subject. Due to the limited amount of functional imaging data from individual brains in the HCP database, this option was not explored in the current study. In the next project, we plan to investigate how individual brain parcellation schemes and individual graph architectures improve brain decoding of complex cognitive processes on individual brains by using fMRI data from a local data collection (www.cneuromod.ca).

Conclusions

In summary, we propose a connectome-based graph neural network to encode the underlying spatiotemporal organization of cognitive processes and to decode multi-domain cognitive functions with fine temporal resolution and fine cognitive granularity. The model propagates information flow of brain dynamics in a multiscale manner, ranging from localized brain areas, to a specific brain network and towards the full brain. The scale of functional integration is largely controlled by the path length, i.e. K-order of ChebNet, and is restricted by the nature of interactions, i.e. brain connectomes. Compared to the topological and morphological constraints, connectome-constrained graph convolutions achieved better decoding of cognitive states. Our findings indicate that multiscale functional integration is critical for large-scale cognitive decoding and better adapts to small fluctuations on the graph architecture including disconnections and lesions in the brain.

Methods

fMRI Datasets and Preprocessing

We used the block-design task-fMRI dataset from the Human Connectome Project S1200 release (https://db.humanconnectome.org/data/projects/HCP_1200). The minimal preprocessed fMRI data in both NIFTI and CIFTI formats were selected. The preprocessing pipelines includes two steps [53]: 1) fMRIVolume pipeline generates "minimally preprocessed" 4D time-series (i.e. ".nii.gz" file) that includes gradient unwarping, motion correction, fieldmap-based EPI distortion correction, brain-boundary-based registration of EPI to structural T1-weighted scan, non-linear (FNIRT) registration into MNI152 space, and grand-mean intensity normalization. 2) fMRISurface pipeline projects fMRI data from the cortical gray matter ribbon onto the individual brain surface and then onto template surface meshes (i.e. "dtseries.nii" file), followed by surface-based smoothing using a geodesic Gaussian algorithm. Further details on fMRI data acquisition, task design and preprocessing can be found in [53,54]. The task fMRI database includes seven cognitive domains, which are emotion, gambling, language, motor, relational, social, and working memory. In total, there are 23 different experimental conditions. Considering the short event design nature of the gambling trials (1.5s for button press, 1s for feedback and 1s for ITI) which constrain the temporal resolution of our brain decoding pipeline (i.e. using a 10s time window), in the following experiments, we excluded the two gambling conditions and only reported results on the remaining 21 cognitive states. The detailed description of the tasks used in this study can be found in [17,54].

Decoding brain activity using graph convolution

As a representative model for brain organization, brain graph has been widely used in the neuroscience literate by associating nodes with brain regions and defining edges via anatomical or functional connections [23]. Graph Laplacian and graph convolution provides a generalized framework to analyze data defined on irregular domains, for instance social networks and brain networks. Thus, a non-linear embedding of brain activity can be learned to project the brain graph onto a low-dimensional representational space [55], for instance mapping the gradients of brain organization [56]. We

recently found that the convolutions on brain graph encoded the within-network interactions of neural dynamics in cognitive tasks [17]. In this study, we applied a generalized form of graph convolution by using high-order Chebyshev polynomials and explored the impact of the following factors: 1) high-order interactions to encode both within- and between-network interactions; 2) different graph architectures, ranging from local regions (spatial graph), to neural circuits (anatomical graph) and to functional networks (functional graph); 3) different brain atlases at various resolutions.

Step 1: construction of brain graph

The decoding pipeline started with a weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$, where $\mathcal{V}$ is a parcellation of cerebral cortex into N regions, $\mathcal{E}$ is a set of connections between each pair of brain regions, with its weights defined as $W_{i,j}$. Many alternative approaches can be used to build such brain graph $\mathcal{G}$, for instance using different brain parcellation schemes and constructing various types of brain connectomes. Here, we investigated multiple choices in both aspects. First, different parcellation schemes were used, including 1) functional subdivision of the cortical surface derived from resting-state functional networks [20]; 2) anatomical parcellation of both cortical and subcortical cortex derived from diffusion tractography [19]; 3) a multimodal cortical parcellation bounded by sharp changes in cortical architecture, function, connectivity, and topography [21]; 4) a multi-scale atlas, consisting of brain parcels at different resolutions [30]. Here, we used Glasser's atlas as the default parcellation scheme, which achieved these highest performance on decoding 21 task states (see Results section). Second, we investigated various types of brain connectivity, i.e. edges between brain parcels, including 1) resting-state functional connectivity (RSFC); 2) anatomical connectivity (AC) derived from whole-brain tractography; 3) structural covariance (SC) of cortical thickness; 4) spatially adjacency (SP) in brain topology. Among which, the graph architecture derived from the anatomical and functional connectivity indicate biological or connectomic constraints, while the spatial and structural graphs represent topological and morphological constraints respectively.

Step 2: mapping of brain activity onto the graph

After the construction of brain graph (i.e. defining brain parcels and edges), for each functional run and each subject, we mapped the preprocessed task-fMRI data (e.g. "dtseries.nii" file for cortical parcellation, and ".nii.gz" file for volume parcellation) onto the set of brain parcels, resulting in a 2-dimensional time-series matrix. We created a task label with the same length as the fMRI time-series based on the experimental designs, e.g. task onsets and durations. Then, we used these task labels to extract a short series of fMRI responses from each functional run, by first splitting the entire run into multiple task blocks and then cutting the blocks into the chosen window sizes (discarding shorter time windows). To be noted that, each functional run usually contains multiple task blocks and each block is split into multiple bins of short time windows. As a result, we generated various numbers of fMRI time-series for each cognitive task, i.e. a short time-series with duration of T for each of N brain parcels, $x \in \mathbb{R}^{N \times T}$, along with a unique task label for each time-series. In total, the entire dataset includes 14,895 functional runs across the six cognitive domains and over 1000 subjects for each domain, and results in 138,662 data samples of fMRI signals $x \in \mathbb{R}^{N \times T}$ when using a 10s time window (i.e. 15 functional volumes at TR=0.72s), and 3,586,670 data samples when using a single-volume prediction.

Step 3: spatiotemporal graph convolutions using ChebNet

Graph convolution relies on the graph Laplacian, which is a smooth operator characterizing the magnitude of signal changes between adjacent nodes. The normalized graph Laplacian is defined as:

$$L = I - D^{-1/2} W D^{-1/2} \tag{1}$$

where D is a diagonal matrix of node degrees, I is the identity matrix, and W is the weight matrix. The eigendecomposition of Lapalcian matrix is defined as $L = \mathrm{U}\Delta\mathrm{U}^T$, where $\mathrm{U} = (u_0, u_1, \cdots u_{N-1})$ is the matrix of Laplacian eigenvectors and is also called graph Fourier modes, and $\Delta = \mathrm{diag}(\lambda_0, \lambda_1, \cdots \lambda_{N-1})$ is a diagonal matrix of the corresponding eigenvalues, specifying the frequency of the graph modes. In other words, the eigenvalues quantify the smoothness of signal changes on the graph, while the eigenvectors indicate the patterns of signal distribution on the graph.

For a signal x defined on graph, i.e. assigning a feature vector to each brain region, the convolution between the graph signal $x \in \mathbb{R}^{N\times T}$ and a graph filter $g_\theta \in \mathbb{R}^{N\times T}$ based on graph $\mathcal{G}$, is defined as their element-wise Hadamard product in the spectral domain, i.e.:

$$x * g_\theta = U(U^T g_\theta) \odot (U^T x) = U G_\theta U^T x \tag{2}$$

where $G_\theta = diag(U^T g_\theta)$ and θ indicate a parametric model for graph convolution g_θ, U = $(u_0, u_1, \cdots u_{N-1})$ is the matrix of Laplacian eigenvectors and $U^T x$ is actually projecting the graph signal onto the full spectrum of graph modes. Equation (2) provided an easy way of calculating the graph convolution through a series of operations of matrix multiplication. To avoid calculating the spectral decomposition of the graph Laplacian, ChebNet convolution (Defferrard et al., 2016) uses a truncated expansion of the Chebychev polynomials, which are defined recursively by:

$$T_k(x) = 2T_{k-1}(x) - T_{k-2}(x), \quad T_0(x) = 1, T_1(x) = x \tag{3}$$

Consequently, the graph convolution is defined as:

$$x * g_\theta = \sum_{k=0}^{K} \theta_k T_k(\widetilde{L}) x \tag{4}$$

where $\tilde{L} = 2L/\lambda_{max} - I$ is a normalized version of graph Laplacian with λ_{max} being the largest eigenvalue, θ_k is the model parameter to be learned at each order of the Chebychev polynomials.

Brain-decoding pipeline

The proposed ChebNet decoding model (as shown in Figure **1**)consists of 6 graph convolutional layers with 32 graph filters at each layer, followed by a flatten layer and 2 fully connected layers (256, 64 units). The model takes in a short series of fMRI volumes as input, maps the fMRI data onto the predefined brain graph and results in a 2-dimensional time-series matrix $X^1 \in \mathbb{R}^{N\times T}$, i.e. a short time-series with duration of T for each of N brain parcels. The first ChebNet layer learns various shapes of temporal convolution kernels by treating multiple time steps as input channels ($t_0, ..., t_T$) and propagates such temporal dynamics within (K=1) and between (K>1) brain networks. As a result, a set of "brain activation" maps are generated and passed on to the second ChebNet layer for higher-order information integration, and so on. The learned graph representations at the last ChebNet layer are then imported into a 2-layer multilayer perceptron (MLP) for classification.

The entire dataset was split into training (64%), validation (16%), test (20%) sets stratified by subject, which ensures that all fMRI data from the same subject was assigned to one of the three sets. Approximately, the training set includes fMRI data from 700 unique subjects (depending on data availability for different cognitive tasks ranging from 1043 to 1085 subjects in total), with 174 subjects for validation set and 218 subjects for test set. Specifically, the training set was used to train/update model parameters at each training epoch, the validation set was used to evaluate the model performance at the end of each training epoch, and the best model with the highest prediction score on the validation set was saved after 100 training epochs. The saved decoding model was then evaluated on the test set and reported the final decoding performance. We used Adam as the optimizer with an initial learning rate of 0.0001 on all cognitive domains. Additional l2 regularization of 0.0005 on weights was used to control model overfitting and the noise effect of fMRI signals. Dropout of 0.5 was additionally applied to the neurons in the last two fully connected layers. The implementation of the ChebNet graph convolution was based on PyTorch 1.1.0, and was made publicly available in the following repository: https://github.com/zhangyu2ustc/gcn_tutorial_test.git .

Effects of *K*-order in ChebNet

As stated in equation (4), the graph convolution can be rewritten as follows at different *K*-orders:

$$x *_G g_\theta = \begin{cases} \theta_0 x & K = 0 \\ \theta_0 x + \theta_1 \tilde{L} x & K = 1 \\ \theta_0 x + \theta_1 \tilde{L} x + \theta_2 \tilde{L}^2 x & K = 2 \end{cases} \tag{5}$$

where $\tilde{L}$ is a normalized version of graph Laplacian and $\{\theta_k\}_{k=1,2,..K}$ are model parameters to be learned. Specifically, *K*=0 indicates a global scaling factor on the input signal x by treating each node independently, similar to the classical mass univariate analysis for brain activation detection; *K*=1 indicates information integration between the direct neighbors and the current node on the graph (integrating signals within the same network); *K*=2 indicates functional integration within a two-step neighborhood on the graph (consisting of information from local area, within network and between networks). Thus, the choice of *K*-order controls the scale of the information integration on the graph. Generally speaking, when *K*>1, the graph convolution integrates information flow within a *K*-step neighborhood by propagating graph signals not only within the same network but also among inter-

682 connected networks. A simulation experiment shown in Figure 2-S1 illustrates that the propagation
683 rate at all brain graphs converges to a similar level after K>10. For cognitive decoding, we explored
684 different choices of K-order in ChebNet spanning over the list of [0,1,2,5,10] and found a significant
685 boost in decoding performance by using high-order graph convolutions (K>1) instead of integrating
686 within the network (K=1) especially on high-order cognition.

687 Definition of edges: brain connectivity

688 Graph architecture is another factor that impacts the propagation of information flow in the brain. We
689 explored different types of edges (or brain connectivity) for brain decoding, including 1) local
690 architecture indicating spatially adjacency in the cortical surface (spatial-graph); 2) community
691 architecture representing neural circuits derived from whole-brain tractography (diffusion-graph); 3)
692 hierarchical network architecture indicating functional correlation of low-frequency fluctuations in
693 BOLD signals (functional-graph); 4) morphological networks derived from structural covariance of
694 cortical thickness across a large population of subjects. These brain graphs were all calculated based
695 on Glasser's multimodal parcellation [21]. Specifically, for each pair of brain parcels in the atlas, the
696 spatial-graph was evaluated by calculating the spatial adjacency matrix with each element indicating
697 whether two brain parcels shared one or more triangles in the cortical surface by using the
698 "wb_commond -cifti-label-adjacency" command from Connectome Workbench software. The
699 diffusion-graph was generated from the whole-cortex probabilistic diffusion tractography based on
700 HCP diffusion-weighted MRI data collected from 1065 subjects, with each element in the
701 connectivity matrix indicating the average proportion of fiber tracts (streamlines) between the seed
702 and target parcels [57]. The functional-graph was calculated based on minimal preprocessed resting-state
703 fMRI data from 1080 HCP subjects, by using the 'tangent' method for group-wise connectivity
704 estimation with the ConnectivityMeasure method in nilearn [58], and then averaged the connectivity
705 matrices across all subjects. The structural-graph was generated by calculating structural covariance
706 of cortical thickness based on the HCP structural MRI database across 1096 subjects. To be noted that,
707 for spatial and structural graphs, only one brain graph was generated from the entire group. For the
708 anatomical and functional graphs, we used the group averaged connectivity matrix and fixed the graph

architecture for all subjects. Moreover, in order to match the sparsity levels across different brain connectomes, a k-nearest-neighbour (k-NN) graph was built based on each brain graph by only connecting each brain region to its 8 neighbours with the highest connectivity strength and ensuring that each pair of regions are connected in the graph.

Definition of nodes: brain atlas

The parcellation scheme controls the scale of the graph. A variety of parcellation schemes have been proposed in the literature by using different imaging modalities and features (see the review paper by [18]). We then investigated different parcellation schemes that have been widely used in the literature, with variable resolutions and connectivity information (e.g. functional, anatomical or multimodal connectivity information). For functional parcellation, we chose Yeo's 7 and 17 resting-state functional networks, which consist of 50 and 112 spatially continuous brain parcels in the cerebral cortex [20]. For anatomical parcellation, we chose the Brainnetome atlas that delineates the differences in connectivity profiles derived from diffusion tractography and consists of 246 brain regions in the cortical and subcortical areas [19]. For parcellation with multimodal information, we chose the Glasser's atlas that consists of 360 areas in the cerebral cortex, bounded by sharp changes in cortical architecture, function, connectivity, and topography [21]. In addition, considering that the resolution of brain parcellation may also impact the homogeneity of parcels, for instance, finer resolution (or smaller parcels) usually has higher internal homogeneity [30,31], we also evaluated brain atlases with multiple resolutions, for instance Schaefer's multiresolution brain parcellation, consisting of 100, 200, 400 and 1000 parcels respectively by integrating both local gradient and global similarity of functional connectivity patterns [30]. For each parcellation map, we evaluated the functional homogeneity by calculating the averaged pairwise Pearson correlations within each brain parcel. We further investigated the relationship between the parcel size and functional homogeneity as well as the decoding performance among all parcellation maps.

Randomized brain graphs: edge rewiring

The robustness of the decoding model was investigated by introducing randomizations on both graph architectures and brain parcels. In order to evaluate whether the decoding model was impact by small fluctuations on the graph architecture, we generated a series of randomized null models that involves rewiring of the functional-graph by swapping a proportion of connections such that local degree is preserved while the global graph architecture is randomized [32]. These randomized null models can preserve graph attributes from the original graph, including local node measures, spatial locations, and wiring cost, and has been widely used as generative null models of the empirical data in network neuroscience literature, for instance [25]. We explored different ratios of random swapping spanning over the list of [0, 0.1, 0.2, 0.5], where ratio=0.5 indicates that half of edges in the functional-graph were rewired. The randomized graphs to some degree represent the inherent mismatch between individuals or task-specific brain organization at a low ratio (e.g. 0.1), and may simulate disconnections in brain networks due to brain disorders by using a high ratio [59–61]. These randomized graphs were then compared with the original functional-graph in terms of decoding accuracy at different *K*-orders.

Randomized brain graphs: node attack and network attack

In order to evaluate whether the decoding model was robust to random or targeted attacks, we manually "silenced" a small portion of nodes (i.e. set their values to zero) and evaluated the reduction in the prediction accuracy by using pre-trained models. This procedure has been commonly used to simulate lesions due to brain injury or neurological disorders [33,34]. For random node attack, we removed brain responses from randomly chosen nodes, ranging from 10% to 80% of brain parcels and repeated the process for 100 times. The vulnerability of each brain region was evaluated by associating the lesion state (i.e. lesion or normal) with the decays in brain decoding. For targeted or network attacks, we removed nodes associated with one intrinsic network at a time, identified by Yeo's 7 or 17 resting-state functional networks [20]. Only brain regions with at least 30% areas affected were considered as lesions by silencing their activities in the following analysis. The vulnerability of

each brain network was measured by the decay of decoding accuracy after applying network lesions as compared to the original decoding model.

Saliency map of graph convolutions

The saliency map analysis aims to locate which discriminative features in the brain help to differentiate between different cognitive tasks. There are several ways to visualize a deep neural network, including visualizing layer activation [38] and heatmaps [62]. Here, we chose the first method due to its easy implementation and generalization to graph convolutions [38]. The basic idea is that if the input from a brain parcel is relevant to the prediction output, a little variation on the input signal will cause high changes in the layer activation. This can be characterized by the gradient of the output given the input, with the positive gradients indicating that a small change to the input signals increases the output value. Specifically, for the graph signal X^l of layer l and its gradient R^l, the overwritten gradient $\nabla_{X^l} R^l$ can be calculated as follows:

$$\nabla_{X^l} R^l = (X^l > 0) \odot (\nabla_{X^{l+1}} R^{l+1} > 0) \odot \nabla_{X^{l+1}} R^{l+1} \tag{6}$$

In order to generate the saliency map, we started with a pre-trained model and used the above chain rule to propagate the gradients from the output layer until reaching the input layer. This guided-backpropagation approach can provide a high-resolution saliency map with the same dimension as the input data. We further calculated a heatmap of saliency maps by taking the variance across the time steps for each parcel, considering that the variance of the saliency curve provides a simplified way to evaluate the contribution of task-evoked hemodynamic response. To make it comparable across subjects, the saliency value was additionally normalized to the range [0,1], with its highest value at 1 (a dominant effect for task prediction) and lowest at 0 (no contribution to task prediction).

Acknowledgment

This work was partially supported by the Major Scientific Project of Zhejiang Lab (No. 2020ND8AD02, No.2021ND0PI01), Courtois foundation through the Courtois NeuroMod Project and the IVADO Postdoctoral Scholarships Program. PB is supported by a salary award of "Fonds de recherche du Québec - Santé", chercheur boursier junior 2.

Author contributions

Conceptualization: YZ, PB; Methodology: YZ, PB; Visualization: YZ, PB;

Investigation: YZ, NF, PB;

Writing—original draft: YZ, NF, PB

Writing—review & editing: YZ, NF, PB

Competing interests

The authors declare no competing financial interests.

Data and materials availability

We use the block-design task-fMRI dataset from the Human Connectome Project S1200 release, downloaded from https://db.humanconnectome.org/data/projects/HCP_1200. In total, fMRI data from 1095 unique subjects under six different task domains and resting-state are used in this study. The minimal preprocessed fMRI data are used, in both CIFTI and NIFTI formats, depending on the choice of brain parcellation schemes. The proposed ChebNet decoding pipeline, as well as the optimized decoding models and the construction of brain graphs, are made publicly available in the following repository: https://github.com/zhangyu2ustc/gcn_tutorial_test.git.

800 # References

801 1. Haxby, J. V. *et al.* Distributed and overlapping representations of faces and objects in
802 ventral temporal cortex. *Science* **293**, 2425–2430 (2001).
803 2. Haxby, J. V., Connolly, A. C. & Guntupalli, J. S. Decoding Neural Representational
804 Spaces Using Multivariate Pattern Analysis. *Annu. Rev. Neurosci.* **37**, 435–456 (2014).
805 3. Poldrack, R. A. Inferring Mental States from Neuroimaging Data: From Reverse
806 Inference to Large-Scale Decoding. *Neuron* **72**, 692–697 (2011).
807 4. Huth, A. G., Nishimoto, S., Vu, A. T. & Gallant, J. L. A continuous semantic space
808 describes the representation of thousands of object and action categories across the
809 human brain. *Neuron* **76**, 1210–1224 (2012).
810 5. Nishimoto, S. *et al.* Reconstructing visual experiences from brain activity evoked by
811 natural movies. *Curr. Biol.* **21**, 1641–1646 (2011).
812 6. Li, H. & Fan, Y. Interpretable, highly accurate brain decoding of subtly distinct brain
813 states from functional MRI using intrinsic functional networks and long short-term
814 memory recurrent neural networks. *NeuroImage* **202**, 116059 (2019).
815 7. Wang, X. *et al.* Decoding and mapping task states of the human brain via deep learning.
816 *Hum. Brain Mapp.* **41**, 1505–1519 (2020).
817 8. Schrimpf, M. *et al.* Brain-Score: Which Artificial Neural Network for Object Recognition is
818 most Brain-Like? *bioRxiv* 407007 (2020) doi:10.1101/407007.
819 9. Haan, E. H. F. de & Cowey, A. On the usefulness of ‘what’ and ‘where’ pathways in
820 vision. *Trends Cogn. Sci.* **15**, 460–466 (2011).
821 10. Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R. & Haynes, J.-D. The
822 Distributed Nature of Working Memory. *Trends Cogn. Sci.* **21**, 111–124 (2017).
823 11. Pulvermüller, F., Tomasello, R., Henningsen-Schomers, M. R. & Wennekers, T.
824 Biological constraints on neural network models of cognitive function. *Nat. Rev. Neurosci.*
825 **22**, 488–502 (2021).

826 12. Quian Quiroga, R. Plugging in to Human Memory: Advantages, Challenges, and Insights
827 from Human Single-Neuron Recordings. *Cell* **179**, 1015–1032 (2019).
828 13. Poldrack, R. A. *et al.* Discovering Relations Between Mind, Brain, and Mental Disorders
829 Using Topic Mapping. *PLoS Comput. Biol.* **8**, e1002707 (2012).
830 14. Poldrack, R. A., Halchenko, Y. & Hanson, S. J. Decoding the Large-Scale Structure of
831 Brain Function by Classifying Mental States Across Individuals. *Psychol. Sci.* **20**, 1364–
832 1372 (2009).
833 15. Cole, M. W., Ito, T., Bassett, D. S. & Schultz, D. H. Activity flow over resting-state
834 networks shapes cognitive task activations. *Nat. Neurosci.* 12 (2016).
835 16. Ito, T., Hearne, L., Mill, R., Cocuzza, C. & Cole, M. W. Discovering the Computational
836 Relevance of Brain Network Organization. *Trends Cogn. Sci.* **24**, 25–38 (2020).
837 17. Zhang, Y., Tetrel, L., Thirion, B. & Bellec, P. Functional annotation of human cognitive
838 states using deep graph convolution. *NeuroImage* **231**, 117847 (2021).
839 18. Eickhoff, S. B., Yeo, B. T. T. & Genon, S. Imaging-based parcellations of the human
840 brain. *Nat. Rev. Neurosci.* **19**, 672–686 (2018).
841 19. Fan, L. *et al.* The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional
842 Architecture. *Cereb. Cortex* **26**, 3508–3526 (2016).
843 20. Yeo, B. T. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic
844 functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).
845 21. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**,
846 171–178 (2016).
847 22. Bassett, D. S. & Bullmore, E. T. Human brain networks in health and disease. *Curr. Opin.*
848 *Neurol.* **22**, 340–347 (2009).
849 23. Bullmore, E. & Sporns, O. Complex brain networks: graph theoretical analysis of
850 structural and functional systems. *Nat. Rev. Neurosci.* **10**, 186–198 (2009).
851 24. Bullmore, E. T. & Bassett, D. S. Brain graphs: graphical models of the human brain
852 connectome. *Annu. Rev. Clin. Psychol.* **7**, 113–140 (2011).
853 25. Betzel, R. F. & Bassett, D. S. Multi-scale brain networks. *NeuroImage* **160**, 73–83 (2017).

854 26. Clos, M., Rottschy, C., Laird, A. R., Fox, P. T. & Eickhoff, S. B. Comparison of structural
855 covariance with functional connectivity approaches exemplified by an investigation of the
856 left anterior insula. *NeuroImage* **99**, 269–280 (2014).
857 27. Zielinski, B. A., Gennatas, E. D., Zhou, J. & Seeley, W. W. Network-level structural
858 covariance in the developing brain. *Proc. Natl. Acad. Sci.* **107**, 18191–18196 (2010).
859 28. Liao, X., Vasilakos, A. V. & He, Y. Small-world human brain networks: Perspectives and
860 challenges. *Neurosci. Biobehav. Rev.* **77**, 286–300 (2017).
861 29. Sakumoto, Y., Kameyama, T., Takano, C. & Aida, M. Information Propagation Analysis
862 of Social Network Using the Universality of Random Matrix. *IEICE Trans. Commun.*
863 **E102.B**, 391–399 (2019).
864 30. Schaefer, A. *et al.* Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic
865 Functional Connectivity MRI. *Cereb. Cortex N. Y. NY* **28**, 3095–3114 (2018).
866 31. Urchs, S. *et al.* MIST: A multi-resolution parcellation of functional brain networks. *MNI*
867 *Open Res.* **1**, 3 (2019).
868 32. Sporns, O. Graph theory methods: applications in brain networks. *Dialogues Clin.*
869 *Neurosci.* **20**, 111–121 (2018).
870 33. Alstott, J., Breakspear, M., Hagmann, P., Cammoun, L. & Sporns, O. Modeling the
871 Impact of Lesions in the Human Brain. *PLOS Comput. Biol.* **5**, e1000408 (2009).
872 34. Honey, C. J. & Sporns, O. Dynamical consequences of lesions in cortical networks. *Hum.*
873 *Brain Mapp.* **29**, 802–809 (2008).
874 35. Christophel, T. B., Hebart, M. N. & Haynes, J.-D. Decoding the Contents of Visual Short-
875 Term Memory from Human Visual and Parietal Cortex. *J. Neurosci.* **32**, 12983–12989
876 (2012).
877 36. Harrison, S. A. & Tong, F. Decoding reveals the contents of visual working memory in
878 early visual areas. *Nature* **458**, 632–635 (2009).
879 37. Mayer, J. S. *et al.* Common neural substrates for visual working memory and attention.
880 *NeuroImage* **36**, 441–453 (2007).

881 38. Springenberg, J. T., Dosovitskiy, A., Brox, T. & Riedmiller, M. Striving for Simplicity: The
882 All Convolutional Net. (2014).
883 39. Eriksson, J., Vogel, E. K., Lansner, A., Bergström, F. & Nyberg, L. Neurocognitive
884 Architecture of Working Memory. *Neuron* **88**, 33–46 (2015).
885 40. Schon, K., Newmark, R. E., Ross, R. S. & Stern, C. E. A Working Memory Buffer in
886 Parahippocampal Regions: Evidence from a Load Effect during the Delay Period. *Cereb.*
887 *Cortex* **26**, 1965–1974 (2016).
888 41. Mitchell, T. M. *et al.* Predicting Human Brain Activity Associated with the Meanings of
889 Nouns. *Science* **320**, 1191–1195 (2008).
890 42. Haxby, J. V. Multivariate pattern analysis of fMRI: The early beginnings. *NeuroImage* **62**,
891 852–855 (2012).
892 43. Varoquaux, G. *et al.* Atlases of cognition with large-scale human brain mapping. *PLOS*
893 *Comput. Biol.* **14**, e1006565 (2018).
894 44. Mottolese, C. *et al.* Mapping motor representations in the human cerebellum. *Brain* **136**,
895 330–342 (2013).
896 45. Rosenberg, M. D. *et al.* Functional connectivity predicts changes in attention observed
897 across minutes, days, and months. *Proc. Natl. Acad. Sci.* **117**, 3797–3807 (2020).
898 46. Yamashita, M. *et al.* A prediction model of working memory across health and psychiatric
899 disease using whole-brain functional connectivity. *eLife* **7**, e38844 (2018).
900 47. Tavor, I. *et al.* Task-free MRI predicts individual differences in brain activity during task
901 performance. *Science* **352**, 216–220 (2016).
902 48. Richiardi, J., Eryilmaz, H., Schwartz, S., Vuilleumier, P. & Van De Ville, D. Decoding
903 brain states from fMRI connectivity graphs. *NeuroImage* **56**, 616–626 (2011).
904 49. Shirer, W. R., Ryali, S., Rykhlevskaia, E., Menon, V. & Greicius, M. D. Decoding
905 Subject-Driven Cognitive States with Whole-Brain Connectivity Patterns. *Cereb. Cortex*
906 **22**, 158–165 (2012).
907 50. Blumensath, T. *et al.* Spatially constrained hierarchical parcellation of the brain with
908 resting-state fMRI. *NeuroImage* **76**, 313–324 (2013).

51. Craddock, R. C., James, G. A., Holtzheimer, P. E., Hu, X. P. & Mayberg, H. S. A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Hum. Brain Mapp.* **33**, 1914–1928 (2012).
52. Ma, L. *et al. BAI-Net: Individualized Human Cerebral Cartography using Graph Convolutional Network*. 2021.07.15.452577 https://www.biorxiv.org/content/10.1101/2021.07.15.452577v2 (2021) doi:10.1101/2021.07.15.452577.
53. Glasser, M. F. *et al.* The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* **80**, 105–124 (2013).
54. Barch, D. M. *et al.* Function in the human connectome: Task-fMRI and individual differences in behavior. *NeuroImage* **80**, 169–189 (2013).
55. Ortega, A., Frossard, P., Kovačević, J., Moura, J. M. F. & Vandergheynst, P. Graph Signal Processing: Overview, Challenges and Applications. *Proc IEEE* **106**, 808–828 (2018).
56. Margulies, D. S. *et al.* Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc. Natl. Acad. Sci.* **113**, 12574–12579 (2016).
57. Rosen, B. Q. & Halgren, E. A Whole-Cortex Probabilistic Diffusion Tractography Connectome. *eNeuro* **8**, (2021).
58. Varoquaux, G., Baronnet, F., Kleinschmidt, A., Fillard, P. & Thirion, B. Detection of Brain Functional-Connectivity Difference in Post-stroke Patients Using Group-Level Covariance Modeling. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010* (eds. Jiang, T., Navab, N., Pluim, J. P. W. & Viergever, M. A.) 200–208 (Springer, 2010). doi:10.1007/978-3-642-15705-9_25.
59. Hearne, L. J. *et al.* Activity flow underlying abnormalities in brain activations and cognition in schizophrenia. *Sci. Adv.* **7**, eabf2513 (2021).
60. Sporns, O. The Non-Random Brain: Efficiency, Economy, and Complex Dynamics. *Front. Comput. Neurosci.* **5**, (2011).

61. Suárez, L. E., Richards, B. A., Lajoie, G. & Misic, B. Learning function from structure in neuromorphic networks. *Nat. Mach. Intell.* **3**, 771–786 (2021).
62. Selvaraju, R. R. *et al.* Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *Int. J. Comput. Vis.* **128**, 336–359 (2020).

Supplementary Materials for

Deep learning models of cognitive processes constrained by human brain connectomes

Authors: Yu Zhang [1,2,3,*], Nicolas Farrugia[4] and Pierre Bellec [2,3,*]

*** Corresponding Authors:**

Yu Zhang, Research Center for Healthcare Data Science, Zhejiang Lab.
yuzhang2bic@gmail.com
Pierre Bellec, Département de Psychologie, Université de Montréal.
pierre.bellec@gmail.com

This PDF file includes:

Seven Supplementary Figures: **Figure 2-S1** to **Figure 8-S1**
Two Supplementary Tables: **Table S1** to **Table S2**

Supplementary Figures

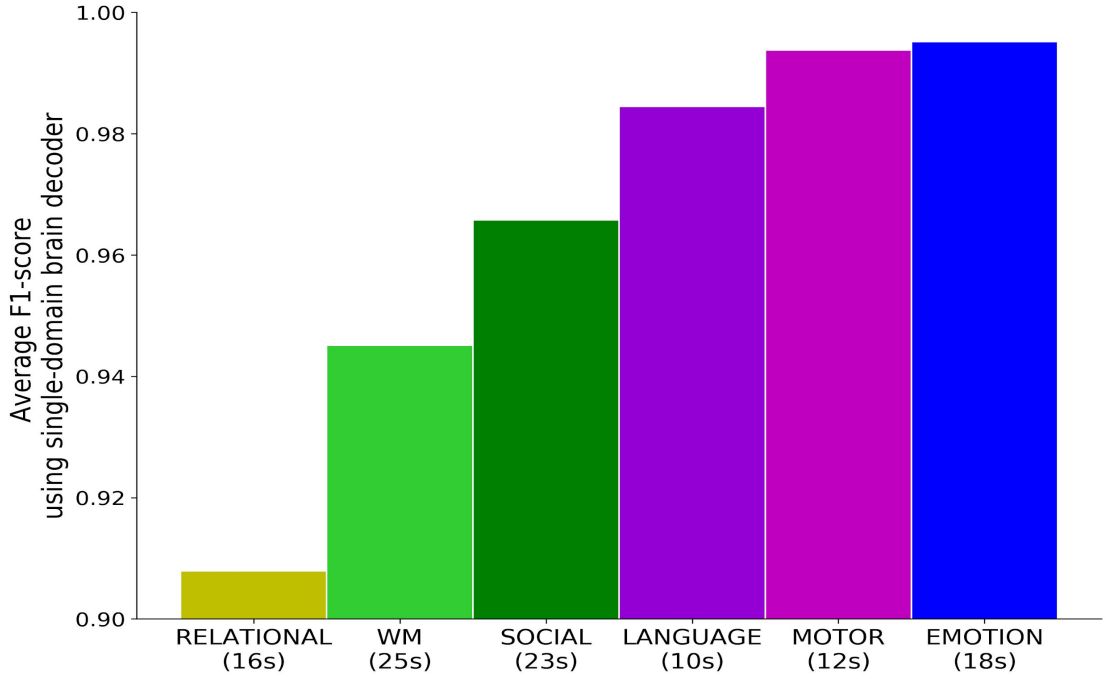

Figure 2-S1. Decoding accuracy of single-domain brain decoders for each of the six cognitive domains.

Variable temporal durations were used for the decoding models of each cognitive domain, according to the maximum length of event trials/blocks among task conditions, for instance 12s for MOTOR tasks and 25s for WM tasks. The same color scheme was used as Figure **2c**. Among the six cognitive domains, the emotion tasks (in blue, fearful face *vs* shape) and motor tasks (in magenta, distinguishing five types of body movements) were the most easily recognizable conditions, with F1-score reaching above 99%.

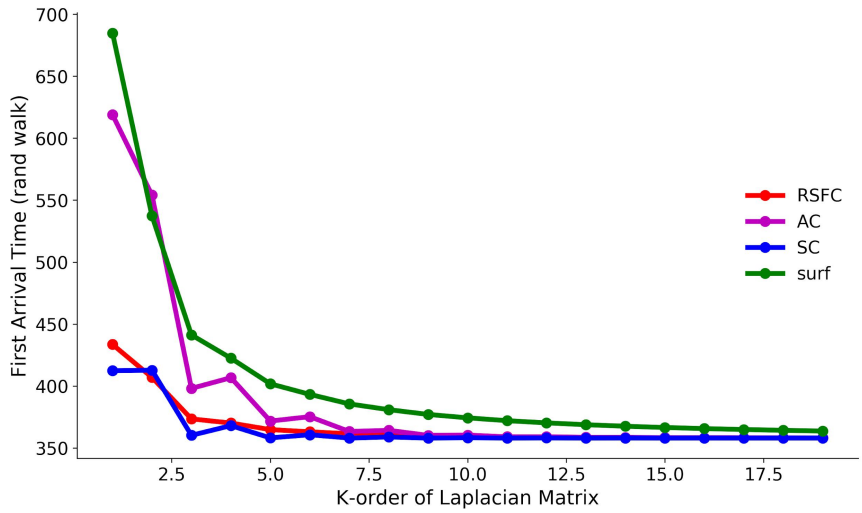

Figure 3-S1. Information propagation speed as a function of the K-order.

We used a metric in information theory, namely the first arrival time, to evaluate the information propagation speed on the brain graph, which represents the time required for the flow of the information first propagated from any source node to a given node. The expected value of first arrival time by taking random walks on the graph was calculated by, according to [29]: $m = \sum_{k=2}^{N} \frac{1}{\lambda_k}$, where λ_k is the eigenvalues of the normalized Lapalican matrix. As we increased the *K*-order, the first arrival time m was largely shortened, with the largest gradient appearing around $K = 5$ and reaching the plateau around $K = 10$ on all brain graphs. Four types of brain graphs were evaluated here, including spatial-graph (surf), structural-graph (SC), diffusion-graph (AC) and functional-graph (RSFC). Different sensitivity levels to the *K*-order were detected, such that information propagation on the spatial-graph was greatly impacted by the choice of *K*-order, followed by the diffusion-graph with a higher convergence speed. The functional-graph and structural-graph quickly converged to the plateau and followed a similar curve when increasing the *K*-order, except that higher randomized effect was observed on the structural graph at small *K*-orders.

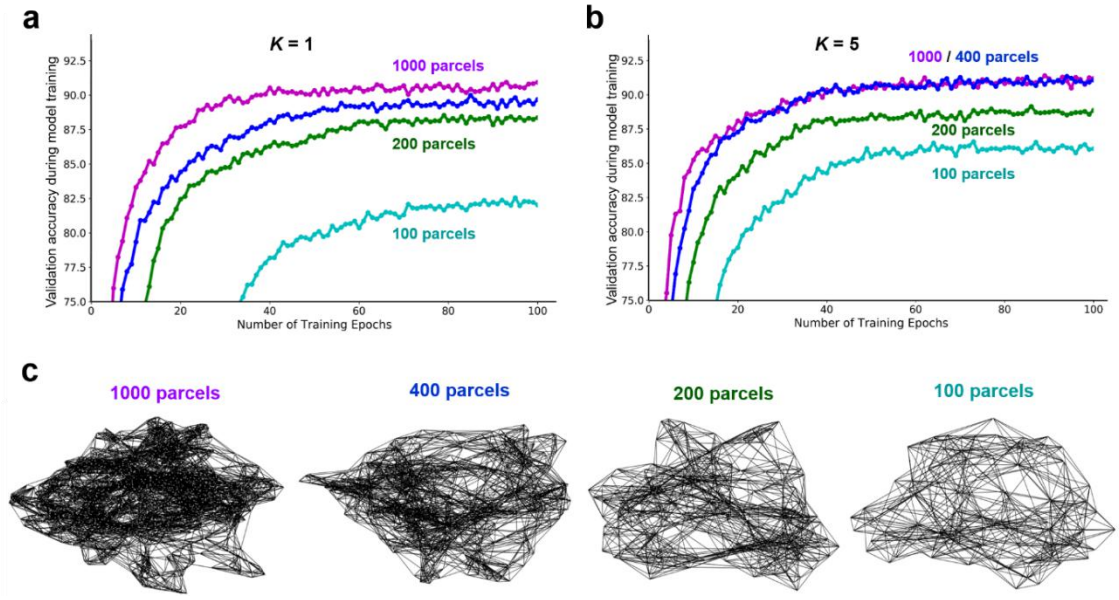

Figure 5-S1. Brain decoding using functional brain parcellation at multiple resolutions. We evaluated different brain atlases derived from Schaefer's multiresolution brain parcellation [30], consisting of 100, 200, 400 and 1000 brain parcels respectively. For each parcellation scheme, we constructed a functional graph by evaluating the region-wise functional connectivity based on resting-state fMRI data (**c**). We found that brain parcellation at higher resolutions (smaller parcel size or more regions) improved the performance of cognitive decoding, especially at a small *K*-order (e.g. K=1 in **a**). By contrast, when using a high-order model (e.g. K=5 in **b**), the model plateaued at 400 parcels with no further improvement on higher resolutions. However, we still observed improvements at small resolutions. It's worth noting that, compared to the Glasser's and Brainnetome atlases, both of which defined based on HCP subjects, the Schaefer's atlas (derived from a different population) showed slightly lower performance (91.11% vs 92.16% vs 93.43%, respectively for Schaefer's (400 parcels), Brainnetome (246 parcels) and Glasser's atlas (360 parcels) when using the ChebNet-*K*5 model), probably due to imperfectly matching between surface templates across populations which consequently impacts the functional homogeneity of brain parcels (see Figure 5-S2).

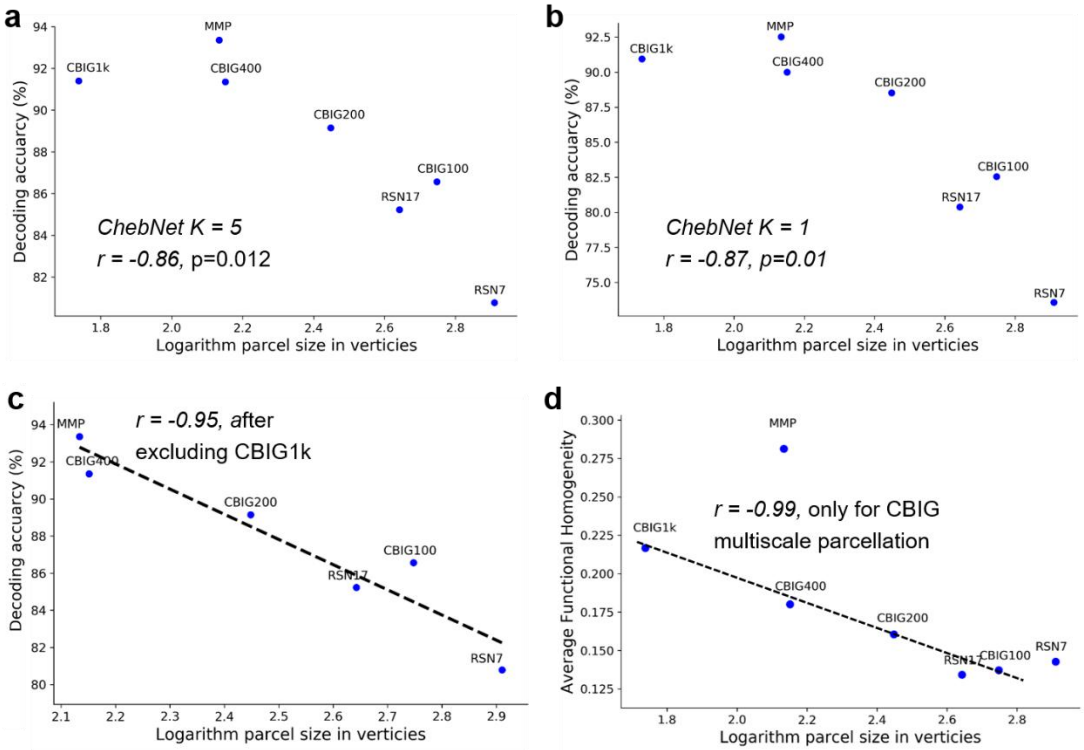

Figure 5-S2. Decoding performance associated with parcel size across different brain atlases.

We evaluated brain atlases at different resolutions for the construction of the brain graph and observed high variability in the decoding performance. The results showed that the parcel size had a significant impact on the decoding accuracy by using either ChebNet-K5 (**a**) or ChebNet-K1 (**b**) models. This effect became more significant when excluding the cases that the decoding model plateaued at more than 400 parcels (r =-0.95, p=0.003 in (**c**), excluding brain parcellation with 1000 regions). The parcel size also showed a significant impact on the functional homogeneity of each brain parcel such that finer-scale atlases (smaller parcel size) have higher functional homogeneity (**d**). For instance, among the Schaefer's multiresolution brain parcellation [30], we found a significant association between parcel size and functional homogeneity (Pearson correlation r = -0.99, p=0.001 in d). Note that we did not include the Brainnetome atlas in this analysis because it is defined on volumetric data while the rest of atlases are defined on the cortical surface. The choice of volume space or cortical surface may influence the calculation of parcel size and functional homogeneity. Our results indicate that high-resolution brain atlases result in high functional homogeneity of brain parcels, which consequently achieve better performance for the decoding of cognitive states.

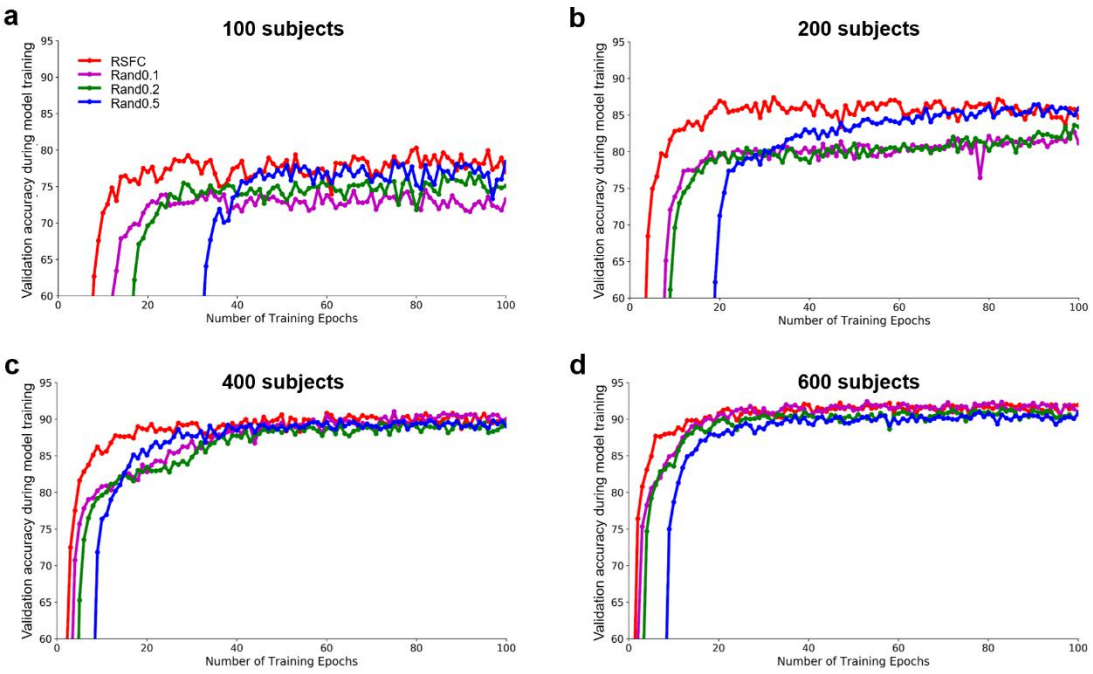

Figure 6-S1. Brain decoding using ChebNet-K5 on randomized functional graphs and small datasets.

We generated three sets of randomized brain graphs by randomly swapping a portion of edges in the functional graph while keeping the node degree unchanged. The network structures of the brain graphs were shown in Figure **6c**. Here, we found that the randomness of brain graphs showed a significant impact on the convergence speed during model training. When fixing the *K*-order, for example using the ChebNet-*K*5 model, and training the model on different sample sizes, for instance 100, 200, 400 and 600 subjects, the original functional graph (RSFC, in red) showed the fastest convergence during the training process at all cases, followed by the graphs of low (Rand0.1, in purple) and moderate randomization (Rand0.2, in green) and finally the highly randomized graph (Rand0.5, in blue). Interestingly, when only a small number of subjects were available, the highly randomized graph (Rand0.5) quickly caught up with the original functional graph during model training and eventually exceeded the decoding performance on lowly randomized graphs (e.g. Rand0.1). This effect gradually disappeared when using a sufficiently large sample size (i.e. more than 600 subjects).

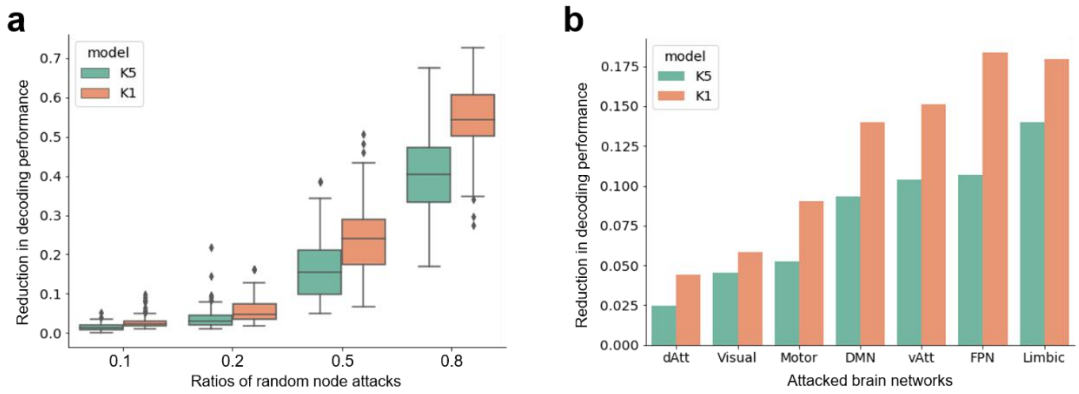

Figure 7-S1. Decays in the decoding of WM tasks using ChebNet-*K*5 and ChebNet-*K*1 models.

Compared to ChebNet-*K*5, the low-order model (e.g. *K*=1) was more vulnerable to random attacks on both brain regions (**a**) and networks (**b**). We simulated different types of brain lesions by silencing brain responses from randomly chosen nodes, ranging from 10% to 80% of brain parcels, or targeted regions from a specific brain network. We found that the decays of cognitive decoding were more severe in the ChebNet-*K*1 model at all levels for random node attacks (T-value=6.27, 4.64, 6.79 and 9.86 respectively, p-values<0.00001) and network attacks. The most affected areas consist of regions in the ventral visual stream and frontoparietal network and the most vulnerable networks are those related to cognitive control and attention, consistent with the findings on the ChebNet-*K*5 model.

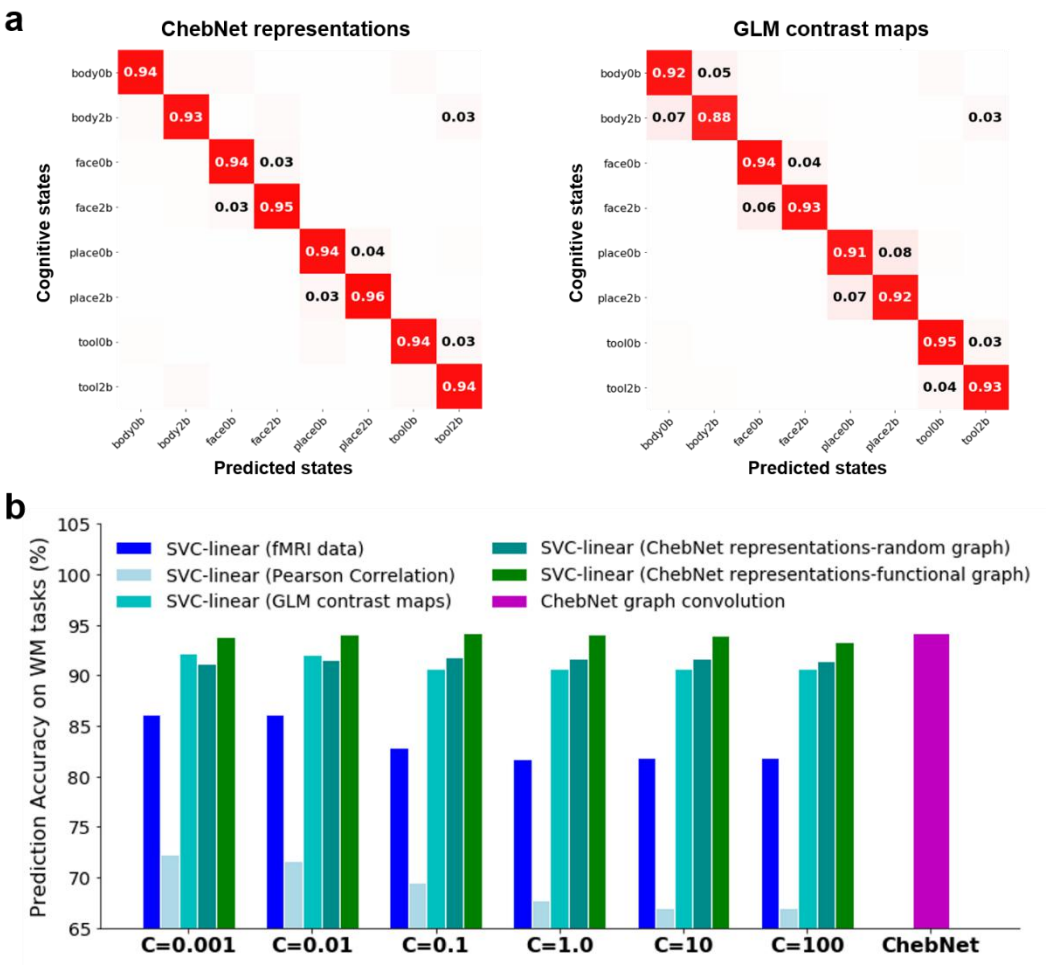

Figure 8-S1. Decoding of WM tasks using different features.

We compared the decoding of WM tasks (b) by using different classifiers, including multi-class support vector machine classification (SVC) with a linear kernel and CheNet graph convolutions, as well as different sets of features, including RAW fMRI data (in blue), functional connectivity (i.e. Pearson correlation, in light blue), contrast maps (in cyan), graph representations resulting from random (in dark cyan) and functional (in green) brain graphs. Among them, graph representations showed the highest decoding accuracy, regardless of the chosen classifiers and parameters, e.g. linear classifier like SVC or nonlinear classifier such as ChebNet. The confusion matrix of decoding eight WM tasks was shown in (a), by applying the SVC-linear classifier on either ChebNet graph representations (left panel) or GLM-derived contrast maps (right panel). The ChebNet decoding model was more powerful in discriminating 0back vs 2back tasks on all types of visual stimuli.

Supplementary Tables

Table S1. Scanning parameters and experimental designs of HCP task-fMRI dataset. The entire dataset includes in total of 14,895 functional runs across the six cognitive domains, and resulted in 138,662 data samples of fMRI signals when using a 10s time window (i.e. 15 functional volumes at TR=0.72s)

Task Domains	#Subjects	#Runs	#Volumes per run	#Trials per run	#Cond.	Min duration per block (sec)
Working-memory	1085	2	405	8	8	25
Motor	1083	2	284	10	5	12
Language	1051	2	316	8	2	10
Social cognition	1051	2	274	5	2	23
Relational processing	1043	2	232	6	2	16
Emotion	1047	2	176	6	2	18

Table S2: Comparison of cognitive decoding by using linear and nonlinear models.

We reported the best performance for the baseline models after a grid search of the hyperparameters. For SVC approaches, we used the one-vs-rest ('ovr') decision function to handle multi-classes and reported the highest accuracy after the grid search for the hyper-parameter (C = [0.0001,0.001,0.1,1,10,100]). For Random Forest, we reported the highest accuracy after evaluating different settings of the classifier including depth of trees: [4,16,64,256,1024] and number of trees: [100,2000]. For MLP (multilayer perceptron), GCN (using first-order graph convolution, [17]) and ChebNet (using 5-order graph convolution), we reported the mean and standard deviation of the decoding accuracies among 10 fold cross-validation with shuffle splits.

Models	**Train Accuracy**	**Validation Accuracy**	**Test Accuracy**
VC-linear	67.2%	63.3%	64.1%
SVC-rbf	99.7%	73.5%	73.8%
Random Forest	100%	48.0%	47.5%
MLP(256-64)	87.9%(+/-1.83%)	83.2%(+/-3.28%)	76.1%(+/-0.41%)
GCN	96.3%(+/-0.42%)	90.2%(+/-0.21%)	90.7%(+/-0.20%)
ChebNet	96.91%(+/-0.18%)	**92.73(+/-0.12%)**	**93.43(+/-0.44%)**