SI-HGN: STRUCTURE-INFORMED HYPERGRAPH NETWORKS FOR VIRTUAL PERTURBATION-BASED EFFECTIVE CONNECTIVITY INFERENCE

Anonymous authors
Paper under double-blind review

ABSTRACT

Effective connectivity (EC) characterizes directed interactions between brain regions, but inferring it from observational BOLD signals remains challenging: invasive stimulation cannot scale to the whole brain, classical methods impose restrictive stationarity or architectural assumptions, and existing data-driven surrogate models lack anatomical grounding and are limited to pairwise interactions.We present SI-HGN, a structure-informed hypergraph network that constructs individualized surrogate brains using structural connectivity (SC) as an anatomical scaffold while learning latent hyperedges to capture coordinated multi-region interactions. Hypergraph-derived edge attention modulates SC-supported anatomical pathways, allowing their contributions to vary across individuals and brain states while preserving anatomical support. Directed EC is then estimated by virtually perturbing individual regions of the frozen surrogate and aggregating the induced response differences across temporal segments of the BOLD time series. On generative benchmarks and macaque connectivity networks, SI-HGN recovers directed topology more accurately than neural perturbational inference and Granger causality (Pearson $r = 0.94$ vs. 0.91 on simulated recurrent systems; $p < 0.001$), with consistent gains across complementary recovery metrics. Individualized EC further reveals disease-related network patterns in Alzheimer's disease and autism cohorts and characterizes directed-network disruption in patients with disorders of consciousness. These results suggest that anatomically grounded, higher-order surrogate brains offer a principled route to interventional-style EC inference, while highlighting the need for validation in larger multicenter and longitudinal cohorts.

1 INTRODUCTION

Effective connectivity (EC) characterizes directed interactions between brain regions and provides a framework for studying information flow and the dynamic organization of large-scale brain networks (Friston, 2011). Direct neural stimulation can provide interventional evidence for these interactions, but invasive stimulation is difficult to apply at the whole-brain scale in humans, while noninvasive approaches remain limited in spatial coverage and stimulation depth (Luo et al., 2025; Mamoon et al., 2025). These limitations underscore the need for methods that infer individualized, directed interactions across the brain from observational neural recordings.

Classical EC methods address this question through statistical dependencies or explicit generative models. Granger causality assesses whether the past activity of one region improves the prediction of another; its conventional implementations rely on assumptions about stationarity and model specification. Dynamic causal modeling instead estimates interactions within a specified generative model and typically requires the definition of candidate network architectures (Friston et al., 2003; Smith et al., 2011; Deshpande et al., 2010; Seth et al., 2015). These approaches provide interpretable estimates under their respective assumptions, but applying them to large-scale, nonlinear, and heterogeneous brain dynamics presents challenges in scalability and model specification.

Neural perturbational inference (NPI) provides an alternative framework for EC inference based on learned surrogate dynamics (Luo et al., 2025). The method first trains a surrogate model to

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

predict subsequent neural activity from historical observations. After training, the model parameters are frozen, and virtual perturbations are applied to the input. Directed influences are then estimated by comparing the predicted responses before and after perturbation. In this way, EC is defined through an explicit perturbation–response procedure, providing a scalable computational framework for whole-brain directed connectivity analysis. However, the reliability of the inferred EC depends on whether the surrogate model accurately captures the dynamics relevant to perturbation responses (Zhang et al., 2026). Strong predictive performance on unperturbed observations does not necessarily guarantee reliable responses under virtual perturbation. Therefore, how the surrogate model represents interregional interactions and incorporates anatomical constraints is critical for perturbation-based EC inference.

Graph neural networks offer a principled way to address both requirements. Structural connectivity (SC) provides an individualized anatomical prior on which interregional communication can plausibly occur, while graph convolution and graph attention supply the machinery to aggregate and re-weight signals along these existing pathways (Kipf & Welling, 2017; Veličković et al., 2018). Accordingly, graph-structured models have been applied to brain-disorder classification (Parisot et al., 2018; Chen et al., 2025; Tang et al., 2025), causal analysis of brain networks (Wein et al., 2021), task-aware functional-network generation (Kan et al., 2022; Thapaliya et al., 2025), and behavioral prediction and individual identification from connectome features (Shen et al., 2017; Finn et al., 2015). These studies establish connectivity structure as a valuable inductive bias for learning brain-network representations. However, their training objectives are predominantly predictive: such models are optimized and evaluated on classification, regression, or reconstruction accuracy. Perturbation-based EC inference instead rests on a different property, namely whether the surrogate yields plausible directed responses to virtual perturbations, and this property has not been directly examined for graph-structured surrogates. Building SC into a surrogate that is explicitly trained to be queried by perturbation therefore remains an open design question.

A second challenge is representational. Brain dynamics are coordinated across groups of regions, and a perturbation to one region propagates through distributed, polysynaptic circuits; pairwise edges, which capture only second-order dependencies, cannot explicitly model such coordinated multi-region effects. Hypergraph neural networks address this limitation directly by allowing a single hyperedge to connect multiple nodes simultaneously (Feng et al., 2019; Gao et al., 2022; Jie et al., 2016; He et al., 2026). Yet in existing neuroimaging applications, hyperedges are typically constructed from predefined similarities or fixed anatomical/functional groupings, so the higher-order structure is imposed a priori and cannot adapt to individual variability or changes in brain state. One might instead learn higher-order interactions freely from data, but unrestricted adaptivity reintroduces the reliability concern raised above: without anatomical grounding, a surrogate can fit the observations while responding implausibly to perturbations. Structural connectivity alone cannot resolve this tension either, because observed functional activity is not a simple readout of anatomy but reflects the joint influence of structural organization and distributed network dynamics (Honey et al., 2009; Popp et al., 2025; Bassett & Sporns, 2017; Xia et al., 2025). What is needed, then, is a surrogate in which individualized higher-order context adaptively re-weights anatomically supported pathways, thereby gaining subject- and state-specific flexibility without surrendering anatomical plausibility.

We introduce **SI-HGN** (Structure-Informed Hypergraph Network), a framework for individualized brain-dynamics modeling and perturbation-based EC inference. Its central design principle is to use structural connectivity (SC) as an anatomical scaffold for learning structured brain representations. SI-HGN predicts future BOLD activity from historical observations through three complementary components. An SC-supported graph branch encodes anatomical structure. Learned latent hyperedges capture multi-region context and generate attention weights that modulate propagation along SC-supported edges. A global prediction head then integrates the resulting regional representations across the whole brain. In this way, SI-HGN combines anatomically informed graph propagation with higher-order regional context and flexible whole-brain integration.

After training, SI-HGN is used as an individualized surrogate brain for virtual perturbation. We systematically perturb each region and quantify the resulting changes in predicted activity across the brain, yielding a directed estimate of regional influence, as illustrated in Figure 1. Because EC is derived from the response of the learned surrogate rather than directly from model parameters, the inferred interactions reflect how regional perturbations propagate through the learned brain dynam-

ics. This perturbation–response formulation provides a unified way to translate structure-informed brain-dynamics modeling into whole-brain directed EC inference.

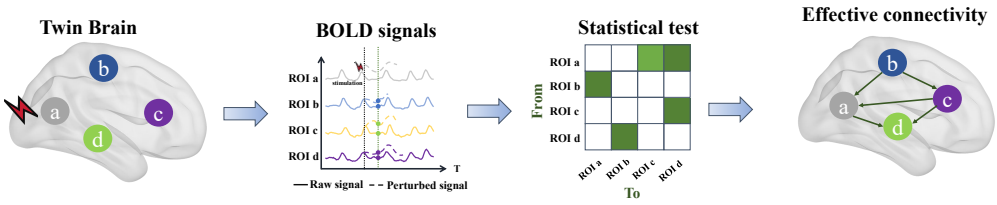

Figure 1: **Virtual-perturbation-based effective connectivity inference.**A virtual perturbation is applied to one region in the trained Surrogate Brain, and the predicted responses are compared with those obtained without perturbation. Response differences are aggregated across BOLD windows to quantify the directed influence of the perturbed region on other regions. Repeating the procedure for all regions yields a whole-brain directed effective connectivity matrix.

The central tenet is to let higher-order hypergraph context gate, rather than supplant, anatomical pathways: structural connectivity defines the support of graph-based message passing, whereas learned latent hyperedges supply the context that adaptively re-weights individual structural edges. The principal contributions of this work are fourfold:

- **A structure-informed surrogate-brain formulation for EC inference.** We incorporate structural connectivity within the learned surrogate itself, so that graph-based propagation is grounded in anatomically plausible pathways while retaining flexible whole-brain integration.
- **Hypergraph-guided adaptive structural modeling.** We introduce learnable latent hyperedges whose derived edge attention re-weights structural pathways in a subject- and state-specific manner, without introducing unsupported edges in the graph-propagation branch.
- **Multi-level validation spanning generative, animal, and human benchmarks.** We benchmark SI-HGN against neural perturbational inference and Granger causality on four complementary datasets, attaining markedly improved EC recovery (Pearson $r = 0.94$ vs. 0.91 on RNN systems; $p < 0.001$).
- **Preliminary evidence for disease-related individualized EC.** We demonstrate, on ADNI, ABIDE, and a DoC cohort, that perturbation-derived EC carries discriminative information for Alzheimer's disease classification, autism screening, and network-disruption characterization in disorders of consciousness.

2 Method

In this section, we detail the SI-HGN framework, with the overall architecture illustrated in Figure 2. The pipeline consists of four stages: (1) formulating the dynamics prediction task over historical BOLD windows (Section 2.1); (2) constructing the SI-HGN surrogate brain, where latent hyperedges capture higher-order regional interactions and hypergraph-derived attention adaptively modulates structurally supported pathways (Section 2.2); (3) jointly optimizing next-step prediction and functional-connectivity consistency objectives (Section 2.2); and (4) performing virtual perturbations on the frozen surrogate to infer directed effective connectivity for downstream analysis (Section 2.3).

2.1 Problem Formulation

Let $\mathcal{B} = (\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(M)})$ denote a batch of M BOLD windows, where each window contains L historical time points from N brain regions, and let $\mathbf{S} \in \mathbb{R}^{N \times N}$ denote structural connectivity (SC). SI-HGN learns a surrogate of brain dynamics that predicts the next whole-brain BOLD observation from historical activity and individualized structural connectivity:

$$\widehat{\mathbf{Y}} = F_\theta(\mathcal{B}, \mathbf{S}) \in \mathbb{R}^{M \times N}. \tag{1}$$

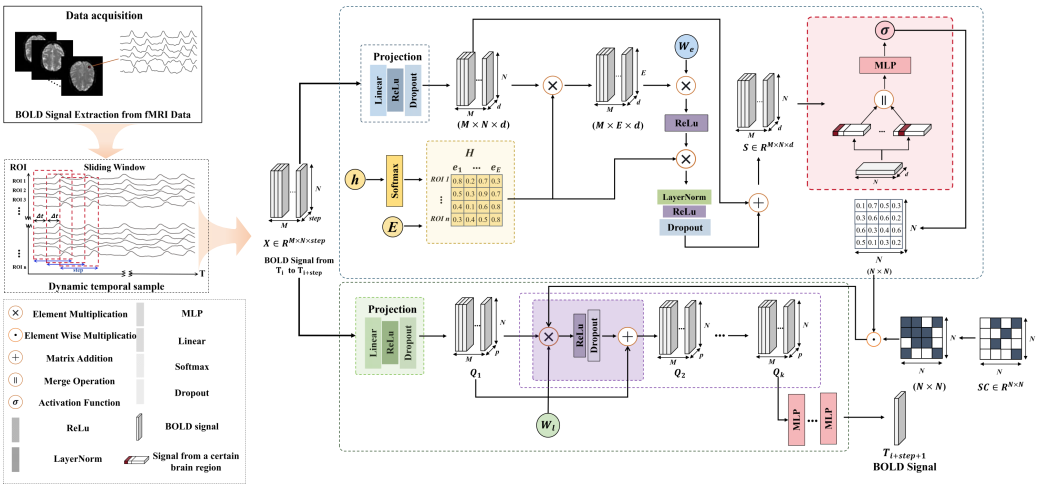

Figure 2: **Overview of SI-HGN.**A learnable hypergraph captures higher-order regional interactions and generates edge-level attention to modulate structural connectivity. Residual graph convolution and a global prediction head predict the next whole-brain BOLD observation. The trained model is then used for virtual-perturbation-based EC inference.

The model combines higher-order regional context with structure-informed propagation to learn individualized brain dynamics, which are subsequently queried through virtual perturbations to estimate directed effective connectivity.

2.2 SI-HGN Surrogate Brain

Higher-order representation. For each BOLD window, regional activity histories are projected into a latent feature space:

$$\mathbf{Z}^{(m)} = \rho_h((\mathbf{X}^{(m)})^\top) \in \mathbb{R}^{N\times d_h}. \tag{2}$$

To capture coordinated interactions among multiple regions, SI-HGN learns E latent hyperedges. The membership of region i in hyperedge e is defined as:

$$H_{i,e} = \frac{\exp(h_{i,e})}{\sum_{r=1}^{N}\exp(h_{r,e})}, \qquad \mathbf{H} \in \mathbb{R}^{N\times E}. \tag{3}$$

The learned incidence matrix supports node–hyperedge–node aggregation:

$$\mathbf{Z}_h^{(m)} = \mathrm{Skip}(\mathbf{Z}^{(m)}) + \mathcal{R}_h\Big(\mathbf{H}\phi_e(\mathbf{H}^\top\mathbf{Z}^{(m)})\Big), \tag{4}$$

where ϕ_e denotes a learnable hyperedge transformation and $\mathcal{R}_h = \mathrm{Dropout}\circ\mathrm{ReLU}\circ\mathrm{LN}$. The operation $\mathbf{H}^\top\mathbf{Z}^{(m)}$ aggregates regional features into latent hyperedge representations, which are transformed and redistributed to individual regions through $\mathbf{H}$. The resulting representations encode multi-region context for adapting structural pathways.

Hypergraph-guided adaptive structural modeling. We normalize the structural connectivity matrix as:

$$\widehat{\mathbf{S}} = \mathbf{D}^{-1/2}(|\mathbf{S}| + \mathbf{I})\mathbf{D}^{-1/2}, \qquad D_{ii} = \sum_j (|S_{ij}| + I_{ij}). \tag{5}$$

The hypergraph-derived regional representations are used to estimate the contribution of each pairwise structural pathway:

$$a_{ij}^{(m)} = \sigma\Big(\mathrm{MLP}_{\mathrm{att}}\Big[\mathbf{z}_{h,i}^{(m)} \| \mathbf{z}_{h,j}^{(m)}\Big]\Big), \tag{6}$$

where σ denotes the sigmoid function and $\|$ denotes concatenation. The attention patterns are aggregated across windows and combined with normalized SC:

$$\mathbf{A}_{\mathrm{adp}}(\mathcal{B}) = \widehat{\mathbf{S}} \odot \left(\frac{1}{M} \sum_{m=1}^{M} \mathbf{A}_{\mathrm{att}}^{(m)} \right), \tag{7}$$

where $\odot$ denotes element-wise multiplication. SC provides the anatomical scaffold for information propagation, while higher-order context adaptively adjusts the contribution of individual structural pathways according to observed brain activity.

Structure-informed whole-brain prediction. The historical activity of each region is projected into graph-node features:

$$\mathbf{Q}_0^{(m)} = \rho_g((\mathbf{X}^{(m)})^\top) \in \mathbb{R}^{N \times d_g}. \tag{8}$$

These regional features are propagated over the adaptive structural connectivity through residual graph layers:

$$\mathbf{Q}_{l+1}^{(m)} = \mathrm{Skip}(\mathbf{Q}_l^{(m)}) + \mathcal{R}_g\Big(\mathbf{A}_{\mathrm{adp}}(\mathcal{B})\mathbf{Q}_l^{(m)}\mathbf{W}_l\Big), \tag{9}$$

where $\mathcal{R}_g$ denotes the nonlinear transformation within each residual graph layer. The resulting regional representations incorporate information propagated through structurally supported pathways modulated by higher-order context. After the final graph layer, these representations are jointly integrated by a global prediction module to estimate the next whole-brain BOLD observation. SI-HGN thereby combines individualized structural connectivity, higher-order multi-region context, and whole-brain integration within a unified surrogate model.

Training objective. For target observations $\mathbf{Y} \in \mathbb{R}^{M \times N}$, the surrogate jointly optimizes next-step prediction accuracy and functional-connectivity consistency:

$$\mathcal{L}(s) = \frac{\lambda_{\mathrm{MSE}}}{MN} \left\| \widehat{\mathbf{Y}} - \mathbf{Y} \right\|_F^2 + \frac{\lambda_{\mathrm{FC}}(s)}{N^2} \left\| \mathrm{Corr}(\widehat{\mathbf{Y}}) - \mathrm{Corr}(\mathbf{Y}) \right\|_F^2. \tag{10}$$

The MSE term minimizes regional prediction error, whereas the FC-consistency term encourages preservation of interregional correlation structure across windows. The FC weight $\lambda_{\mathrm{FC}}(s)$ increases after a warm-up stage, with s denoting the training epoch. Neither reference EC matrices nor diagnostic labels are used during surrogate training; directed connectivity is inferred after the brain-dynamics model has been learned.

2.3 Virtual Perturbation-Based EC Inference

After training, all surrogate parameters are fixed and directed EC is estimated through virtual regional perturbations. Let $\mathcal{P} = (\mathcal{B}_1, \ldots, \mathcal{B}_B)$ contain $M_{\mathrm{EC}} = \sum_b M_b$ inference windows. For each source region i, a perturbation of magnitude δ is applied to its final input time point in every window, producing $\mathcal{B}_b^{(i,\delta)}$. The surrogate evaluates both baseline and perturbed conditions using the same brain-dynamics modeling procedure.

The directed influence from source region i to target region j is defined as:

$$E_{ij}(\delta; \mathcal{P}) = \frac{1}{M_{\mathrm{EC}}} \sum_{b=1}^{B} \sum_{m=1}^{M_b} \left([F_\theta(\mathcal{B}_b^{(i,\delta)}, \mathbf{S})]_{m,j} - [F_\theta(\mathcal{B}_b, \mathbf{S})]_{m,j} \right). \tag{11}$$

Thus, E_{ij} measures the average change in the predicted activity of target region j induced by perturbing source region i. Repeating this procedure for all source regions yields the directed EC matrix $\mathbf{E} \in \mathbb{R}^{N \times N}$, where rows index perturbed source regions and columns index response regions. We set $E_{ii} = 0$. The learned individualized brain dynamics are thereby translated into whole-brain directed effective connectivity through an explicit perturbation–response procedure.

3 Experiments And Results

3.1 Experimental Setup

Baselines and protocol. We compare SI-HGN with NPI and Granger causality (GC), representing neural-surrogate-based perturbation inference and conventional time-series-based directed connec-

tivity estimation, respectively (Luo et al., 2025; Seth et al., 2015). NPI and SI-HGN are evaluated using the same BOLD datasets and data partitions.Pearson correlation between the inferred and reference effective connectivity, denoted by EC_r, is used as the primary recovery metric, with complementary strength- and topology-based metrics reported where appropriate. For experiments evaluating EC_r, results from five independent training runs are first averaged within each independently generated system or simulated subject. Two-sided paired t-tests are then used to assess differences in EC_r between methods.

3.2 Effective Connectivity Validation

3.2.1 Experiment on RNN-Simulated Data

The RNN benchmark consists of 50 independently simulated recurrent systems, each containing 20 nodes and 8,000 BOLD time points (Luo et al., 2025). Each system provides a structural connectivity matrix and a continuous-valued ground-truth EC matrix determined by the underlying recurrent interactions. The explicitly specified directed interactions provide a controlled setting for evaluating the recovery of EC direction and connection strength.SI-HGN achieves the strongest overall EC recovery on the RNN benchmark, consistently outperforming NPI and GC across the reported recovery metrics. At the level of the 50 independently generated recurrent systems, the improvement in EC_r over both NPI and GC is statistically significant (two-sided paired t-tests, $p < 0.001$; Table 1).

Table 1: Quantitative comparison of effective connectivity inference performance on the RNN dataset.

Model	$\mathbf{EC}_r$	$\mathbf{R}^2$	MAE	Spearman	CosSim
GC	0.74 ± 0.02	-117.89 ± 16.15	0.40 ± 0.05	0.73 ± 0.02	0.74 ± 0.02
NPI	0.91 ± 0.01	0.71 ± 0.03	0.02 ± 0.01	0.90 ± 0.01	0.91 ± 0.01
SI-HGN	$\mathbf{0.94} \pm 0.01$	$\mathbf{0.88} \pm 0.02$	$\mathbf{0.02} \pm 0.01$	$\mathbf{0.94} \pm 0.01$	$\mathbf{0.94} \pm 0.01$

3.2.2 Experiment on Whole-Brain Model Data

The whole-brain model (WBM) benchmark contains 66 brain regions, with structural connectivity derived from human diffusion MRI (Luo et al., 2025; Sanz Leon et al., 2013). BOLD signals are generated using a dynamic mean-field model incorporating neural population activity and hemodynamic transformation. A continuous-valued reference EC matrix obtained through virtual stimulation is used to evaluate whole-brain directed connectivity recovery.SI-HGN provides the best overall agreement with the reference EC on the WBM benchmark and better preserves the functional correlation structure of the simulated signals. Its consistent advantages across complementary metrics indicate improved whole-brain connectivity recovery over the evaluated baselines (two-sided paired t-tests, $p < 0.001$; Table 2).

Table 2: Quantitative comparison of effective connectivity inference performance on the WBM dataset.

Model	$\mathbf{EC}_r$	$\mathbf{FC}_r$	MAE	Spearman	CosSim
GC	0.02	N/A	0.13	-0.01	0.02
NPI	0.38 ± 0.01	0.56 ± 0.01	0.05 ± 0.01	0.11 ± 0.01	0.40 ± 0.01
SI-HGN	$\mathbf{0.43} \pm 0.01$	$\mathbf{0.62} \pm 0.08$	$\mathbf{0.01} \pm 0.01$	$\mathbf{0.18} \pm 0.02$	$\mathbf{0.44} \pm 0.01$

Note: N/A indicates that FC_r is not applicable to GC. GC directly estimates directed connectivity and does not generate surrogate-predicted BOLD signals from which model-based functional connectivity can be computed.

3.2.3 Experiment on Macaque Synthetic Data

We use the SmallDegree network from the Feedback-Discovery benchmark (Sanchez-Romero et al., 2019). The network is based on experimentally derived macaque interareal connectivity (Markov et al., 2014) and contains 28 cortical regions and 52 directed connections. The benchmark includes

60 simulated subjects, each with 500 BOLD time points and a corresponding ground-truth directed network. The sparse reference network allows evaluation of both EC agreement and directed connection identification.SI-HGN achieves the strongest overall performance on the macaque benchmark, outperforming NPI and GC in both EC agreement and directed topology recovery. After averaging the five training runs within each simulated subject, the improvement in EC_r over both NPI and GC is statistically significant (two-sided paired t-tests, $p < 0.001$; Table 3).

Table 3: Quantitative comparison of effective connectivity inference performance on the macaque synthetic dataset.

Model	**EC$_r$**	**ACC**	**PRE**	**SEN**	**SPE**	**MCC**
GC	0.48 ± 0.03	0.84 ± 0.01	0.27 ± 0.03	0.76 ± 0.06	0.85 ± 0.01	0.39 ± 0.04
NPI	0.50 ± 0.02	0.86 ± 0.01	0.31 ± 0.02	0.84 ± 0.04	0.86 ± 0.01	0.45 ± 0.03
SI-HGN	$\mathbf{0.58} \pm 0.02$	$\mathbf{0.89} \pm 0.01$	$\mathbf{0.37} \pm 0.02$	$\mathbf{0.92} \pm 0.04$	$\mathbf{0.88} \pm 0.01$	$\mathbf{0.54} \pm 0.03$

3.2.4 Experiment on Human Medial Temporal Lobe Data

We further evaluate SI-HGN on a high-resolution 7-T resting-state fMRI dataset of the human medial temporal lobe (MTL) (Shah et al., 2018; Sanchez-Romero et al., 2019). The dataset contains 23 participants with 421 time points per participant and a repetition time of 1 s. Seven MTL regions are defined separately in the left and right hemispheres. The processed representation adopted by the Feedback-Discovery benchmark provides reference connectivity for evaluating EC recovery in the two hemispheres. SI-HGN achieves the best overall performance on the left MTL network. On the right MTL network, NPI shows a marginally higher EC correlation (0.58 vs. 0.57), whereas SI-HGN performs better on topology-identification metrics, including accuracy, precision, and specificity (Table 4). This slight reduction in EC correlation may be related to the sparser structural scaffold in the right hemisphere (Appendix A.2.4). Overall, SI-HGN provides more balanced EC recovery across hemispheres. Given the small sample size and the use of a group-level consensus reference, the MTL experiment should be interpreted primarily as a biological validation rather than a precise subject-level EC evaluation.

Table 4: Quantitative comparison of effective connectivity inference performance on the real human MTL fMRI dataset.

Task	Model	EC_r	ACC	PRE	SEN	SPE
MTL-L	GC	0.25 ± 0.19	0.53 ± 0.05	0.42 ± 0.22	0.25 ± 0.31	0.76 ± 0.30
	NPI	0.57 ± 0.03	0.69 ± 0.04	0.73 ± 0.10	$\mathbf{0.53} \pm 0.09$	0.82 ± 0.10
	SI-HGN	$\mathbf{0.66} \pm 0.04$	$\mathbf{0.75} \pm 0.03$	$\mathbf{0.89} \pm 0.10$	$\mathbf{0.53} \pm 0.05$	$\mathbf{0.94} \pm 0.05$
MTL-R	GC	0.27 ± 0.17	0.53 ± 0.07	0.58 ± 0.03	0.19 ± 0.03	0.81 ± 0.03
	NPI	$\mathbf{0.58} \pm 0.04$	0.76 ± 0.04	0.89 ± 0.10	0.54 ± 0.05	0.94 ± 0.05
	SI-HGN	0.57 ± 0.02	$\mathbf{0.78} \pm 0.02$	$\mathbf{0.93} \pm 0.04$	$\mathbf{0.56} \pm 0.04$	$\mathbf{0.97} \pm 0.02$

3.3 Individualized Clinical Analysis

3.3.1 Experiment on ADNI Dataset

The final Alzheimer's Disease Neuroimaging Initiative (ADNI) study cohort comprises 49 participants, including 19 patients with Alzheimer's disease (AD), 12 participants with mild cognitive impairment (MCI), and 18 cognitively normal controls (CN) (Petersen et al., 2010; Weiner et al., 2010). All included participants have resting-state fMRI, diffusion MRI, and diagnostic information. Resting-state fMRI is parcellated into 116 regions using the AAL116 atlas (Tzourio-Mazoyer et al., 2002), while diffusion MRI is used to derive subject-specific structural connectivity. Individualized EC is subsequently evaluated in AD vs. CN, AD vs. MCI, and MCI vs. CN classification. SI-HGN achieves the highest classification accuracy across all three ADNI tasks. Although performance varies across individual classification metrics, the results suggest that individualized EC contains disease-related information with discriminative value for subject-level analysis (Table 5).

Table 5: Quantitative comparison of classification performance on the ADNI dataset.

Metric	AD vs. CN			AD vs. MCI			MCI vs. CN		
	GC	NPI	SI-HGN	GC	NPI	SI-HGN	GC	NPI	SI-HGN
ACC(%) ↑	54.32	56.59	**60.00**	53.39	60.32	**63.87**	55.17	58.00	**59.00**
PRE(%) ↑	57.14	57.98	**60.42**	62.27	**68.86**	68.54	41.94	47.62	**49.44**
SEN(%) ↑	50.00	53.95	**63.68**	62.11	65.53	**77.11**	35.00	**42.92**	35.42
SPE(%) ↑	58.89	**59.17**	56.11	39.58	**52.08**	42.92	68.61	68.06	**74.72**

3.3.2 Experiment on ABIDE-Yale Dataset

The final ABIDE-Yale cohort comprises 40 participants from the Yale site, including 22 individuals with autism spectrum disorder (ASD) and 18 typically developing controls (Di Martino et al., 2014). Resting-state fMRI data are parcellated according to the AAL116 atlas to extract regional BOLD time series. Subject-specific morphological features are derived from T1-weighted structural MRI and used to construct individualized morphological similarity networks (Tzourio-Mazoyer et al., 2002; Kong et al., 2015; Seidlitz et al., 2018). The resulting individualized EC matrices are then used for ASD-versus-control classification. SI-HGN achieves the strongest performance in accuracy, AUC, and precision, while GC performs better in F1 score. Given the limited cohort size, these results provide preliminary evidence that individualized EC captures brain-network differences associated with ASD (Table 6).

Table 6: Quantitative comparison of classification performance on the ABIDE-Yale dataset.

Model	**ACC(%) ↑**	**AUC(%) ↑**	**PRE(%) ↑**
GC	55.00	50.00	55.00
NPI	52.50	39.42	53.43
SI-HGN	**57.50**	**57.83**	**59.33**

3.3.3 Experiment on DoC Cohort

We further evaluate SI-HGN in a hospital-based cohort of 39 patients with DoC. Patient-specific EC is estimated from individual MRI-derived features and compared with a normative EC template derived from healthy controls. EC deviations are further examined in relation to CRS-R scores and clinically identified injury-related regions. DoC patients show reduced EC strength relative to healthy controls, indicating disrupted directed brain-network organization associated with impaired consciousness (Figure 3). Additional interpretation is provided in Appendix A.4.1.

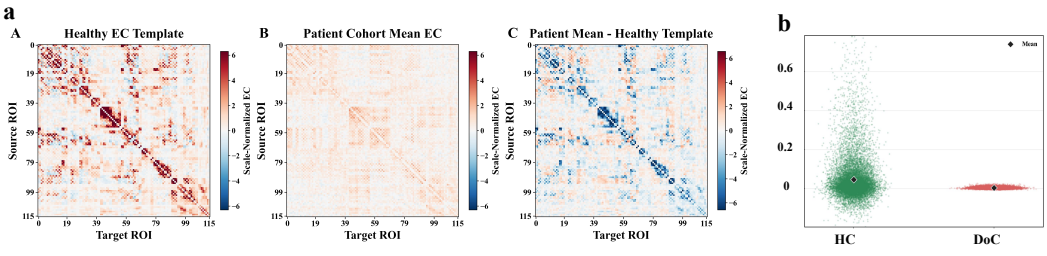

Figure 3: **EC alterations in DoC.** (a) Normative EC in healthy controls, mean EC in DoC patients, and the patient-minus-control difference. (b) Distributions of EC values in healthy controls and DoC patients. DoC patients exhibit a compressed EC distribution, fewer high-strength connections, and lower overall EC values.

3.3.4 SUMMARY OF CLINICAL EVIDENCE

Taken together, the results across ADNI, ABIDE-Yale, and the DoC cohort form a graded pattern of evidence. SI-HGN's advantage is largest and statistically significant on benchmarks with explicitly specified ground-truth connectivity, remains consistent on biological networks with tracer-derived or consensus references, and becomes modest and exploratory on small clinical cohorts where neither ground-truth EC nor external validation is available. The clinical classification gains (e.g., 60.0% vs. 56.6% accuracy in AD vs. CN) are directionally consistent across tasks but should be interpreted as hypothesis-generating rather than diagnostic evidence, and they motivate the larger multicenter validation discussed in Section 4.

3.4 ABLATION STUDY

We evaluate the contributions of structural connectivity, hypergraph modeling, and the nonlinear prediction module by removing each component from the full SI-HGN. All variants use identical data partitions, training settings, and EC inference procedures, with EC correlation used as the primary metric.

Table 7: Ablation study of SI-HGN based on effective connectivity recovery performance. All values are reported as percentages (%).

Dataset	Full SI-HGN	Ablation-SC	Ablation-HGNN	Ablation-MLP
WBM	**43.69**	35.79	36.28	29.94
Macaque	**57.94**	56.99	57.43	15.74
MTL-L	**66.14**	43.72	51.66	51.55
MTL-R	**57.32**	44.96	50.92	51.66

The complete SI-HGN achieves the best EC recovery across all reported benchmarks. Removing the prediction module produces the largest degradation on WBM and macaque, while structural connectivity and hypergraph modeling make larger contributions on the MTL networks. These results support complementary roles for anatomical constraints, higher-order representation, and nonlinear prediction in SI-HGN (Table 7).

4 CONCLUSION

We presented SI-HGN, a structure-informed hypergraph network for individualized surrogate-brain modeling and directed EC inference from BOLD dynamics. Structural connectivity guides graph propagation, latent hyperedges capture multi-region context, and regional perturbations query the frozen surrogate to estimate directed EC.

Methodological implications: Beyond the specific architecture, our results suggest a general principle for building perturbation-queryable surrogates: anatomical priors help most when they constrain where information can flow rather than how much. SC restricts propagation to anatomically supported pathways, while higher-order context adapts the strengths of these pathways within this support. The right MTL network delineates the failure regime of this principle: when the scaffold is sparse or poorly differentiated, the prior adds bias without reducing variance, and a structure-free surrogate may remain preferable. Cohorts with degraded structural data may thus require softer or group-level priors. Finally, EC inference requires only 2N forward passes through a lightweight single-layer surrogate, keeping whole-brain inference practical at standard parcellation scales.

Limitations and outlook: Clinical validation remains limited by the absence of subject-level ground-truth EC, small cohorts, and the lack of external replication; disease-related patterns should therefore be regarded as model-derived network representations rather than established biomarkers. Future work will validate inferred EC against stimulation or invasive recordings, extend evaluation to larger multicenter and longitudinal cohorts, and explore multimodal integration for more faithful surrogate-brain modeling and individualized response simulation.

ETHICS STATEMENT

This study uses a combination of publicly available neuroimaging datasets and a hospital-based cohort of patients with disorders of consciousness (DoC). The ADNI and ABIDE datasets were accessed and analyzed in accordance with their respective data-use policies and ethical requirements. For the hospital-based DoC cohort, all data were collected and handled in accordance with the ethical standards of the responsible institutional review board or ethics committee, and all analyses were performed on de-identified data. Appropriate informed consent was obtained from participants or their legally authorized representatives where required by the approved study protocol.

The present work focuses on computational modeling and retrospective analysis of neuroimaging data. The proposed method does not directly determine clinical diagnosis or treatment decisions, and the individualized effective-connectivity estimates should be interpreted as model-derived research outputs rather than validated clinical biomarkers. Particular care was taken to avoid over-interpreting results from the relatively small clinical cohorts. Future clinical use of the proposed framework would require independent validation, appropriate regulatory oversight, and prospective evaluation of its safety, reliability, and clinical utility.

REPRODUCIBILITY STATEMENT

The SI-HGN implementation is available at `https://anonymous.4open.science/r/SI-HGN-2E6E`. Section 2 specifies the hypergraph representation, batch-averaged adaptive graph, global prediction head, training objective, and perturbational EC estimator. Appendix A.1 provides mathematical details and hyperparameter settings. Data sources, reference connectivity, comparison methods, and evaluation metrics are described in Section 3; additional results are provided in Appendices A.2–A.4.

For EC inference, reproducibility requires matching the input-window collection and ordering, perturbation amplitude and location, and batch partition. Model parameters remain fixed, while the adaptive graph is recomputed separately for the unperturbed and perturbed batches. The implementation's partition rule and window-weighted aggregation are specified in Appendix A.1.4. Reproducing a reported comparison also requires matching signal preprocessing, data splits, model settings, and the unit over which results are aggregated. In particular, repeated training runs characterize optimization variability and should be distinguished from independently generated systems or individual participants.

AI USE STATEMENT

Generative AI tools were used to assist with research ideation, language polishing and translation, code development and debugging, and the refinement of figures and manuscript structure. This assistance included drafting alternative formulations, improving the clarity of technical descriptions, and helping identify inconsistencies between mathematical notation and implementation. AI-assisted suggestions were treated as material for author evaluation, not as independent scientific evidence.

The research questions, methodological choices, experimental settings, data analysis, and interpretation of results remained the responsibility of the authors. AI-assisted text was reviewed and revised by the authors. AI-generated or AI-assisted code was inspected, executed, debugged, and tested within the research workflow, and proposed research ideas were evaluated against the literature, mathematical reasoning, and experimental evidence. The use of AI assistance does not itself establish the correctness of an implementation, the reproducibility of a result, or the validity of a scientific claim.

Generative AI tools were not used to fabricate experimental data, supply numerical results in place of executed analyses, or make final scientific decisions. The authors take full responsibility for the manuscript and associated research outputs, including the reported methods, results, citations, code, figures, and conclusions.

REFERENCES

Danielle S Bassett and Olaf Sporns. Network neuroscience. *Nature neuroscience*, 20(3):353–364, 2017.

Dongdong Chen, Mengjun Liu, Sheng Wang, Zheren Li, Lu Bai, Qian Wang, Dinggang Shen, and Lichi Zhang. Guiding fusion of dynamic functional and effective connectivity in spatio-temporal graph neural network for brain disorder classification. *Knowledge-Based Systems*, 309:112856, 2025.

Gopikrishna Deshpande, Krish Sathian, and Xiaoping Hu. Effect of hemodynamic variability on granger causality analysis of fmri. *Neuroimage*, 52(3):884–896, 2010.

Adriana Di Martino, Chao-Gan Yan, Qingyang Li, Erin Denio, Francisco X Castellanos, Kaat Alaerts, Jeffrey S Anderson, Michal Assaf, Susan Y Bookheimer, Mirella Dapretto, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 19(6):659–667, 2014.

Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 3558–3565, 2019. doi: 10.1609/aaai.v33i01.33013558.

Emily S Finn, Xilin Shen, Dustin Scheinost, Monica D Rosenberg, Jessica Huang, Marvin M Chun, Xenophon Papademetris, and R Todd Constable. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature neuroscience*, 18(11):1664–1671, 2015.

Karl J Friston. Functional and effective connectivity: a review. *Brain Connectivity*, 1(1):13–36, 2011. doi: 10.1089/brain.2011.0008.

Karl J Friston, Lee Harrison, and Will Penny. Dynamic causal modelling. *Neuroimage*, 19(4): 1273–1302, 2003.

Yue Gao, Zizhao Zhang, Haojie Lin, Xibin Zhao, Shaoyi Du, and Changqing Zou. Hypergraph learning: Methods and practices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5):2548–2566, 2022. doi: 10.1109/TPAMI.2020.3039374.

Mengshen He, Jianxin Wang, and Jin Liu. Applications of hypergraph learning for brain disorder diagnosis with neuroimaging: A survey. In *AI in MRI-based Brain Disease Prediction*, pp. 296–327. CRC Press, 2026.

Christopher J Honey, Olaf Sporns, Leila Cammoun, Xavier Gigandet, Jean-Philippe Thiran, Reto Meuli, and Patric Hagmann. Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences*, 106(6):2035–2040, 2009.

Biao Jie, Chong-Yaw Wee, Dinggang Shen, and Daoqiang Zhang. Hyper-connectivity of functional networks for brain disease diagnosis. *Medical image analysis*, 32:84–100, 2016.

Xuan Kan, Hejie Cui, Joshua Lukemire, Ying Guo, and Carl Yang. FBNetGen: Task-aware GNN-based fMRI analysis via functional brain network generation. In *Proceedings of the 5th International Conference on Medical Imaging with Deep Learning*, volume 172 of *Proceedings of Machine Learning Research*, pp. 618–637. PMLR, 2022. URL https://proceedings.mlr.press/v172/kan22a.html.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=SJU4ayYgl.

Xiangzhen Kong, Zhaoguo Liu, Lijie Huang, Xu Wang, Zetian Yang, Guangfu Zhou, Zonglei Zhen, and Jia Liu. Mapping individual brain networks using statistical similarity in regional morphology from MRI. *PLOS ONE*, 10(11):e0141840, 2015. doi: 10.1371/journal.pone.0141840.

Zixiang Luo, Kaining Peng, Zhichao Liang, Shengyuan Cai, Chenyu Xu, Dan Li, Yu Hu, Changsong Zhou, and Quanying Liu. Mapping effective connectivity by virtually perturbing a surrogate brain. *Nature Methods*, 22(6):1376–1385, 2025.

Saqib Mamoon, Zhengwang Xia, Wang Jin, Amani Alfakih, and Jianfeng Lu. Joint dynamic brain network estimation and graph representation learning for the recognition of neurological disorders. *IEEE Journal of Biomedical and Health Informatics*, 2025.

Nikola T Markov, Maria M Ercsey-Ravasz, AR Ribeiro Gomes, Camille Lamy, Loic Magrou, Julien Vezoli, Pierre Misery, Arnaud Falchier, Rene Quilodran, Marie-Alice Gariel, et al. A weighted and directed interareal connectivity matrix for macaque cerebral cortex. *Cerebral cortex*, 24(1): 17–36, 2014.

Sarah Parisot, Sofia Ira Ktena, Enzo Ferrante, Matthew Lee, Ricardo Guerrero, Ben Glocker, and Daniel Rueckert. Disease prediction using graph convolutional networks: application to autism spectrum disorder and alzheimer's disease. *Medical image analysis*, 48:117–130, 2018.

Ronald Carl Petersen, Paul S Aisen, Laurel A Beckett, Michael C Donohue, Anthony Collins Gamst, Danielle J Harvey, Clifford R Jack Jr, William J Jagust, Leslie M Shaw, Arthur W Toga, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology*, 74 (3):201–209, 2010. doi: 10.1212/WNL.0b013e3181cb3e25.

Johanna L Popp, Jonas A Thiele, Joshua Faskowitz, Caio Seguin, Olaf Sporns, and Kirsten Hilger. Structural-functional brain network coupling during cognitive demand reveals intelligence-relevant communication strategies. *Communications Biology*, 8(1):855, 2025.

Ruben Sanchez-Romero, Joseph D Ramsey, Kun Zhang, Madelyn RK Glymour, Biwei Huang, and Clark Glymour. Estimating feedforward and feedback effective connections from fMRI time series: Assessments of statistical methods. *Network Neuroscience*, 3(2):274–306, 2019.

Paula Sanz Leon, Stuart A Knock, M Marmaduke Woodman, Lia Domide, Jochen Mersmann, Anthony R McIntosh, and Viktor Jirsa. The virtual brain: a simulator of primate brain network dynamics. *Frontiers in neuroinformatics*, 7:10, 2013.

Jakob Seidlitz, František Váša, Maxwell Shinn, Rafael Romero-Garcia, Kirstie J Whitaker, Petra E Vértes, Konrad Wagstyl, Paul Kirkpatrick Reardon, Liv Clasen, Siyuan Liu, et al. Morphometric similarity networks detect microscale cortical organization and predict inter-individual cognitive variation. *Neuron*, 97(1):231–247.e7, 2018. doi: 10.1016/j.neuron.2017.11.039.

Anil K Seth, Adam B Barrett, and Lionel Barnett. Granger causality analysis in neuroscience and neuroimaging. *Journal of Neuroscience*, 35(8):3293–3297, 2015.

Preya Shah, Danielle S Bassett, Laura EM Wisse, John A Detre, Joel M Stein, Paul A Yushkevich, Russell T Shinohara, John B Pluta, Elijah Valenciano, Molly Daffner, et al. Mapping the structural and functional network architecture of the medial temporal lobe using 7t mri. *Human Brain Mapping*, 39(2):851–865, 2018.

Xilin Shen, Emily S Finn, Dustin Scheinost, Monica D Rosenberg, Marvin M Chun, Xenophon Papademetris, and R Todd Constable. Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *nature protocols*, 12(3):506–518, 2017.

Stephen M Smith, Karla L Miller, Gholamreza Salimi-Khorshidi, Matthew Webster, Christian F Beckmann, Thomas E Nichols, Joseph D Ramsey, and Mark W Woolrich. Network modelling methods for fmri. *Neuroimage*, 54(2):875–891, 2011.

Jingye Tang, Tianqing Zhu, Wanlei Zhou, and Wei Zhao. Graph neural networks for fmri functional brain networks: A survey. *Neural Networks*, pp. 108137, 2025.

Bishal Thapaliya, Robyn Miller, Jiayu Chen, Yu Ping Wang, Esra Akbas, Ram Sapkota, Bhaskar Ray, Pranav Suresh, Santosh Ghimire, Vince D Calhoun, et al. Dsam: A deep learning framework for analyzing temporal and spatial dynamics in brain networks. *Medical image analysis*, 101: 103462, 2025.

Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Octave Etard, Nicolas Delcroix, Bernard Mazoyer, and Marc Joliot. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, 15(1):273–289, 2002.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=rJXMpikCZ.

Simon Wein, Wilhelm M Malloni, Ana Maria Tomé, Sebastian M Frank, G-I Henze, Stefan Wüst, Mark W Greenlee, and Elmar W Lang. A graph neural network framework for causal inference in brain networks. *Scientific reports*, 11(1):8061, 2021.

Michael W Weiner, Paul S Aisen, Clifford R Jack Jr, William J Jagust, John Q Trojanowski, Leslie Shaw, Andrew J Saykin, John C Morris, Nigel Cairns, Laurel A Beckett, et al. The Alzheimer's Disease Neuroimaging Initiative: progress report and future plans. *Alzheimer's & Dementia*, 6 (3):202–211, 2010. doi: 10.1016/j.jalz.2010.03.007.

Jing Xia, Yi Hao Chan, Deepank Girish, and Jagath C Rajapakse. Interpretable modality-specific and interactive graph convolutional network on brain functional and structural connectomes. *Medical Image Analysis*, 102:103509, 2025.

Yinuo Zhang, Demao Liu, Zhichao Liang, Jiani Cheng, Kexin Lou, Jinqiao Duan, Ting Gao, Bin Hu, and Quanying Liu. Artificial intelligence as a surrogate brain: bridging neural dynamical models and data. *National Science Review*, 13(3):nwaf457, 2026.

A APPENDIX

Appendix includes the following sections:

- Sec. A.1: Additional mathematical details.
- Sec. A.2: Additional experimental results.
- Sec. A.3: Additional clinical classification results.
- Sec. A.4: Additional analysis of disorders of consciousness.

A.1 ADDITIONAL MATHEMATICAL DETAILS

This section provides additional mathematical details for the hypergraph propagation, adaptive structural connectivity, functional-connectivity regularization, and virtual-perturbation-based EC inference introduced in Section 2.

A.1.1 NODE–HYPEREDGE–NODE PROPAGATION

For the m-th input window, let $\mathbf{Z}^{(m)} \in \mathbb{R}^{N\times d_h}$ denote the projected regional features. The incidence matrix $\mathbf{H} \in \mathbb{R}^{N\times E}$ is shared across input windows and is learned jointly with the surrogate model. Equation 3 applies a softmax over the node dimension, so each latent hyperedge satisfies:

$$H_{i,e} > 0, \qquad \sum_{i=1}^{N} H_{i,e} = 1. \tag{12}$$

Thus, $H_{i,e}$ represents the normalized membership of region i in latent hyperedge e. The node-to-hyperedge aggregation is defined as:

$$\mathbf{u}_e^{(m)} = \sum_{i=1}^{N} H_{i,e}\mathbf{z}_i^{(m)}, \qquad \mathbf{U}^{(m)} = \mathbf{H}^\top \mathbf{Z}^{(m)} \in \mathbb{R}^{E\times d_h}. \tag{13}$$

Each row of $\mathbf{U}^{(m)}$ therefore represents a weighted aggregation of regional features associated with one latent hyperedge. The transformed hyperedge features are subsequently propagated back to the regional nodes:

$$\overline{\mathbf{Z}}^{(m)} = \mathbf{H}\phi_e\left(\mathbf{H}^\top \mathbf{Z}^{(m)}\right) \in \mathbb{R}^{N\times d_h}. \tag{14}$$

Here, $\phi_e(\cdot)$ denotes the learnable transformation applied in the hyperedge space. The multiplication by $\mathbf{H}^\top$ aggregates node features into hyperedges, whereas multiplication by $\mathbf{H}$ redistributes the transformed hyperedge information back to the nodes.

The induced higher-order coupling becomes explicit when ϕ_e is locally approximated by the identity mapping:

$$\overline{\mathbf{Z}}^{(m)} \approx \mathbf{H}\mathbf{H}^\top \mathbf{Z}^{(m)}, \qquad (\mathbf{H}\mathbf{H}^\top)_{ij} = \sum_{e=1}^{E} H_{i,e}H_{j,e}. \tag{15}$$

The quantity $(\mathbf{H}\mathbf{H}^\top)_{ij}$ is large when regions i and j have strong memberships in the same latent hyperedges. The hypergraph branch can therefore exchange information between regions through shared higher-order memberships rather than only through pairwise edges. The residual connection, layer normalization, nonlinearity, and dropout in Equation 4 are applied to this propagated representation.

A.1.2 ADAPTIVE STRUCTURAL CONNECTIVITY

The structural graph defines the anatomical support for subsequent graph propagation. Let

$$\widetilde{\mathbf{S}} = |\mathbf{S}| + \mathbf{I}, \qquad D_{ii} = \sum_{j=1}^{N} \widetilde{S}_{ij}. \tag{16}$$

The normalized structural weight between regions i and j is defined as:

$$\widehat{S}_{ij} = \frac{|S_{ij}| + \mathbb{I}[i=j]}{\sqrt{D_{ii}D_{jj}}}. \tag{17}$$

For each input window m, the hypergraph representation generates an edge-attention value $a_{ij}^{(m)}$. Let M denote the number of samples used to construct the batch-level adaptive graph. The adaptive structural weight is defined as:

$$A_{\mathrm{adp},ij}(\mathcal{B}) = \frac{|S_{ij}| + \mathbb{I}[i=j]}{\sqrt{D_{ii}D_{jj}}} \left(\frac{1}{M} \sum_{m=1}^{M} a_{ij}^{(m)} \right). \tag{18}$$

This expression separates two roles. The normalized structural term $\widehat{S}_{ij}$ determines whether and how strongly a pathway is anatomically supported, while the hypergraph-derived attention term adjusts its contribution according to the learned higher-order representations averaged across the batch. Thus, each window's prediction is conditioned on the other windows through this shared graph. During perturbation inference, the same averaging operation is applied separately to the unperturbed and perturbed batches.

For off-diagonal entries, $S_{ij} = 0$ directly implies:

$$A_{\mathrm{adp},ij} = 0, \qquad i \neq j. \tag{19}$$

Therefore, the attention mechanism cannot create an off-diagonal connection that is absent from the structural scaffold. Before the residual, normalization, and nonlinear operations, the graph-convolution message received by region i is:

$$\left(\mathbf{A}_{\mathrm{adp}}\mathbf{Q}_l\mathbf{W}_l\right)_i = \sum_{j=1}^{N} A_{\mathrm{adp},ij}\mathbf{Q}_{l,j}\mathbf{W}_l. \tag{20}$$

Thus, information from region j contributes to region i according to both anatomical support and the higher-order attention assigned to that pathway.

A.1.3 FUNCTIONAL-CONNECTIVITY CONSISTENCY

Let $\widehat{\mathbf{Y}}, \mathbf{Y} \in \mathbb{R}^{M\times N}$ denote the predicted and target BOLD observations used to compute the training objective. Each column corresponds to one brain region, and the correlation is computed across the M observations.

Let $\widehat{\mathbf{y}}_i^c$ and $\widehat{\mathbf{y}}_j^c$ denote the centered predicted signals of regions i and j. Their predicted Pearson correlation is defined as:

$$\widehat{C}_{ij} = \frac{(\widehat{\mathbf{y}}_i^c)^\top \widehat{\mathbf{y}}_j^c}{\|\widehat{\mathbf{y}}_i^c\|_2 \|\widehat{\mathbf{y}}_j^c\|_2}. \tag{21}$$

The same operation applied to $\mathbf{Y}$ produces the target correlation matrix $\mathbf{C}$. Equation 10 therefore gives the expanded FC consistency loss:

$$\mathcal{L}_{\mathrm{FC}} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\widehat{C}_{ij} - C_{ij} \right)^2 . \tag{22}$$

The MSE term constrains the prediction of individual BOLD observations, whereas the FC term additionally constrains their interregional correlation structure. The two objectives therefore provide complementary supervision for learning the surrogate dynamics.

A.1.4 Batch-Conditioned Virtual Perturbation as a Finite Difference

All surrogate-model parameters remain fixed during EC inference. The ordered inference partition is $\mathcal{P} = (\mathcal{B}_1, \ldots, \mathcal{B}_B)$, where batch b contains M_b windows. For every window in a batch, source region i is perturbed at the final input time point:

$$\widetilde{\mathbf{X}}_b^{(m,i)} = \mathbf{X}_b^{(m)} + \delta \mathbf{e}_L \mathbf{e}_i^{\top}, \tag{23}$$

where $\mathbf{e}_L \in \mathbb{R}^L$ and $\mathbf{e}_i \in \mathbb{R}^N$ are standard basis vectors selecting the final time point and source region i, respectively. This operation is applied to all M_b windows, yielding the perturbed batch $\mathcal{B}_b^{(i,\delta)}$. Each forward pass computes its own batch-averaged graph; no baseline graph is cached or held fixed between the two conditions.

The resulting change in the predicted activity of target region j is defined as:

$$\Delta y_{b,m,j}^{(i)} = [F_\theta(\mathcal{B}_b^{(i,\delta)}, \mathbf{S})]_{m,j} - [F_\theta(\mathcal{B}_b, \mathbf{S})]_{m,j}. \tag{24}$$

For a fixed partition, define $\phi_{b,m,j}(\mathcal{B}) = [F_\theta(\mathcal{B}, \mathbf{S})]_{m,j}$. Where this function is twice continuously differentiable in a neighborhood of the unperturbed batch, a sufficiently small perturbation admits the first-order expansion:

$$\Delta y_{b,m,j}^{(i)} = \delta \sum_{n=1}^{M_b} \frac{\partial \phi_{b,m,j}}{\partial X_{b,L,i}^{(n)}} \bigg|_{\mathcal{B}_b} + \mathcal{O}(\delta^2). \tag{25}$$

The sum over n is required because all windows in the batch are perturbed. Terms with $n \neq m$ describe cross-window dependence mediated by the shared adaptive graph. Summing the responses over $M_{\mathrm{EC}} = \sum_b M_b$ inference windows gives:

$$E_{ij}(\delta; \mathcal{P}) = \frac{\delta}{M_{\mathrm{EC}}} \sum_{b=1}^{B} \sum_{m=1}^{M_b} \sum_{n=1}^{M_b} \frac{\partial \phi_{b,m,j}}{\partial X_{b,L,i}^{(n)}} \bigg|_{\mathcal{B}_b} + \mathcal{O}(\delta^2). \tag{26}$$

Thus, in the small-perturbation limit, E_{ij}/δ approximates a directional sensitivity of the batch-conditioned model under a joint shift of source region i across windows. It is not generally an average of isolated-window derivatives. At the experimental perturbation amplitude, EC is defined by the finite differences in Equation 11; the local expansion is a conditional interpretation, not an assumption needed to compute the estimator. In general,

$$E_{ij} \neq E_{ji}, \tag{27}$$

so the resulting EC matrix is directed. The response includes adaptive graph changes and the global readout, and is not restricted to direct anatomical pathways.

Inference partition. The current inference routine preserves the order of the supplied windows. For networks with $N \leq 100$, all supplied windows form one batch. For $N > 100$, consecutive batches contain at most 500 windows, with the final batch containing the remainder. The same partition is used for the baseline and every source-region perturbation. Responses are averaged over individual windows, so a shorter final batch is weighted by its window count. The resulting estimate is conditional on the supplied window collection and this partition.

A.1.5 HYPERPARAMETER SETTINGS OF SI-HGN.

Table 8: Hyperparameter settings of SI-HGN.

Category	Hyperparameter	Value
Input	Temporal window length L	3
	Prediction horizon T_{pred}	1
Architecture	Hidden dimension	$d_g = \min(16, N), d_h = \min(8, \max(4, \lfloor N/3 \rfloor))$
	Residual GCN layers	1
	Hypergraph layers	1
	Hyperedge number E	$\min(8, \max(3, \lfloor N/4 \rfloor))$
	Hyperedge embedding dimension	64
	Attention heads	4
Training	Optimizer	Adam
	Learning rate	5×10^{-4}
	Weight decay	1×10^{-4}
	Batch size	32
	Epochs	80
	Dropout	0.2
Objective	λ_{MSE}	1.0
	λ_{FC}	0→0.3 (warm-up)
Perturbation	Perturbation strength δ	1.0
	Independent runs	5

A.2 Additional Experimental Results

This appendix provides additional qualitative results corresponding to the quantitative comparisons reported in Section 3. It includes visual comparisons of the reference and inferred effective connectivity networks, network-level representations, and supplementary clinical classification results.

A.2.1 RNN-Simulated Dataset

Table 1 reports the quantitative results on the RNN-simulated dataset. Figure 4 compares the inferred EC values and matrices with the ground-truth EC, while Figure 5 provides complementary network-level visualizations.

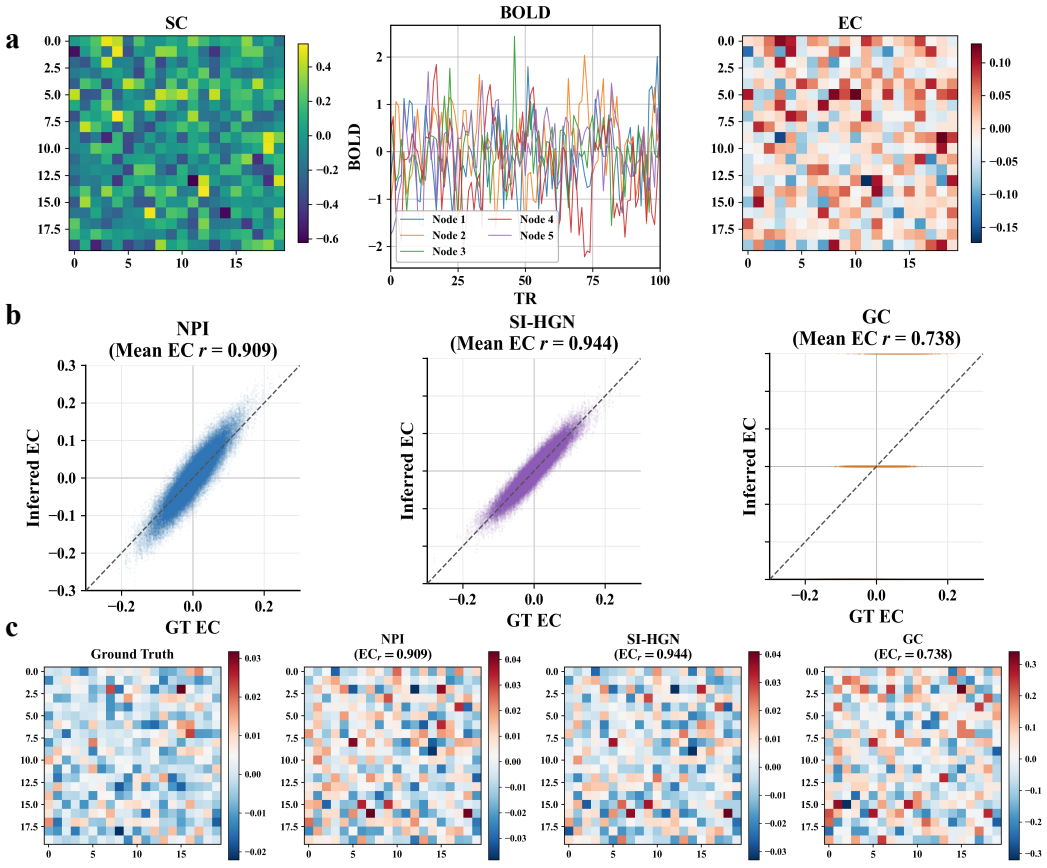

Figure 4: (a) Structural connectivity (SC), representative BOLD time series from selected nodes, and the ground-truth effective connectivity (EC) matrix of the dataset. (b) Scatter plots showing the relationships between the inferred EC and the ground-truth EC for NPI, SI-HGN, and Granger causality (GC), where the horizontal axis denotes the ground-truth EC and the vertical axis denotes the inferred EC. (c) The ground-truth EC matrix and the mean EC matrices inferred by the three methods. Red and blue indicate positive and negative effects, respectively.

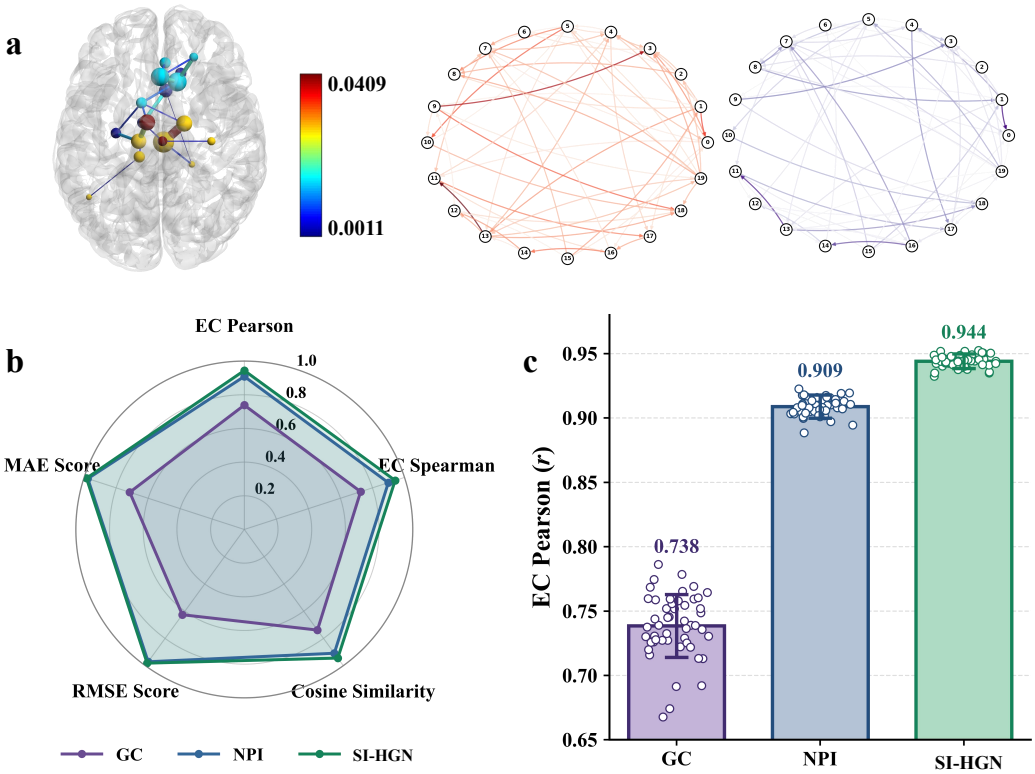

Figure 5: (a) Brain-space visualization and circular network representation of the ground-truth EC, together with a comparison to the directed connectivity topology inferred by SI-HGN. Nodes represent brain regions, and edges represent effective connections between brain regions. (b) Radar chart comparing the three methods across different evaluation metrics. To ensure that all metrics have the same direction, where higher values indicate better performance, the reciprocal values of MAE and RMSE are used. (c) Comparison of the EC correlation coefficients inferred by the three methods.

A.2.2 WHOLE-BRAIN MODEL DATASET

Table 2 reports the quantitative results on the whole-brain model dataset. Figure 6 presents the reference and inferred EC matrices, and Figure 7 shows the corresponding brain-space and lobe-level connectivity patterns.

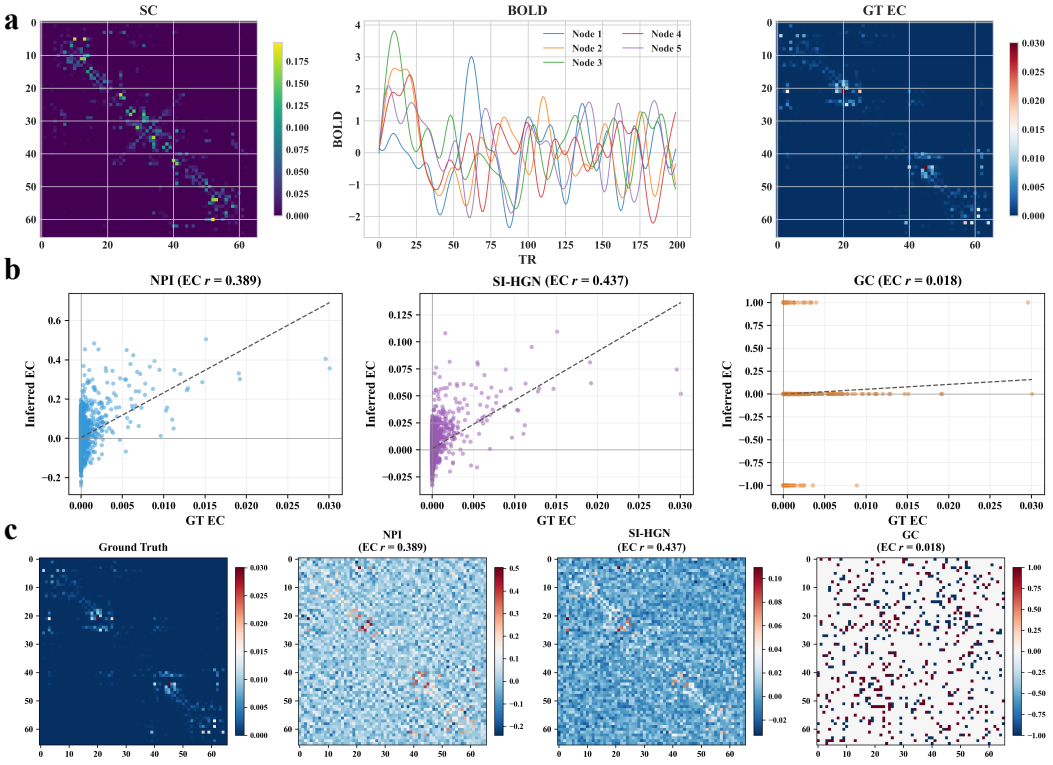

Figure 6: (a) Structural connectivity (SC), representative BOLD time series from selected brain regions, and the ground-truth effective connectivity (GT EC) obtained through virtual stimulation in the WBM dataset. (b) Scatter plots showing the relationships between the EC values inferred by NPI, SI-HGN, and Granger causality (GC) and the ground-truth EC. (c) The ground-truth EC matrix and the whole-brain EC matrices inferred by the three methods, where the colors indicate the direction and strength of the connections.

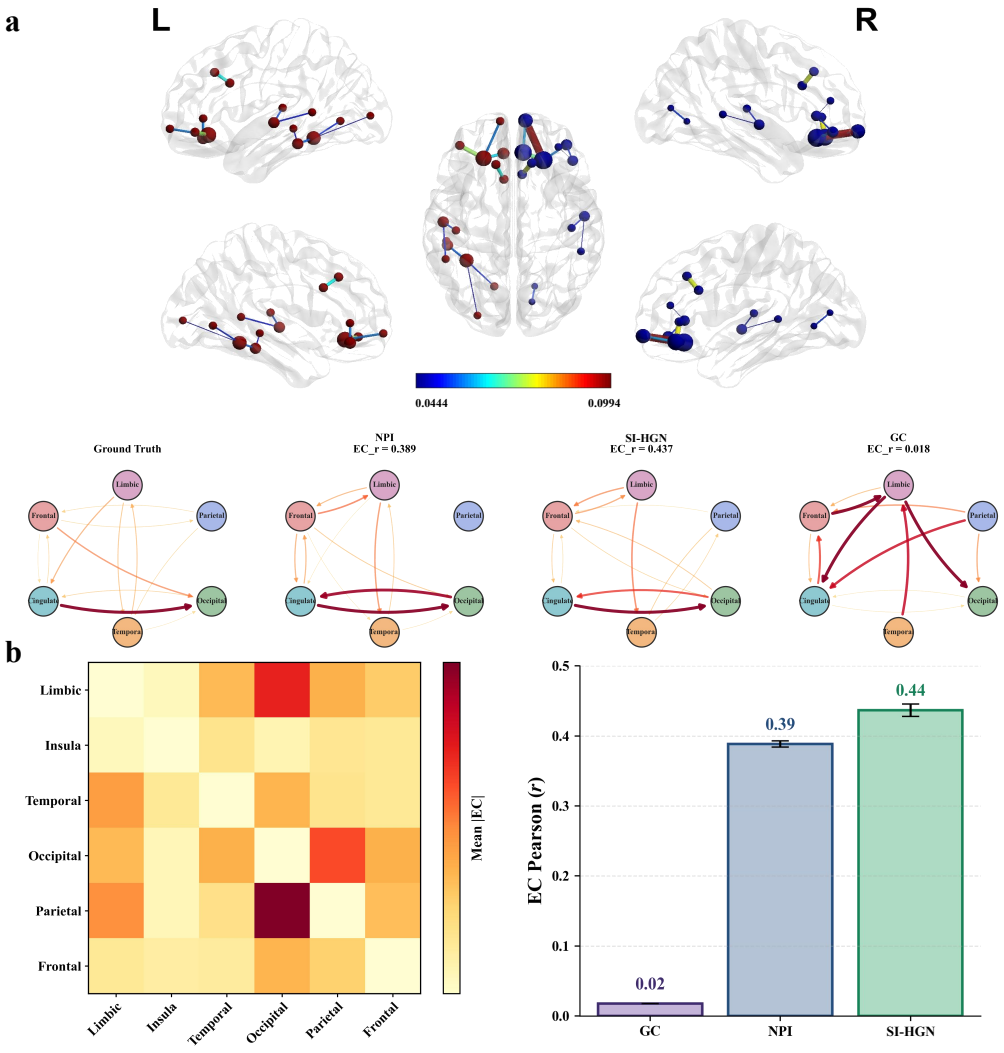

Figure 7: (a) Spatial distributions of effective connectivity in the left and right cerebral hemispheres. Nodes represent brain regions, and edges represent directed interactions between brain regions. The lower panels show the ground-truth network of regions with relatively strong EC, together with the lobe-level aggregated networks inferred by NPI, SI-HGN, and GC. Edge color and width indicate the direction and strength of the connections. (b) Lobe-level EC analysis. The upper panels present the inter-lobe connectivity-strength matrices and the EC Pearson correlation coefficients obtained by the three methods.

A.2.3 MACAQUE SYNTHETIC DATASET

Table 3 reports the quantitative results on the macaque synthetic dataset. Figure 8 compares the inferred matrices with the tracer-derived reference network, while Figure 9 presents the sample-level distributions and network topology comparisons.

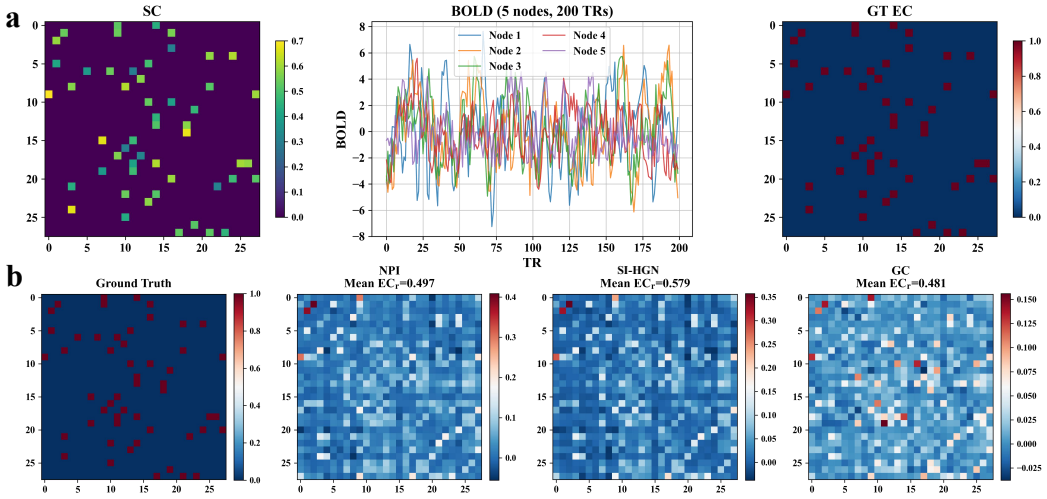

Figure 8: (a) Macaque tracer-derived structural connectivity (SC), representative simulated BOLD time series from selected brain regions, and the binary ground-truth effective connectivity (GT EC). (b) The ground-truth EC matrix and the mean EC matrices inferred by NPI, SI-HGN, and Granger causality (GC), where the colors indicate the direction and strength of the connections.

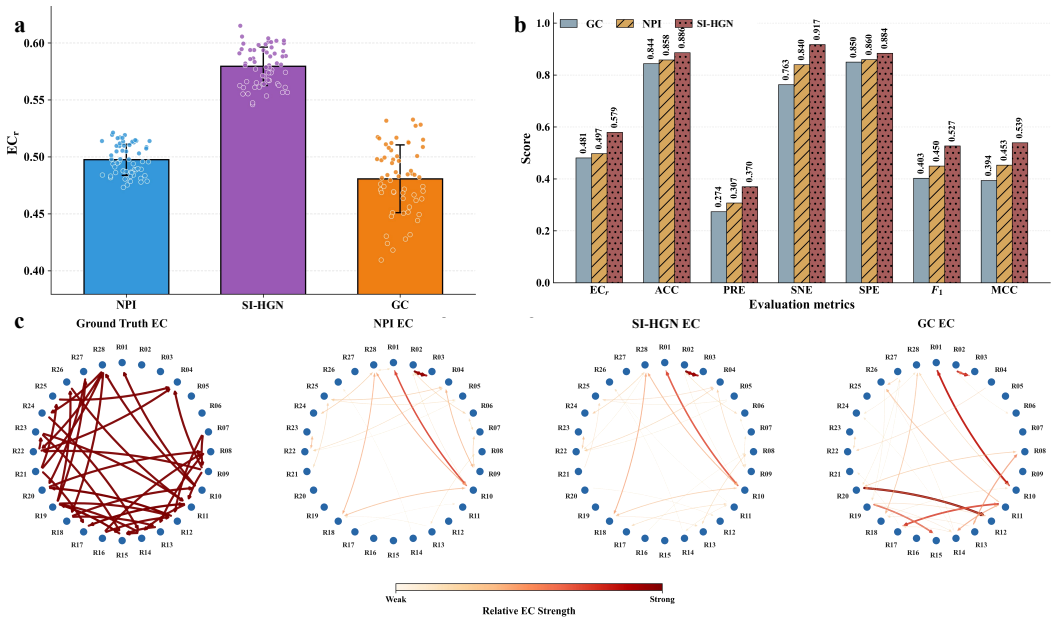

Figure 9: (a) Distributions of EC Pearson correlation coefficients across 60 simulated BOLD samples. The bars and error bars represent the mean and variability, respectively, while the scatter points denote individual independent samples. (b) Comparison of the three methods in terms of EC correlation, accuracy (ACC), precision (PRE), sensitivity (SEN), specificity (SPE), F1 score, and Matthews correlation coefficient (MCC). (c) Circular visualizations of the reference network and the networks inferred by the three methods. Nodes represent brain regions, and edges represent directed effective connections, with edge color and width indicating relative connection strength.

A.2.4 HUMAN MEDIAL TEMPORAL LOBE DATASET

Table 4 reports the quantitative results for the left and right MTL networks. Figure 10 presents the structural covariance, reference EC, and inferred EC matrices. Figure 11 provides additional topology and metric comparisons.

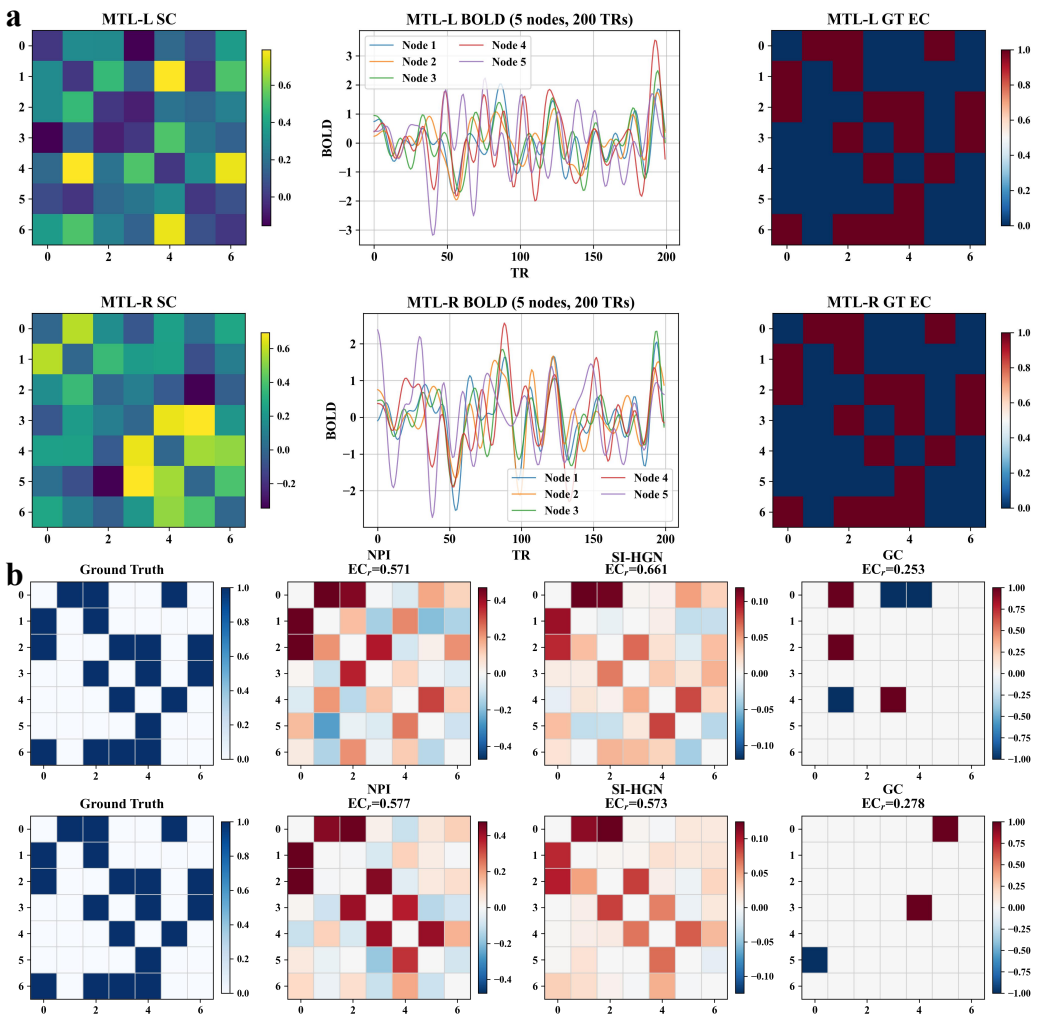

Figure 10: (a) Structural covariance matrices (SCs) of the left- and right-hemispheric medial temporal lobe (MTL) networks, BOLD time series from selected brain regions, and the consensus ground-truth effective connectivity (GT EC). (b) Reference EC matrices for the left and right hemispheres, together with the EC matrices inferred by NPI, SI-HGN, and Granger causality (GC); the Pearson correlation coefficients between the inferred results and the reference networks are reported in the panel titles.

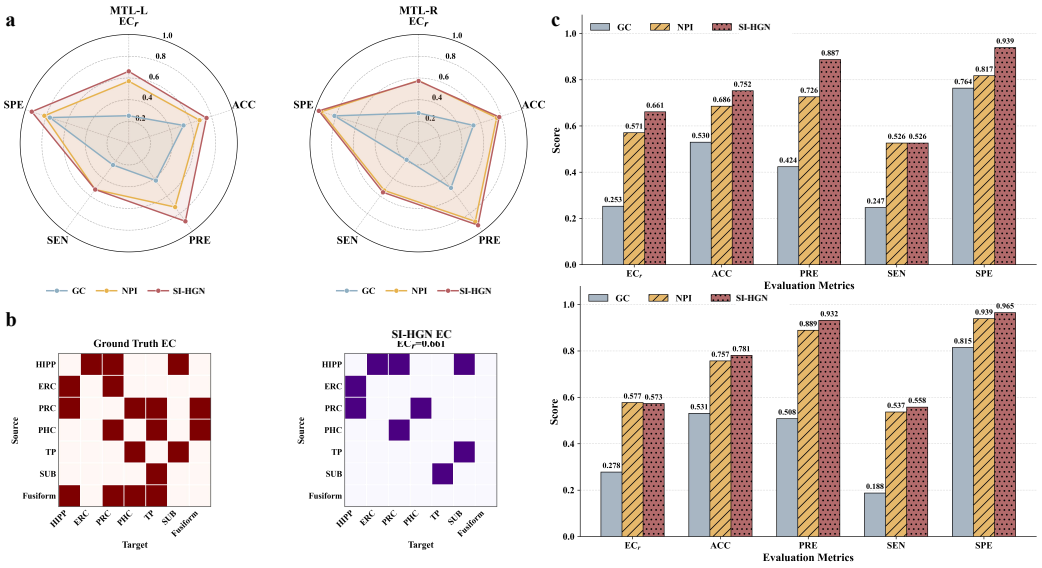

Figure 11: (a) Multi-metric radar plots comparing the three methods in the left and right hemispheres, including EC correlation, accuracy, precision, sensitivity, specificity, F1 score, and Matthews correlation coefficient (MCC). (b) Brain-region-level comparison of the directed topologies of the reference EC and the EC inferred by SI-HGN, where the horizontal and vertical axes represent the target and source brain regions, respectively. (c) Comparison of the three methods in terms of EC Pearson correlation, accuracy (ACC), precision (PRE), sensitivity (SEN), and specificity (SPE) for the left and right hemispheres.

A.3 Additional Clinical Classification Results

A.3.1 ADNI Dataset

Table 5 reports the classification performance for the three ADNI diagnostic comparisons. Figures 12–14 provide the corresponding confusion matrices, repeated nested cross-validation results, and spatial representations of classification-related effective connections.

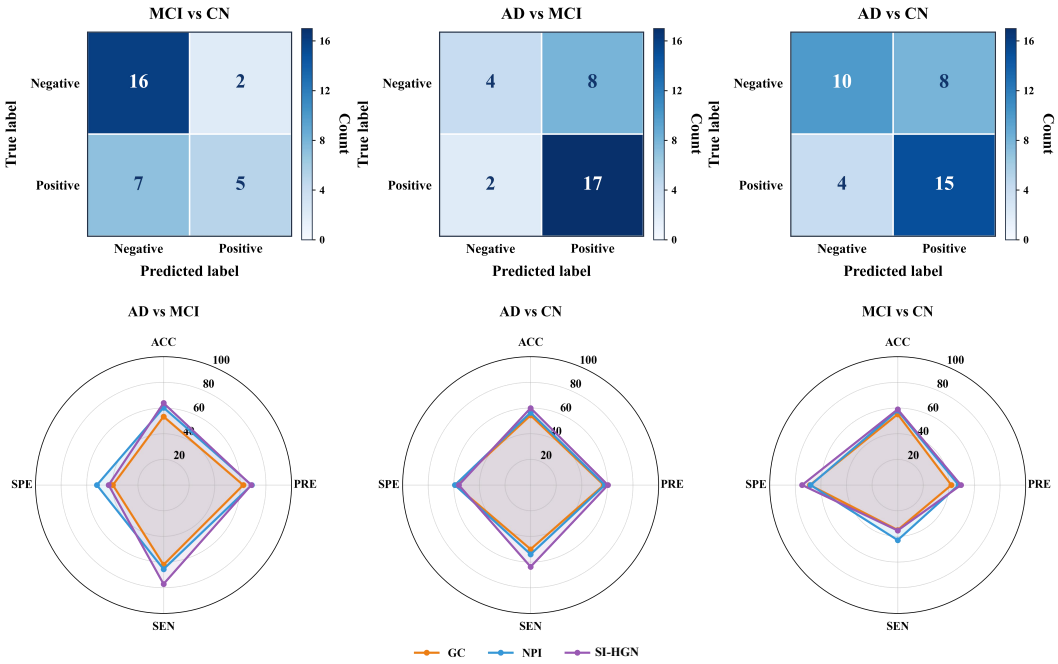

Figure 12: The upper panel shows the confusion matrices of SI-HGN for the three binary classification tasks: MCI vs. cognitively normal controls (CN), AD vs. MCI, and AD vs. CN. The lower panel compares GC, NPI, and SI-HGN across the three classification tasks in terms of accuracy (ACC), precision (PRE), sensitivity (SEN), specificity (SPE), and F1 score.

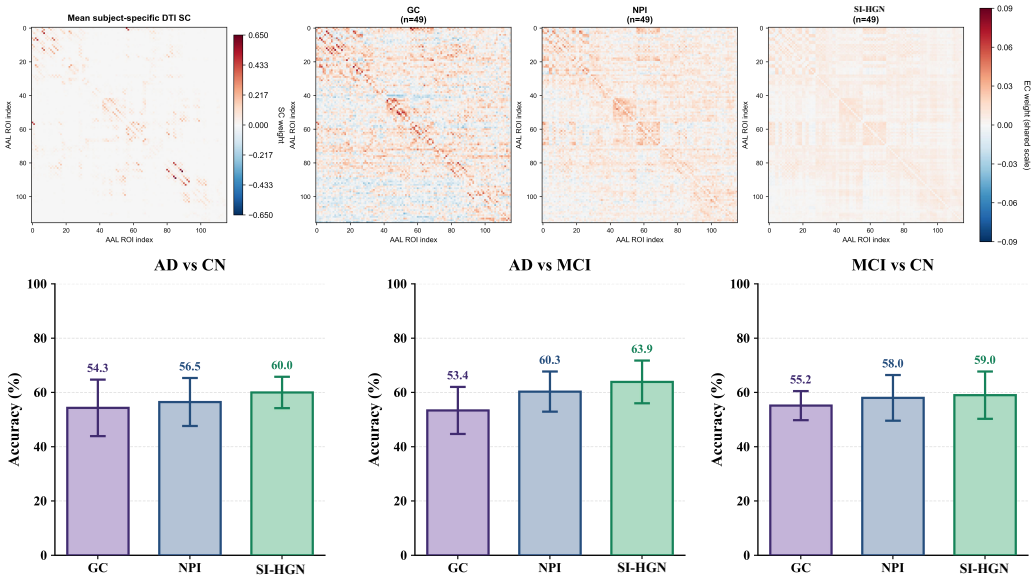

Figure 13: The upper panel shows the spatial distribution of representative classification-related effective connections in brain space, where nodes denote AAL brain regions and edge colors indicate connection strength. The lower panel shows the mean classification accuracy and its variation range obtained from repeated nested cross-validation.

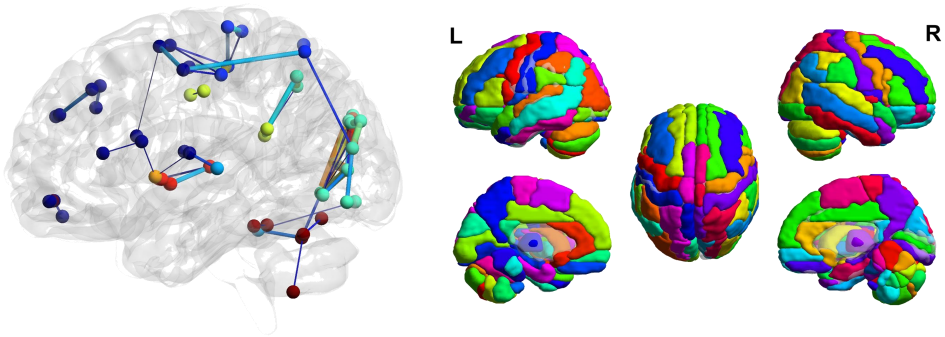

Figure 14: The left panel shows the spatial distribution of representative classification-related effective connections in the brain. Nodes represent AAL brain regions, and edge colors indicate connection strength. The right panel shows the spatial distribution of the AAL116 brain atlas across the left and right hemispheres from different viewing orientations.

A.3.2 ABIDE-YALE DATASET

Table 6 reports the classification results on the ABIDE-Yale dataset. Figure 15 presents the performance comparison, the AAL116 atlas, and the two-dimensional representation of individualized EC features.

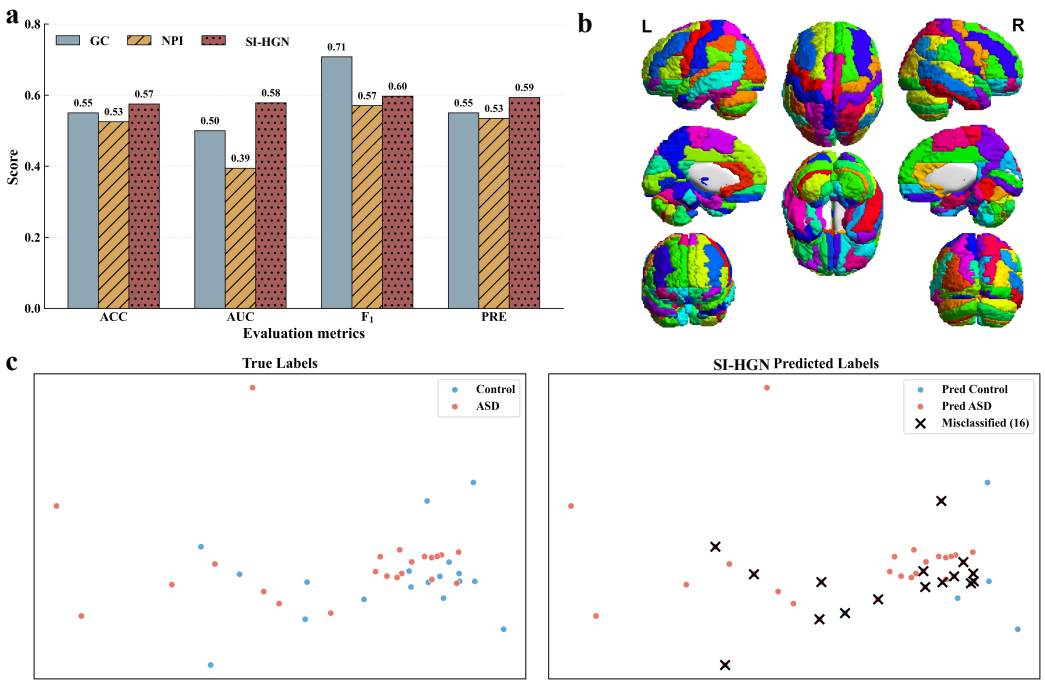

Figure 15: (a) Comparison of GC, NPI, and SI-HGN in terms of accuracy (ACC), area under the ROC curve (AUC), F1 score, and precision (PRE) for the classification of autism spectrum disorder (ASD) versus healthy controls (HCs). (b) Spatial distribution of the AAL116 brain atlas across the left and right hemispheres and from multiple viewing angles, used to define the brain-region nodes for BOLD signal extraction and effective connectivity analysis. (c) Two-dimensional distribution of individual effective connectivity features after dimensionality reduction using principal component analysis (PCA): the left panel is colored according to the ground-truth diagnostic labels, whereas the right panel shows the labels predicted by SI-HGN and the misclassified samples.

A.4 Additional Analysis of Disorders of Consciousness

We further evaluated SI-HGN in a hospital-based cohort of 39 patients with disorders of consciousness (DoC). Individual MRI-derived features were integrated to estimate patient-specific EC profiles, which were compared with normative EC patterns derived from healthy individuals. The deviation from healthy connectivity organization was used to characterize network disruption. Model-inferred abnormal connectivity patterns were also examined in relation to CRS-R scores and potential injury-related brain regions.

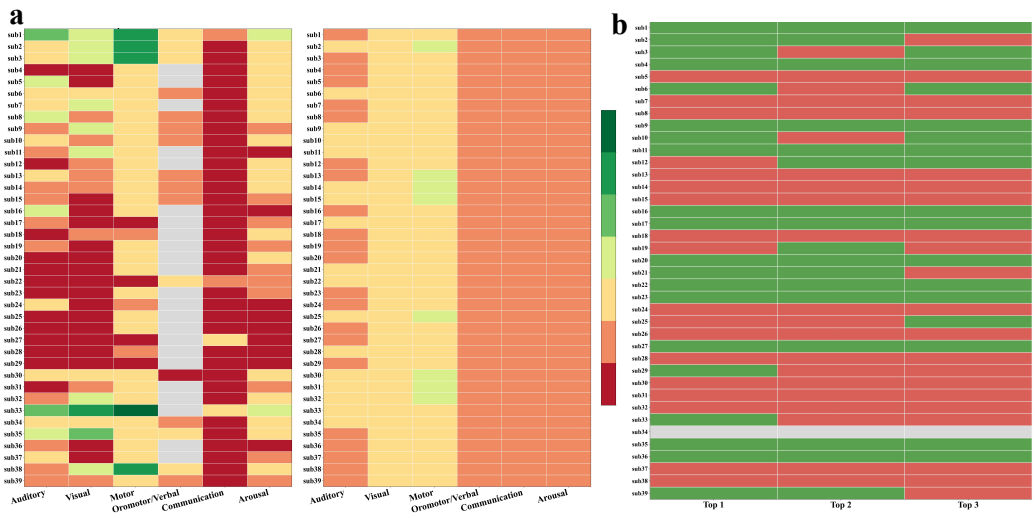

Figure 16: (a) Comparison between clinically assigned CRS-R scores and EC-derived scores for patients with disorders of consciousness (DoC). The left panel shows the CRS-R scores assigned by hospital clinicians, whereas the right panel shows the scores estimated by our framework based on the comparison between the predicted patient-specific EC and the healthy EC template. Gray areas indicate missing values. Scores range from 0 to 6 from bottom to top. (b) Evaluation of lesion-region prediction results. We quantified the degree of EC impairment across brain regions and selected the top three regions with the most severe EC abnormalities for assessment. Green indicates correct predictions, red indicates incorrect predictions, and gray indicates cases that could not be determined.

A.4.1 Interpretation of the DoC Connectivity Alterations

Figure 3 provides a cohort-level comparison between the normative EC organization and the EC estimated for the DoC cohort. Let $\mathbf{E}_{\mathrm{HC}}$ denote the normative EC template and $\mathbf{E}_{\mathrm{DoC}}^{(s)}$ denote the estimated EC of patient s. The cohort-level DoC matrix and its deviation from the healthy template are

$$\overline{\mathbf{E}}_{\mathrm{DoC}} = \frac{1}{N_{\mathrm{DoC}}} \sum_{s=1}^{N_{\mathrm{DoC}}} \mathbf{E}_{\mathrm{DoC}}^{(s)}, \qquad \Delta\mathbf{E} = \overline{\mathbf{E}}_{\mathrm{DoC}} - \mathbf{E}_{\mathrm{HC}}. \tag{28}$$

In Figure 3(a), negative entries of $\Delta\mathbf{E}$ indicate connections whose model-estimated strength is lower in the DoC cohort than in the normative template. The spatially distributed negative deviations suggest that the group difference is not confined to a single directed pathway, but is expressed across multiple parts of the estimated network. Because these are cohort-level averages, the panel summarizes a group pattern and should not be used to infer that every patient has the same altered connections.

Figure 3(b) provides a complementary distribution-level view. Relative to the healthy template, the DoC estimates occupy a narrower range and contain fewer high-strength EC values. This compression is consistent with reduced differentiation among strongly and weakly coupled directed path-

ways. The matrix and distribution panels therefore describe two aspects of the same result: where EC deviations occur and how the overall range of estimated EC values changes.

The cohort-level result in Figure 3 and the patient-level analyses in Figure 16 are complementary. Figure 3 describes the average network alteration relative to the normative template, whereas Figure 16 examines how individualized deviations relate to CRS-R scores and clinically identified injury-related regions. These results should be interpreted as properties of model-estimated perturbational responses. They do not by themselves establish direct anatomical causality, a clinically validated diagnostic biomarker, or the effect of a real neurostimulation intervention.