Unified Generative Latent Representation for Functional Brain Graphs

Subati Abulikemu ss2905@cam.ac.uk
Tiago Azevedo tiago.azevedo@cst.cam.ac.uk
Michail Mamalakis mm2703@cam.ac.uk
John Suckling js369@cam.ac.uk
University of Cambridge, Cambridge, United Kingdom

Editors: List of editors' names

Abstract

Functional brain graphs are often characterized with separate graph-theoretic or spectral descriptors, overlooking how these properties covary and partially overlap across brains and conditions. We anticipate that dense, weighted functional connectivity graphs occupy a low-dimensional latent geometry along which both topological and spectral structures display graded variations. Here, we estimated this unified graph representation and enabled generation of dense functional brain graphs through a graph transformer autoencoder with latent diffusion, with spectral geometry providing an inductive bias to guide learning. This geometry-aware latent representation, although unsupervised, meaningfully separated working-memory states and decoded visual stimuli, with performance further enhanced by incorporating neural dynamics. From the diffusion modeled distribution, we were able to sample biologically plausible and structurally grounded synthetic dense graphs.

Keywords: Graph Generative Model, Latent Diffusion, Functional Connectivity

1. Introduction

Large-scale functional brain systems are modeled as weighted graphs via statistical dependencies between distributed neural signals, i.e., functional connectivity (FC) (Biswal et al., 1995; Hallquist and Hillary, 2018). Classical graph-theoretic methods have effectively revealed the small-world, modular and hub-dominance architecture of FC graphs (Achard et al., 2006; Meunier et al., 2010), while graph spectral decomposition has complementarily identified low-frequency, computationally meaningful eigenmodes anchoring cortical organization (Margulies et al., 2016). Rather than summarizing functional brain graphs with separate topological and spectral descriptors, we propose learning a unified connectome embedding within a geometrically organized and generative latent space, such that movements in this space mirror structured transformations in graph topology and spectral gradients.

Current graph learning in connectomes is largely discriminative, with embedding strategies often adopting fixed density sparsification that equalises edge counts across graphs and may obscure global weight geometry (Li et al., 2021; Thapaliya et al., 2025; Wang et al., 2025). Instead, we learn a compact graph-level embedding from dense, weighted connectivity through joint node–edge encoding, using spectral geometry as inductive bias. A unified whole-brain representation provides geometric underpinnings of cognitive states and coordinates along which complex brain conditions vary smoothly.

For generation, we model the distribution of the learned embeddings with latent diffusion rather than operating on raw graphs (Rombach et al., 2021; Zhou et al., 2024). This mitigates challenges of non-Gaussian edge distributions and edgewise generation burdens, while allowing traversal within the interpretable latent geometry before decoding to dense FC graphs. Sampling structurally grounded brain graphs from this latent geometry then enables data augmentation and mechanistic explorations of connectivity-to-computation with connectome-constrained neural networks (Lappalainen et al., 2024; Suárez et al., 2024).

2. Methods

2.1. Problem set-up

We define each individual brain as a dense graph $\mathcal{G} = (W, X)$ with weighted adjacency $W \in \mathbb{R}^{N\times N}$ and node features $X \in \mathbb{R}^{N\times F}$. The objective was an unsupervised learning, through a transformer-based autoencoder, of a graph-level latent representation $z_g \in \mathbb{R}^{d_g}$ that: (1) is geometrically organized with graded variations along spectral and graph-theoretic axes with edge-aware encoding and spectral geometry inducting the representation; (2) supports downstream decoding of cognitive states despite being unsupervised; and (3) enables a generative prior through diffusion on z_g. We used Human Connectome Project resting-state and working-memory (WM) fMRI data (Essen et al., 2013).

2.2. Edge-conditioned self-attention encoder

We mapped $(W, X) \mapsto z_g$ with L edge-conditioned transformer layers similar to (Ma et al., 2023). Node and edge initializations followed $h^{(0)} = XW_{\text{init}}$, $e_{ij}^{(0)} = W_{ij}E_{\text{init}}$ (i, j index nodes). At layer l, we constructed queries, keys, and values as $Q_i{=}h_iW_Q$, $K_i{=}h_iW_K$, $V_i{=}h_iW_V$ and computed edge-conditioned logits

$$\hat{e}_{ij} = \text{GELU}\Big(\rho\big((Q_i{+}K_j) \odot (e_{ij}E_w)\big) + (e_{ij}E_b)\Big), \quad \alpha_{ij} = \text{softmax}_j\Big(\tfrac{\hat{e}_{ij}\cdot w_A}{\sqrt{d_k}}\Big), \tag{1}$$

with $\rho(x) = \text{sign}(x)\sqrt{|x| + \epsilon}$ for stabilization. Values injected edge information with $m_i = \sum_j \alpha_{ij}\big(V_j + (\hat{e}_{ij}E_v)\big)$. Multi-head outputs were updated through residual connections and feedforward networks for node and edge representations. After L_e layers, we obtained $h^{L_e} \in \mathbb{R}^{N\times d_h}$ and pooled with global attention

$$\alpha = \text{softmax}\big(\tanh(h^{L_e}W_{\text{pool}}) \cdot c_{\text{pool}}\big), \qquad s = \sum_{i=1}^{N} \alpha_i h_i^{L_e}, \qquad z_g = W_\mu s + b_\mu. \tag{2}$$

2.3. Cross-attention graph decoder

We reconstructed $\hat{\mathcal{G}} = (\hat{W}, \hat{X})$ from the highly compressed z_g using a proposed *memory mechanism*. We define a learnable memory $M \in \mathbb{R}^{N\times d_m}$, where M_i is a persistent embedding for node i. The memory M acts as a learned prior over regional characteristics and typical inter-regional affinities. Cross-attention uses z_g as a routing signal that selectively retrieves and combines these priors to instantiate a graph-specific realization. Here, keys and values

were constructed as $K = MW_K$ and $V = MW_V$, with node state initialized as $h^{(0)}{=}MW_{\text{init}}$. At decoder layer l,

$$Q_i = W_{q,z} z_g + W_{q,h} h_i, \quad \alpha_{ij} = \text{softmax}_j\Big(\frac{Q_i \cdot K_j}{\sqrt{d_k}}\Big), \quad m_i = \sum_j \alpha_{ij} V_j, \tag{3}$$

followed by residual and feedforward updates. After L_d layers we derived node representations $r_i = \phi_r(h_i^{L_d})$ and reconstructed

$$\hat{X}_i = \phi_X([r_i; z_g]), \qquad \hat{W}_{ij} = \phi_E([r_i; r_j; r_i \odot r_j; |r_i - r_j|; z_g]). \tag{4}$$

2.4. Latent diffusion on z_g

After training the autoencoder deterministically, we trained a denoising diffusion probabilistic model (Ho et al., 2020) on z_g with a linear noise schedule $\{\beta_t\}_{t=1}^T$: $q(z_t|z_{t-1}) = \mathcal{N}(\sqrt{1-\beta_t}\, z_{t-1}, \beta_t I)$, with the denoising network ϵ_θ trained on $\mathbb{E}_{t,z_0,\epsilon}\|\epsilon - \epsilon_\theta(z_t, t)\|^2$. Sampling ran the reverse process to obtain z_0; the decoder was *frozen* to map $z_0 \mapsto (\hat{W}, \hat{X})$.

3. Results

3.1. Resting-state graph learning and latent space

We compared dense FC graph reconstruction using three node feature sets: spectral embeddings via diffusion maps (first ten functional gradients) (Margulies et al., 2016); graph features (degree, clustering, participation coefficients); and edge-only encoding. We included a baseline graph convolutional autoencoder (GAE) with spectral embeddings. Spectral geometry improved learning (Figure 1A; edge MSE=0.012, R^2=0.80) compared to graph features (MSE=0.017, R^2=0.74), edge-only (MSE=0.016, R^2=0.74), and the baseline GAE (MSE=0.018, R^2=0.72). The latent z_g captured graded variations in spectral domains—association-sensory (ψ_{AS}) and visual-sensorimotor (ψ_{VS}) patterns quantified by gradient ranges, and graph-theoretic properties including mean degree, modularity, and small-worldness (Figure 1B).

3.2. Task-state graph learning and cognitive state separability

For heterogeneous task FC graphs, spectral embeddings (d=10/30) achieved superior reconstruction (MSE=0.032/0.029, R^2=0.67/0.71) versus graph features and edge-only encoding (MSE=0.044, R^2=0.55/0.56), and GAE (d=30; MSE=0.049, R^2=0.51; Figure 1C). The unsupervised latent space separated cognitive states; SVM on z_g classified 0- vs 2-back WM loads with 80% accuracy (AUC=0.87), and visual stimuli (4 types) presented during WM tasks with 60.9% (0-back) and 66.6% (2-back) accuracies. Further embedding neural activities improved classification to 86% (AUC=0.93) for cognitive load and 73.9%/70.6% for visual stimuli (Figure 1D).

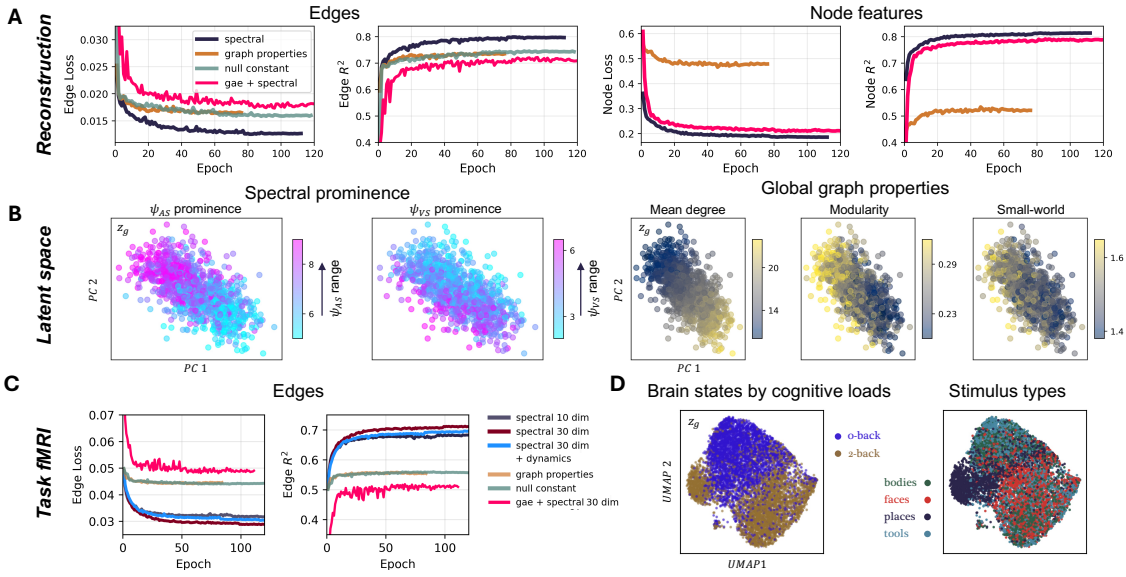

Figure 1: (**A**) Resting-state FC graph reconstruction comparing spectral, graph features, edge-only encodings and baseline. (**B**) Latent space color coded by spectral prominence (of association-sensory and visual-sensorimotor gradients) and global graph properties. (**C**) Task-fMRI FC reconstruction across different feature encodings. (**D**) UMAP showing cognitive load and stimulus type separation in z_g.

3.3. Latent diffusion and brain graph generation

We modeled $p(z_g)$ through diffusion on the pre-trained encoder's latent space from resting-state FCs (Figure 2A), then sampled and generated new dense FC graphs (Figure 2B). Generated graphs exhibited structured variations along the latent distribution; for instance, samples from high association-sensory (AS) dominance region displayed an extended AS axis in generated spectral gradients and greater AS system segregation in generated connectivity (Figure 2B). Distributions of graph statistics including mean degree across nodes (KS statistic=0.037, p=0.99), degree standard deviation (KS=0.079, p=0.51), and modularity (KS=0.098, p=0.25) aligned between test and generated sets (Figure 2C).

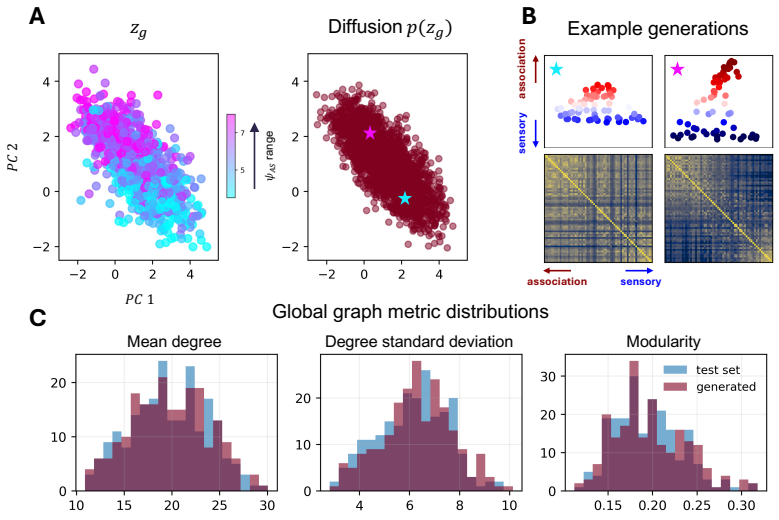

Figure 2: (**A**) Latent space of resting-state FC colored by association-sensory gradient (ψ_{AS}) prominence and diffusion-learned distribution $p(z_g)$. (**B**) Example generation of low- and high-ψ_{AS} prominence spectral embeddings and connectivity matrices. (**C**) Distribution alignment of graph statistics (mean degree, degree variability, modularity) between test and generated sets.

4. Conclusions

This work presents a functional brain-graph learning framework that maps smooth variations in spectral and topological structure through a compact latent geometry. Spectral embeddings as node features strengthened graph learning, especially for heterogeneous task-based graphs. In addition to being geometrically aware, this unsupervised latent space demonstrated functional relevance by meaningfully separating working-memory loads and stimulus types from task graphs. Diffusion sampling from this space generated graphs whose statistics aligning with real graphs. Future work will systematically map extended cognitive states and disease conditions to establish the geometry's functional significance, developing dual representation connecting graph structure with behavioral and clinical domains.

References

Sophie Achard, Raymond Salvador, Brandon Whitcher, John Suckling, and Ed Bullmore. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *Journal of Neuroscience*, 26:63–72, 1 2006. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.3874-05.2006. URL https://www.jneurosci.org/content/26/1/63https://www.jneurosci.org/content/26/1/63.abstract.

Bharat Biswal, F. Zerrin Yetkin, Victor M. Haughton, and James S. Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic Resonance in Medicine*, 34:537–541, 1995. ISSN 15222594. doi: 10.1002/MRM.1910340409,. URL https://pubmed.ncbi.nlm.nih.gov/8524021/.

David C. Van Essen, Stephen M. Smith, Deanna M. Barch, Timothy E.J. Behrens, Essa Yacoub, and Kamil Ugurbil. The wu-minn human connectome project: An overview. *NeuroImage*, 80:62, 10 2013. ISSN 10538119. doi: 10.1016/J.NEUROIMAGE.2013.05.041. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC3724347/.

Michael N. Hallquist and Frank G. Hillary. Graph theory approaches to functional network organization in brain disorders: A critique for a brave new small-world. *Network Neuroscience*, 3:1, 1 2018. ISSN 24721751. doi: 10.1162/NETN_A_00054. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC6326733/.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 2020-December, 6 2020. ISSN 10495258. URL https://arxiv.org/pdf/2006.11239.

Janne K. Lappalainen, Fabian D. Tschopp, Sridhama Prakhya, Mason McGill, Aljoscha Nern, Kazunori Shinomiya, Shin Ya Takemura, Eyal Gruntman, Jakob H. Macke, and Srinivas C. Turaga. Connectome-constrained networks predict neural activity across the fly visual system. *Nature*, 634:1132–1140, 10 2024. ISSN 14764687. doi: 10.1038/S41586-024-07939-3;TECHMETA=119;SUBJMETA=116,1483,1925,2613,378,3920,631;KWRD=MOTION+DETECTION,NETWORK+MODELS,NEURAL+CIRCUITS. URL https://www.nature.com/articles/s41586-024-07939-3.

Xiaoxiao Li, Yuan Zhou, Nicha Dvornek, Muhan Zhang, Siyuan Gao, Juntang Zhuang, Dustin Scheinost, Lawrence H. Staib, Pamela Ventola, and James S. Duncan. Braingnn: Interpretable brain graph neural network for fmri analysis. *Medical Image Analysis*, 74:102233, 12 2021. ISSN 1361-8415. doi: 10.1016/J.MEDIA.2021.102233. URL https://www.sciencedirect.com/science/article/pii/S1361841521002784?via%3Dihub.

Liheng Ma, Chen Lin, Derek Lim, Adriana Romero-Soriano, Puneet K. Dokania, Mark Coates, Philip H.S. Torr, and Ser Nam Lim. Graph inductive biases in transformers without message passing. *Proceedings of Machine Learning Research*, 202:23321–23337, 5 2023. ISSN 26403498. URL https://arxiv.org/pdf/2305.17589.

Daniel S. Margulies, Satrajit S. Ghosh, Alexandros Goulas, Marcel Falkiewicz, Julia M. Huntenburg, Georg Langs, Gleb Bezgin, Simon B. Eickhoff, F. Xavier Castellanos,

Michael Petrides, Elizabeth Jefferies, and Jonathan Smallwood. Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences of the United States of America*, 113:12574–12579, 11 2016. ISSN 10916490. doi: 10.1073/PNAS.1608282113/SUPPL_FILE/PNAS.201608282SI.PDF. URL /doi/pdf/10.1073/pnas.1608282113?download=true.

David Meunier, Renaud Lambiotte, and Edward T. Bullmore. Modular and hierarchically modular organization of brain networks. *Frontiers in Neuroscience*, 4, 2010. ISSN 16624548. doi: 10.3389/FNINS.2010.00200,. URL https://pubmed.ncbi.nlm.nih.gov/21151783/.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Bjorn Ommer. High-resolution image synthesis with latent diffusion models. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2022-June: 10674–10685, 12 2021. ISSN 10636919. doi: 10.1109/CVPR52688.2022.01042. URL https://arxiv.org/pdf/2112.10752.

Laura E. Suárez, Agoston Mihalik, Filip Milisav, Kenji Marshall, Mingze Li, Petra E. Vértes, Guillaume Lajoie, and Bratislav Misic. Connectome-based reservoir computing with the conn2res toolbox. *Nature Communications 2024 15:1*, 15:1–14, 1 2024. ISSN 2041-1723. doi: 10.1038/s41467-024-44900-4. URL https://www.nature.com/articles/s41467-024-44900-4.

Bishal Thapaliya, Esra Akbas, Jiayu Chen, Ram Sapkota, Bhaskar Ray, Pranav Suresh, Vince D. Calhoun, and Jingyu Liu. Brain networks and intelligence: A graph neural network based approach to resting state fmri data. *Medical Image Analysis*, 101:103433, 4 2025. ISSN 1361-8415. doi: 10.1016/J.MEDIA.2024.103433. URL https://www.sciencedirect.com/science/article/pii/S136184152400358X.

Yulong Wang, Vince D. Calhoun, Godfrey D. Pearlson, Peter Kochunov, Theo G.M. van Erp, and Yuhui Du. A graph transformer-based foundation model for brain functional connectivity network. *Pattern Recognition*, 169:111988, 1 2025. ISSN 0031-3203. doi: 10.1016/J.PATCOG.2025.111988. URL https://www.sciencedirect.com/science/article/pii/S003132032500648X.

Cai Zhou, Xiyuan Wang, and Muhan Zhang. Unifying generation and prediction on graphs with latent graph diffusion. *Advances in Neural Information Processing Systems*, 37: 61963–61999, 12 2024. URL https://github.com/zhouc20/LatentGraphDiffusion.

Appendix A. Extended Methods

Subjects and connectivity graphs. Functional MRI data from 1067 subjects were available. Training and test splits (80/20) were performed on subject indices and fixed across all training and analyses (including the transformer-autoencoder, diffusion model, and cognitive-state classifiers), resulting in 853 training and 214 test subjects with no subject overlap. For the working memory (WM) task, each subject's time series was partitioned into eight segments corresponding to task-block conditions (2 cognitive loads * 4 stimulus types). Functional connectivity (FC) matrices were computed as the Pearson correlation of regional time series under a 64-region parcellation. Therefore, the resting-state dataset contains one FC matrix per subject, whereas the WM dataset contains eight matrices per subject, all adopting the same subject-wise train–test split.

Spectral embeddings. Pairwise affinity A for each FC matrix was computed using normalized angle kernel. To extract the eigenmodes, we first constructed a kernel $K = D^{-\alpha}AD^{-\alpha}$, where D is the degree matrix $D_{ii} = \sum_j A_{ij}$, and $\alpha = 0.5$, approximating Fokker–Planck diffusion. A Markov diffusion operator was then computed as $P = \left(\sum_j K_{ij}\right)^{-1} K$, whose eigenvectors were scaled with multiscale eigenvalue aggregates, $\sum_{t=1}^{\infty} \lambda_k^t = \lambda_k/(1-\lambda_k)$, producing the set $\{\psi_k\}$ that defines functional gradients (Figure 3).

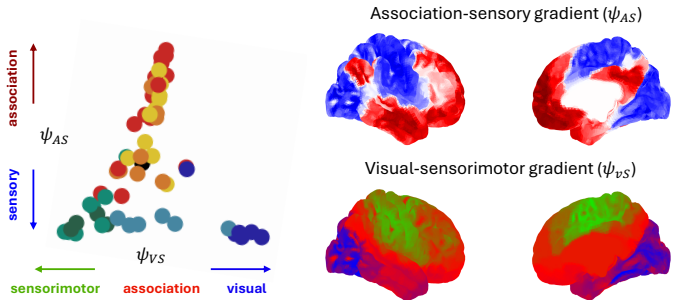

Figure 3: Diffusion-map embeddings (first two gradients). The ψ_{AS} continuously spans from transmodal association to unimodal sensory networks, while ψ_{VS} separates lower-order visual and somatosensory/motor systems.

The hierarchical organization captured by each gradient can be quantified through its diffusion range, $\text{range}(\psi_k) = \max(\psi_k) - \min(\psi_k)$. Greater range corresponds to more pronounced differentiation and steeper transition across network components. To establish correspondence between individual and group-level embeddings, we used orthogonal Procrustes alignment preserving individual geometric structures while ensuring comparable dimensions. The group template was computed strictly from the resting-state training set (mean FC) and adopted in all gradient computations.

Feature normalizations. For each training pipeline (rest and WM), all node features were normalized using statistics from the training set only and then applied unchanged to the test set. For spectral embeddings, all dimensions were scaled by a single global training

standard deviation to preserve the eigenvalue induced relative scaling and hierarchy across dimensions.

Graph transformer autoencoder loss function. For a batch $\{(X^{(b)}, W^{(b)})\}_{b=1}^{B}$ with reconstructions $(\hat{X}^{(b)}, \hat{W}^{(b)})$, we minimized the weighted mean-squared error on nodes and edges:

$$\mathcal{L} = \frac{1}{B}\sum_{b=1}^{B}\left[\lambda_x \frac{1}{nd}\sum_{i=1}^{n}\sum_{k=1}^{d}\left(X_{ik}^{(b)} - \hat{X}_{ik}^{(b)}\right)^2 + \lambda_a \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(W_{ij}^{(b)} - \hat{W}_{ij}^{(b)}\right)^2\right].$$

Appendix B. Extended Results

Function–geometry dual representation. When embedding FC graphs from the seven cognitive tasks, a linear probe on the learned latent representation z_g achieved 85% accuracy (Figure 4A,B). Geometrically, the latent space displayed a radial integration–segregation organization, where the latent distance to the centroid ($r = \|z_g - \bar{z}\|_2$) correlated with modularity ($\rho = 0.34$), small-worldness ($\rho = 0.44$), and mean clustering coefficient ($\rho = 0.53$). Peripheral positions in latent space therefore corresponded to more segregated network configurations. Functionally, this global axis further branched, where greater eigenmode prominence—i.e., more segregated organization along association–sensory, visual–sensorimotor, and attention/frontoparietal–default mode gradients—occupied distinct regions of the latent space according to the cognitive task performed (Figure 4C).

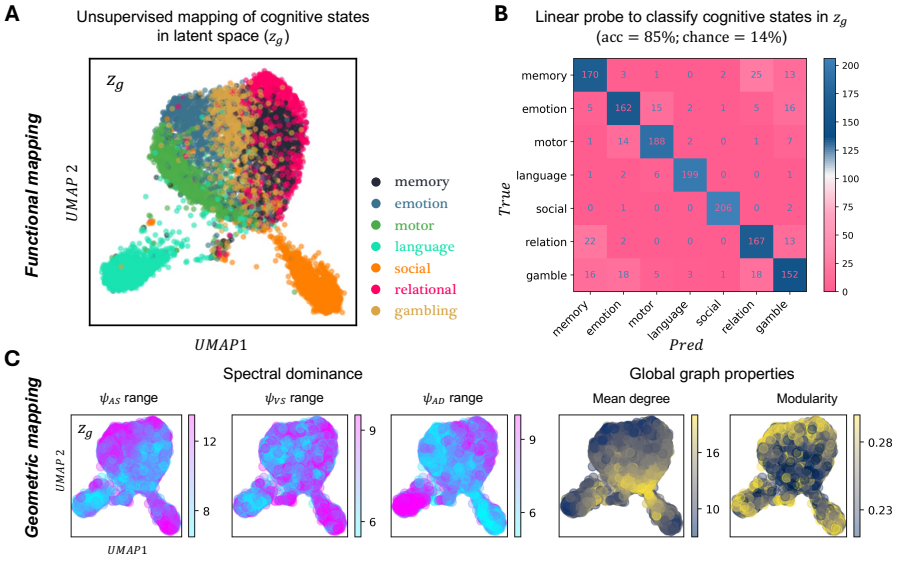

Figure 4: (**A**) Latent space embedding z_g of FC graphs from seven cognitive tasks. (**B**) Confusion matrix for linear classifier on test set z_g. (**C**) Radial integration–segregation origination in z_g.

Latent diffusion quality. For assessing the quality of latent diffusion prior, we compared the original training distribution against the diffusion generated set. The median KS statistic across latent dimensions was 0.040, with all dimensions having $p > 0.05$. The covariance structure across latent dimensions was also accurately preserved, with a correlation of 0.99 between training and generated sets (Figure 5B).

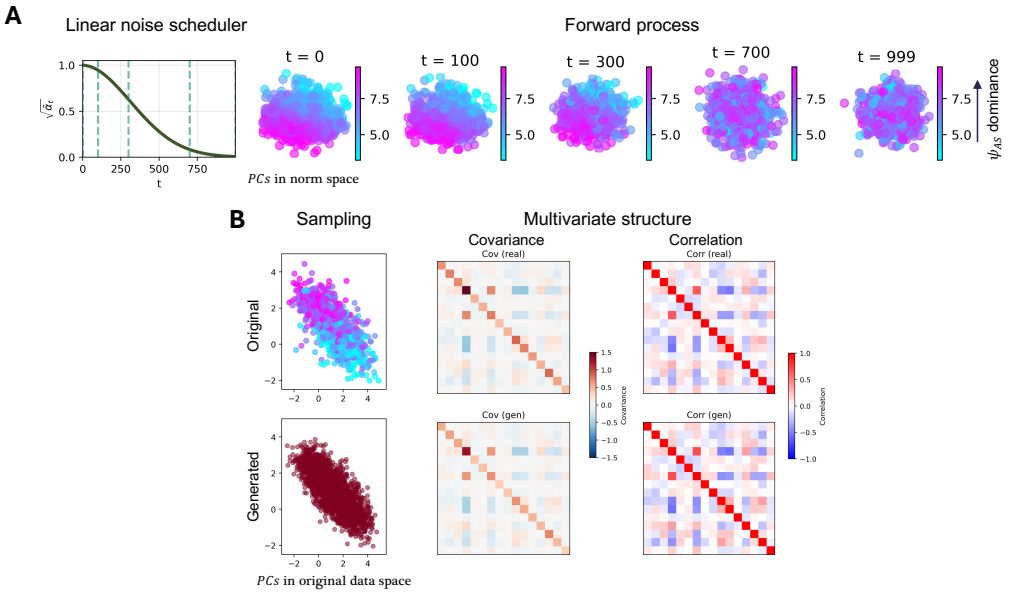

Figure 5: (A) Linear noise scheduler showing the signal retention $\sqrt{\bar{\alpha}_t}$ across diffusion steps t, with the corresponding forward noising process in latent space. (**B**) The covariance structure of latent dimensions in the original training and generated sets.