MAGNET: Manifold-Aware Graph Diffusion Network for Connectome Generation

Protyay Dey[1*†] **Ayush Roy**[1*] **Hyuna Cho**[2] **Won Hwa Kim**[3] **Vishnu Suresh Lokhande**[1]

[1]University at Buffalo, SUNY, USA [2]National Institutes of Health, USA [3]POSTECH, South Korea

Abstract

Functional brain connectome represents neural connectivity as a matrix of pairwise interactions between brain regions. Generation of functional connectomes is not only a question of validity; rather, having satisfied the constraints on the correlation matrix, the next step is to recover the class-conditional geometry buried under coarse labels. We propose MAGNET, a Manifold-Aware Graph Diffusion Network which uses a normalized-Cholesky representation of the manifold of correlation matrices that guarantees validity. MAGNET lifts noisy latent states into ROI-level region tokens and performs denoising with a relational inductive bias over brain atlas regions. To deal with structural problems induced by coarse labels, MAGNET employs class-anchored conditioning, amortized structural bridge, and relevance-preserving corruption. Across ABIDE, ADNI, and OASIS-3, MAGNET consistently achieves favorable results compared to previous manifold-aware approaches, demonstrating improvements of $7-21\%$ in class-conditional fidelity (α, β-F1) across three cohorts and better sampling efficiency. Moreover, while training only with strict binary labels, MAGNET is capable of maintaining clinical heterogeneity through fine substructure of the connectomes in a zero-shot setting, improving subclass covariance alignment (λ-MSE) by over 30%. These results suggest that geometric validity is a necessary but insufficient condition for clinical utility in connectome synthesis. Moreover, efforts in making the diffusion denoising class-conditional manifold aware finds utility beyond the highly curved brain connectome generation as this is a critical problem in various general settings. The code repository can be found at `https://github.com/protyayofficial/MAGNET`.

1 Introduction

A brain connectome is a comprehensive map of neural connectivity [76], formalized as a relational object where each entry encodes interactions between pairs of distinct brain regions. Resting-state functional connectomes play a central role in biomarker discovery, cohort analysis, and personalized prognosis in modern neuroimaging [74, 15]. Conditional generative model for connectomes enables key applications in data-scarce settings, including augmentation [25, 1], synthesis of underrepresented populations [59], and construction of synthetic cohorts [79]. Related conditional diffusion approaches have also been explored for longitudinal neurodegenerative data synthesis [18]. Such a model must satisfy two criterias: (i) samples must be valid correlation matrices (symmetric, positive definite, with unit diagonals) [64, 3]; and (ii) must lie within the valid class-conditional regions of the connectome manifold, ensuring both geometric validity and clinical consistency [8, 67, 21]. While manifold-aware generative models address validity, accurately recovering class-conditional support remains a key challenge.

Riemannian score models approximate distributions with manifold support [26], while Riemannian flow matching and pullback-based pipelines enable efficient transport without leaving the mani-

*Equal contribution.

†Corresponding authors: `protyayd@buffalo.edu` and `vishnulo@buffalo.edu`.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

fold [41, 21]. Recent work on manifold-preserving diffusion, guidance, and Lie group methods further emphasizes that geometry is intrinsic to the learning problem [36, 19, 9, 35]. For connectomes, the normalized-Cholesky pullback [3, 21] guarantees validity by construction, ensuring every decoded sample is a proper correlation matrix. As illustrated in Fig. 1, valid structures are clinically meaningful, and the primary manifold direction smoothly *stratifies* diagnostic groups across cohorts.

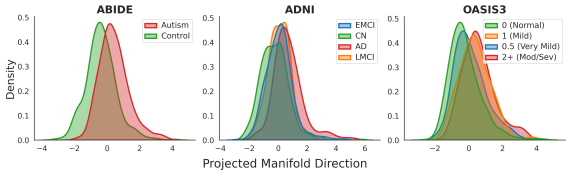

Figure 1: **1D manifold projection.** Density estimates along the primary manifold direction across ABIDE, ADNI, and OASIS-3. A consistent diagnostic shift along the dominant geometric axis confirms that valid manifold structure carries clinical signal. Healthy controls are anchored in green.

In neuroimaging, binary labels often collapse structurally diverse subpopulations into a single conditioning signal [79, 85, 84]. Prior brain network classification work has highlighted the importance of modeling progressive diagnostic relationships and heterogeneous graph structure [17]. This challenge is further exacerbated by inter-cohort distribution shifts and covariate imbalances, which can obscure clinically meaningful heterogeneity in neuroimaging datasets [20]. Fig. 2 shows *class-conditional distributions are not centered around a single prototype but instead span directions orthogonal to the main diagnostic axis, reflecting substantial variability that binary labels cannot capture.* Conditioning a generator on such coarse labels (e.g., labeling a scan simply as "Alzheimer's" without accounting for different subtypes or disease stages) forces the model to blur these distinct subgroups together into an unrealistic average [34]. Moreover, even when the generator produces samples in the correct geometric region, it may still fail to capture the underlying structure [8, 67]. Thus, the challenge of support recovery, i.e., *correctly assigning probability mass to the appropriate class-conditional regions on the manifold of valid connectomes*, remains unresolved (cannot be assessed by geometric validity alone).

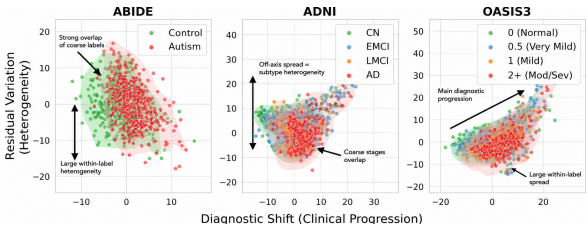

Figure 2: **Latent structural heterogeneity.** 2D projections of connectomes across ABIDE, ADNI, and OASIS-3. *Key observation:* The extensive vertical spread (residual variation) and severe overlap within individual clinical labels demonstrate that coarse categories collapse diverse subpopulations. This off-axis variation proves classes lack a single prototype, visually confirming the core challenge of conditional support recovery.

Core challenges in conditional support recovery. While a valid chart is necessary, conditional connectome synthesis faces three inherent challenges. The *label challenge* arises because coarse binary labels collapse structurally heterogeneous subpopulations, forcing feature blending [79, 85, 84, 42, 13, 45, 14]. The *global-context challenge* occurs because local noisy states lack sufficient information to recover global structure under limited reverse steps, while standard classifier-free guidance can overspecialize and reduce coverage [52, 47, 19, 39, 48]. The *coordinate challenge* stems from isotropic corruption in a nonlinear chart, where equal-variance noise distorts informative coordinates and practically degrades class-conditional structure, making recovery increasingly difficult under constrained few-step sampling regimes [41, 9, 62, 77, 86, 75].

All three challenges are compounded by a structural mismatch inherent to the chart: connectomes are relational objects [7], but the normalized-Cholesky parameterization imposes a strict ordering on brain regions, making latent coordinates explicitly not permutation-equivariant [88, 53]. Graph-based inductive biases are therefore motivated not by permutation invariance [12] but by the structured dependencies among ordered Cholesky coefficients. Each z_{ij} encodes the orthogonalized partial correlation of region i with respect to the preceding $j-1$ regions in the atlas ordering. Moreover, ROI-level attention with learned anatomical positional embeddings [35, 33] consistently outperforms flat denoisers that ignore this relational structure [44, 10, 24, 11]. Together, these shift the focus beyond geometric feasibility: *the core difficulty is conditional support recovery*.

Keeping the aforementioned points in mind, we introduce **MAGNET**, a **M**anifold-**A**ware **G**raph **Net**work diffusion framework operating in a normalized-Cholesky chart [3, 21], where denoising is performed over region tokens rather than flattened coordinates. The central insight is that geometric validity is a necessary but insufficient condition and the binding constraint is conditional support

recovery. It requires jointly resolving label ambiguity, missing global context, and coordinate corruption within the valid manifold. MAGNET addresses this through three coupled mechanisms:

- We address the *label challenge* via **class-anchored conditioning**: each class is modeled as a Gaussian mixture [54, 32, 27], assigning subtype tokens to explicitly factor latent structural heterogeneity before conditioning reaches the denoiser [79, 85, 29, 89].
- For the *global-context challenge*, we introduce an **amortized structural bridge**, a lightweight predictor that provides subtype-specific structural summaries derived from correlation statistics at each reverse step. This compensates for the inability of local noisy states to encode global connectivity under few-step regimes, while avoiding the diversity collapse typically induced by classifier-free guidance in limited-step inference [39, 19, 48, 70].
- Finally, the *coordinate challenge* is handled via **relevance-preserving corruption**, a condition-adaptive forward process that uses the conditional–unconditional denoising gap to identify informative chart coordinates and reduce their corruption variance early. Unlike auxiliary-prior or conditioning-annealing approaches [77, 70], we directly modulate corruption covariance, ensuring class-relevant structure survives into the reverse trajectory where it is most recoverable [41, 62].

Sec. 2 formalizes generation as a support-recovery problem, decoupling geometric validity from clinical fidelity. We establish theoretical guarantees and provide ablations (App. F) validating our class-anchored conditioning, amortized bridge, and relevance-preserving corruption. Empirically (see Sec. 3), MAGNET demonstrates robust conditional fidelity across ABIDE, ADNI, and OASIS-3, achieving superior zero-shot subclass recovery alongside accelerated sampling times.

2 Methodology

We construct MAGNET using four coupled components, targeting the three challenges identified in Sec. 1 - label ambiguity from coarse supervision, missing global context under few-step inference, and coordinate-space corruption in the nonlinear chart. Sec. 2.1–2.2 establish the validity-preserving chart and ROI-level graph denoiser; Sec. 2.3.1–2.3.3 introduce the three controllers that address label ambiguity, missing global context, and coordinate-space corruption respectively. Sec. 3.2 evaluates zero-shot subclass preservation. All proofs, ablations, and geometric analyses are in App. B–H.

2.1 From validity to support recovery

Let $\mathcal{C}_d = \left\{C \in \mathbb{R}^{d\times d} : C = C^\top,\ C \succ 0,\ \mathrm{diag}(C) = \mathbf{1}\right\}$ denote the manifold of valid correlation connectomes [64]. We study conditional generation $C \sim p_\theta(C \mid y)$, where y is a coarse label. Once validity is ensured, the central challenge shifts to *support recovery*: generating samples from the correct class-conditional regions of the connectome manifold (Figure 2). This is difficult under coarse labels, which collapses structurally heterogeneous subpopulations into a single label [79, 85, 84]. As this issue is orthogonal to geometric validity, validity-preserving transport alone is insufficient [21]. Consequently, models may produce geometrically valid but clinically implausible samples by averaging across distinct modes [34, 8], reducing diversity needed for downstream tasks [59, 85]. MAGNET addresses this via three components: (i) a validity-preserving latent chart, (ii) an ROI-level graph denoiser, and (iii) controllers targeting coarse-label ambiguity, missing global context, and early corruption of condition-relevant coordinates.

Geometry-preserving latent chart. Instead of diffusing directly on C, we operate in the normalized-Cholesky chart [21], which bijectively parameterizes the free degrees of freedom of correlation matrices and guarantees valid outputs upon inversion [3]. Let $C = LL^\top$ with L lower triangular and positive diagonal, and define $\widetilde{L} = \mathrm{diag}(L)^{-1}L$. The chart coordinate is $z = \chi(C) := \mathrm{vech}_0(\widetilde{L}) \in \mathbb{R}^m$, where $m = \frac{d(d-1)}{2}$. Diffusion is performed on standardized coordinates $x = (z - \mu_z)/\sigma_z$, with statistics computed from training data [3, 66, 21]. Decoding reverses this process: we reconstruct $\widehat{L}$, form $G = \widehat{L}\widehat{L}^\top$, and renormalize to obtain $\widehat{C} = \mathrm{diag}(G)^{-1/2}\, G\, \mathrm{diag}(G)^{-1/2}$. By construction, $\widehat{C}$ is symmetric, positive definite, and unit-diagonal [3, 21], so validity is enforced intrinsically rather than via post-hoc projection. We reserve χ for this chart, while structural feature extractors used in the bridge and spectral loss are denoted separately by $\varphi(\cdot)$.

2.2 ROI-level graph denoising in the validity-preserving chart

Validity-preserving charts guarantee geometrical feasibility but lack the relational inductive biases required to recover class-conditional interactions in flat latent spaces [7, 33, 35, 21, 9]. Because

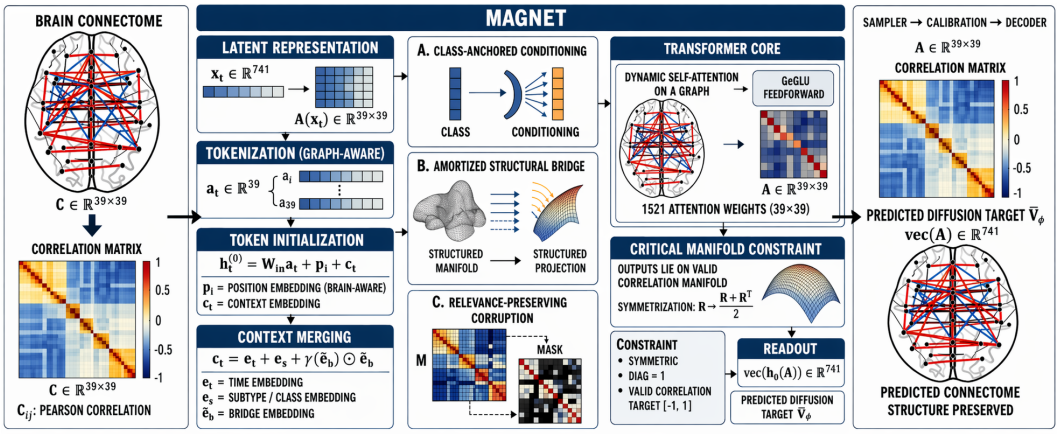

Figure 3: **Illustration of MAGNET Architecture.** The pipeline maps raw correlation matrices to validity-preserving coordinates (x_t), which are tokenized and merged with additive context embeddings (time, subtype, and bridge). The Transformer core then performs graph-aware denoising, modulated by three dedicated controllers: **Class-Anchored Conditioning, Amortized Structural Bridge, and Relevance-Preserving Corruption**. Finally, row projection and symmetrization guarantee the predicted structure remains on the valid correlation manifold.

functional connectomes inherently require representing nodes alongside relationship patterns [44, 10, 24], MAGNET denoises ROI tokens directly. This successfully combines geometric feasibility with necessary structural biases [7] (validated against flat charts in Appendix 3.3).

Given x_t, we reshape x_t into a symmetric, zero-diagonal matrix $A(x_t) \in \mathbb{R}^{d\times d}$. Each row of this matrix is then used as an initial region token [44, 81]. Concatenating orthogonal signals like timestep, class, and context often entangles their semantic roles. Instead, MAGNET uses additive token-level conditioning, ensuring each channel is independently optimizable [65, 63, 61, 40]. Defining p_i as the learned anatomical positional embedding, e_t as the timestep embedding [40], e_s as the subtype embedding, and $\widetilde{e}_b = \mathrm{Embed}_{\mathrm{spec}}(b_t)$ as the bridge embedding, where $b_t = B_\psi(x_t, t, s)$. Here, x_t is the noisy normalized cholesky state, t is the diffusion timestep, s is the class-anchored subtype, and B_ψ (.) is the amortized bridge predictor (see sec. 2.3.2). MAGNET forms a shared conditioning vector and initializes ROI tokens as:

$$c_t = e_t + e_s + \xi(\widetilde{e}_b) \odot \widetilde{e}_b, \qquad h_i^{(0)} = W_{\mathrm{in}} a_i(x_t) + p_i + c_t, \tag{1}$$

where $\xi(\cdot)$ is a learned gating map and W_{in} is a row projection. A pre-norm transformer stack [81, 87] with QK-normalized attention [37] and GeGLU feed-forward layers [73] produces the final token matrix $H^{(L)}$. MAGNET then applies a row-wise output projection followed by symmetrization:

$$R = W_{\mathrm{out}} H^{(L)}, \qquad E = \frac{1}{2}\left(R + R^\top\right), \qquad \widehat{v}_\theta = \mathrm{vech}_0(E). \tag{2}$$

Eq. (2) maps the region-level representation back to chart coordinates. Fig. 3 summarizes this full architecture. The chart enforces valid correlation matrix output geometry [21], while Transformer backbone performing ROI-level graph denoising supplies the relational inductive bias [7] needed for conditional support recovery. Detailed component implementations are provided in App. E.

2.3 MAGNET Framework

2.3.1 Class-anchored conditioning

First problem arises due to coarse labels, where the application of one label to multiple latent structural subpopulations [79, 85, 84] results in sample points located in different off-axis subregions rather than a single prototype (Fig. 2). Conditional on such labels, generator has no choice but to create a mean representation for each mode within the class [34], generating geometrically valid yet clinically unrealistic samples [8, 67, 59]. Existing works on subpopulation identification and disease heterogeneity modeling shows that an explicit decomposition of the coarse labels into latent substructures significantly alleviates mode-averaging [79, 85, 84]. Ignoring this sub-manifold structure harms generalization, analogous to label conflation in classification [42, 13], and biases the denoiser toward majority modes, compressing diversity toward dominant structural patterns [14, 45].

MAGNET refines each coarse class y with a *train-only subtype variable* s. For each class, we fit a diagonal-covariance Gaussian mixture in standardized chart coordinates via expectation maximization [27]: $p(x \mid y) \approx \sum_{s=1}^{K_y} \pi_{y,s}\, p(x \mid y, s)$. We select K_y by BIC [71, 32] and assign hard subtype

labels $s_i = \arg\max_s p(s \mid x_i, y_i)$ [54]. At generation, we sample $s \sim p(s \mid y)$ and condition the denoiser on the corresponding token. This variable s captures latent within-class structure rather than a validated biological taxonomy [85, 84].

Proposition 1 (Subtype refinement monotonically reduces Bayes denoising risk)**.** *Assume $\mathbb{E}\|X_0\|_2^2 < \infty$. Let X_0 be the clean latent variable, X_t its noisy version, Y the coarse class label, and S a class-anchored subtype variable. Under squared loss, we define the Bayes denoising risk*

$$\mathcal{R}_Y^\star = \inf_f \mathbb{E}\big[\|X_0 - f(X_t, Y)\|_2^2\big], \qquad \mathcal{R}_{Y,S}^\star = \inf_g \mathbb{E}\big[\|X_0 - g(X_t, Y, S)\|_2^2\big]. \tag{3}$$

Then

$$\mathcal{R}_Y^\star - \mathcal{R}_{Y,S}^\star = \mathbb{E}\Big[\|\mathbb{E}[X_0 \mid X_t, Y, S] - \mathbb{E}[X_0 \mid X_t, Y]\|_2^2\Big] \geq 0. \tag{4}$$

In particular, subtype refinement is beneficial exactly when it captures within-class structure unresolved by the coarse label.

Remark. Eq. 4 demonstrates that subtype refinement improves denoising exactly when S captures within-class structure unresolved by the coarse label Y (Fig. 2). Proofs are in App. B.1 and F.4.

2.3.2 Amortized structural bridge

Subtype conditioning is a solution for the problem of labeling, but it does not have the required global structural knowledge to use a few-step reverse sampling technique, hence creating a dilemma for the choice of either out-of-distribution averaging [8] or constrained manifold learning [70, 48, 19]. In order to resolve this ambiguity without additional forward samples, we condense a structural feature of the subtype from a pre-trained dictionary of data into an amortized predictor [38, 16]. Unlike self-conditioning, which recycles recursive denoiser outputs [16], this bridge provides an independently supervised, compact structural prior to offset the limitations of the noisy input x_t.

Let $\varphi(C) \in \mathbb{R}^{16}$ denote the proxy structural feature extractor, combining spectral properties (e.g., moments, rank, condition number) and off-diagonal connectivity statistics. For each subtype s, MAGNET fits a summary atom set $\mathcal{A}_s = \{a_{s,1}, \ldots, a_{s,K_s}\} \subset \mathbb{R}^{16}$ ($K_s \in \{1, 2\}$) using either the subtype mean or two principal-split means, depending on training support. Given a clean latent x_0, the teacher bridge target is a distance-weighted convex combination of these atoms: $b^\star(x_0, s, t) = \sum_{k=1}^{K_s} w_k(x_0, s, t)\, a_{s,k}$. Because x_0 is unavailable at inference, MAGNET amortizes this process by predicting the target directly from the noisy state: $b_t = B_\psi(x_t, t, s) = f_\psi(\varphi_{\text{cheap}}(x_t), e_t, e_s)$. Here, φ_{cheap} is a low-cost proxy of the noisy chart coordinates. Trained to match the teacher target, B_ψ is injected via Eq. 1. By entirely replacing the teacher during sampling, B_ψ bypasses repeated clean-state estimations, yielding a $24 - 33\times$ speedup (Table 5).

$$L_{\text{bridge}} = \|B_\psi(x_t, t, s) - b^\star(x_0, s, t)\|_2^2. \tag{5}$$

At inference time, B_ψ replaces the teacher entirely, yielding the approximately $25 - 32\times$ sampling speedup over teacher-bridge alternatives shown in Tab. 5.

Proposition 2 (Bridge-assisted denoising risk and amortization error)**.** *Let $U = (X_t, t, S)$, where X_t is the noisy latent state, t is diffusion time, and S is the subtype token. Let $B^\star = b^\star(X_0, S, t)$ denote the teacher bridge target. Under squared loss, define*

$$\mathcal{R}_U^\star = \inf_f \mathbb{E}\big[\|X_0 - f(U)\|_2^2\big], \qquad \mathcal{R}_{U,B^\star}^\star = \inf_g \mathbb{E}\big[\|X_0 - g(U, B^\star)\|_2^2\big]. \tag{6}$$

Then

$$\mathcal{R}_U^\star - \mathcal{R}_{U,B^\star}^\star = \mathbb{E}\Big[\|\mathbb{E}[X_0 \mid U, B^\star] - \mathbb{E}[X_0 \mid U]\|_2^2\Big] \geq 0. \tag{7}$$

Moreover, if the Bayes regressor $\eta(u, b) = \mathbb{E}[X_0 \mid U = u, B^\star = b]$ is L-Lipschitz in b, then the plug-in estimator $\widehat{X}_0(U) = \eta(U, B_\psi(U))$ satisfies

$$\mathbb{E}\Big[\|X_0 - \widehat{X}_0(U)\|_2^2\Big] - \mathcal{R}_{U,B^\star}^\star \leq L^2\, \mathbb{E}\big[\|B_\psi(U) - B^\star\|_2^2\big]. \tag{8}$$

Remark. Prop. 2 shows that conditioning on the teacher bridge $B^\star$ reduces denoising risk [56, 31] (Eq. 7), while minimizing L_{bridge} bounds the inference gap (Eq. 5). As shown later in Tab. 5 (App. F.1), lower distillation error directly improves α, β-F1, empririically validating the amortization bound (Eq. 8). Furthermore, omitting L_{bridge} spikes tracking error by two orders of magnitude with no fidelity gain, confirming this active bridge mode is essential rather than optional (App. F.1, B.2).

2.3.3 Relevance-preserving corruption

The third failure mode is geometric. Although the chart χ guarantees validity after decoding, isotropic corruption in chart coordinates is not geometry-neutral: because χ is nonlinear, equal-variance perturbations can move the noisy state away from the local class-conditional support that the reverse process must recover. This is especially harmful in the few-step regime where interpolation steps are larger. Motivated by geometry-aware diffusion and on-manifold correction methods [41, 9, 36, 50, 62, 68], we therefore replace isotropic corruption with a manifold-faithful, condition-adaptive noising rule as trajectory deviation at any timestep leads to accumulation of deviations away from the manifold [68, 62].

Starting from a provisional isotropic sample $\tilde{x}_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \epsilon$, $\epsilon \sim \mathcal{N}(0, I)$, we estimate condition-relevant directions from the conditional-unconditional denoising gap

$$d_t = G_\theta(\tilde{x}_t, t, s, b_t) - G_\theta(\tilde{x}_t, t, \varnothing, \varnothing), \quad m_{t,j} = \frac{|d_{t,j}|}{\max_k |d_{t,k}| + \varepsilon}, \tag{9}$$

where $m_t \in [0,1]^m$ scores chart coordinates by conditional relevance. Following the auxiliary-prior view of diffusion guidance [77], we define a task-specific manifold-faithfulness prior

$$\pi_t(x_t \mid x_0, \tilde{x}_t, s, b_t) \propto \exp\Big(-\lambda_t \left\| \sqrt{m_t} \odot \left(x_t - \sqrt{\bar{\alpha}_t} x_0\right) \right\|_2^2\Big), \qquad \lambda_t = \frac{\kappa \rho(t)}{2(1-\bar{\alpha}_t)}, \tag{10}$$

where $\rho(t) \in [0,1]$ is a monotonically decreasing protection schedule. Equation (10) penalizes corruption away from the support-aligned coordinates indicated by m_t: directions judged condition-relevant are protected most strongly early in diffusion. [3]

Multiplying the isotropic Gaussian kernel by this prior yields a diagonal Gaussian corruption law, $q_{\mathrm{mf}}(x_t \mid x_0, \tilde{x}_t, s, b_t) = \mathcal{N}\big(\sqrt{\bar{\alpha}_t} x_0,\, (1-\bar{\alpha}_t)\,\mathrm{Diag}(\sigma_t^2)\big)$, with coordinate-wise variance $\sigma_{t,j}^2 = \mathrm{clip}\big(1 - \kappa\rho(t) m_{t,j},\ \sigma_{\min}^2,\ 1\big)$. The implemented forward corruption is therefore $x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, (\sigma_t \odot \epsilon),\ \epsilon \sim \mathcal{N}(0, I)$.

This rule is annealed in the spirit of [70]: protection is strongest at early timesteps, where coarse corruption can erase condition-defining structure, and vanishes later so that the process reverts to standard isotropic noising near the clean end. Unlike [70], however, we anneal the corruption covariance rather than the conditioning signal itself. Thus MAGNET reduces off-support drift at its source: coordinates that matter most for conditional denoising retain higher signal-to-noise ratio early in the trajectory, while asymptotically recovering the standard diffusion kernel as $\rho(t) \to 0$. *Intuitively, preserving condition-relevant structure for longer keeps the noisy state closer to the local class-conditional manifold at each timestep, making it easier for the reverse process to remain on, and return to, the correct support.* A Gaussian surrogate analysis formalizing this is in App. B.3 and corresponding ablations are reported in App. F.2.

2.4 Full Training Objective

Our training objective is $L = L_{\mathrm{diff}} + \lambda_{\mathrm{bridge}} L_{\mathrm{bridge}} + \lambda_{\mathrm{spec}} L_{\mathrm{spec}}$. Here, L_{diff} corresponds to the diffusion loss, L_{bridge} is the bridge distillation error (Eq. 5), and $L_{\mathrm{spec}} = \|\varphi(\widehat{x}_0) - \varphi(x_0)\|_2^2$ steers the predicted latent to retain structural features. L_{diff} uses min-SNR reweighting and inverse subject-frequency weights for longitudinal cohorts. L_{bridge} is additionally weighted by $\bar{\alpha}_t$ and masked to conditioned examples only and L_{spec} is mean-reduced over the minibatch.

We sample with DDIM [75], using six reverse steps as the primary setting. Generation begins by sampling a subtype $s \sim p(s \mid y)$ and initializing from Gaussian noise, followed by reverse denoising with subtype and bridge conditioning at every step. Instead of the standard linear CFG approach, we use a diversity-preserving variant that orthogonalizes the unconditional residual from the guidance direction (see App. E for details), using $\gamma = 1.5$ (App. F.3). This directly addresses the global-context failure discussed in Sec. 2.3.2: plain linear CFG [39] adds a fixed displacement along the guidance direction with a scale that can cause over-specialization toward narrow manifold regions [19, 48]. Our formulation decomposes the unconditional prediction into components parallel and orthogonal to g, and *replaces* the parallel variant with $\tanh$-saturated guidance instead of adding it. In standard CFG, the full unconditional prediction $\hat{v}_\emptyset$ is retained and γg is added which can keep the parallel

[3] During training, G_θ in Equation 9 is evaluated with a stop-gradient, so the relevance mask m_t does not back propagate through the denoizer and diffusion marginals remain tractable.

Table 1: **Quantitative comparison of validity-preserving generative models under matched low-step budgets.** We compare MAGNET against GDD-proj[43], TriangCFM, DiffeoGauss, and DiffeoCFM[21] on ABIDE [30], ADNI [83], and OASIS-3 [49]. We report α-Precision, β-Recall, and α, β-F1 [2] to evaluate class-conditional support fidelity, ROC-AUC and F1 to evaluate downstream utility, and wall-clock sampling time per split. The Real Data row provides the empirical oracle from split-wise real-data comparisons, and the MAGNET row highlights our method. Bold denotes the best generative model for each metric and underlining denotes the runner-up. Supporting geometric analyses are given in Appendix G–H.

Dataset	Method	Quality Metrics			CAS Metrics		Time (s.)
		α-Precision ↑	β-Recall ↑	α, β-F1 ↑	ROC-AUC ↑	F1 ↑	Sampling ↓
ABIDE	Real Data	0.80 ± 0.04	0.81 ± 0.04	0.80 ± 0.01	0.66 ± 0.04	0.58 ± 0.06	N/A
	TriangCFM	0.21 ± 0.21	0.24 ± 0.35	0.20 ± 0.29	0.50 ± 0.05	0.40 ± 0.13	7.16 ± 1.17
	TriangDDPM	0.25 ± 0.04	0.22 ± 0.06	0.23 ± 0.04	0.52 ± 0.05	0.43 ± 0.08	5.45 ± 0.19
	DiffeoGauss	0.51 ± 0.04	0.23 ± 0.03	0.32 ± 0.03	**0.66 ± 0.05**	**0.56 ± 0.06**	**0.06 ± 0.01**
	DiffeoCFM	**0.81 ± 0.05**	0.50 ± 0.04	0.62 ± 0.05	0.60 ± 0.04	0.52 ± 0.07	9.04 ± 0.86
	GDSS-proj	0.00 ± 0.00	0.28 ± 0.07	0.00 ± 0.00	0.47 ± 0.05	0.29 ± 0.31	1.05 ± 0.02
	MAGNET (ours)	0.80 ± 0.04	**0.59 ± 0.04**	**0.68 ± 0.03**	0.64 ± 0.07	**0.56 ± 0.06**	0.38 ± 0.02
ADNI	Real Data	0.87 ± 0.09	0.87 ± 0.02	0.87 ± 0.05	0.58 ± 0.03	0.59 ± 0.03	N/A
	TriangCFM	0.23 ± 0.31	0.25 ± 0.35	0.23 ± 0.33	0.52 ± 0.05	0.37 ± 0.10	21.70 ± 3.35
	TriangDDPM	0.27 ± 0.06	0.30 ± 0.02	0.28 ± 0.03	0.54 ± 0.08	0.43 ± 0.05	13.92 ± 0.93
	DiffeoGauss	0.02 ± 0.01	0.50 ± 0.08	0.04 ± 0.02	**0.58 ± 0.03**	0.28 ± 0.13	1.42 ± 0.40
	DiffeoCFM	0.68 ± 0.12	0.75 ± 0.04	0.71 ± 0.05	0.57 ± 0.04	0.40 ± 0.09	36.08 ± 5.84
	GDSS-proj	0.57 ± 0.05	0.38 ± 0.07	0.45 ± 0.05	0.48 ± 0.07	0.53 ± 0.36	1.70 ± 0.01
	MAGNET (ours)	**0.71 ± 0.13**	**0.84 ± 0.06**	**0.76 ± 0.06**	0.57 ± 0.07	**0.61 ± 0.05**	**0.86 ± 0.28**
OASIS-3	Real Data	0.89 ± 0.04	0.88 ± 0.05	0.88 ± 0.01	0.71 ± 0.04	0.61 ± 0.06	N/A
	TriangCFM	0.22 ± 0.27	0.26 ± 0.37	0.23 ± 0.32	0.51 ± 0.05	0.45 ± 0.17	15.29 ± 2.51
	TriangDDPM	0.26 ± 0.08	0.31 ± 0.05	0.28 ± 0.05	0.55 ± 0.07	0.48 ± 0.03	8.79 ± 1.31
	DiffeoGauss	0.52 ± 0.06	0.31 ± 0.02	0.39 ± 0.03	**0.68 ± 0.03**	0.40 ± 0.09	1.88 ± 1.10
	DiffeoCFM	0.63 ± 0.06	0.38 ± 0.04	0.47 ± 0.05	**0.68 ± 0.03**	0.53 ± 0.04	23.24 ± 2.55
	GDSS-proj	0.01 ± 0.02	0.04 ± 0.01	0.01 ± 0.01	0.51 ± 0.03	0.30 ± 0.32	1.44 ± 0.01
	MAGNET (ours)	**0.75 ± 0.07**	**0.47 ± 0.06**	**0.57 ± 0.06**	**0.68 ± 0.03**	**0.56 ± 0.04**	**0.63 ± 0.00**

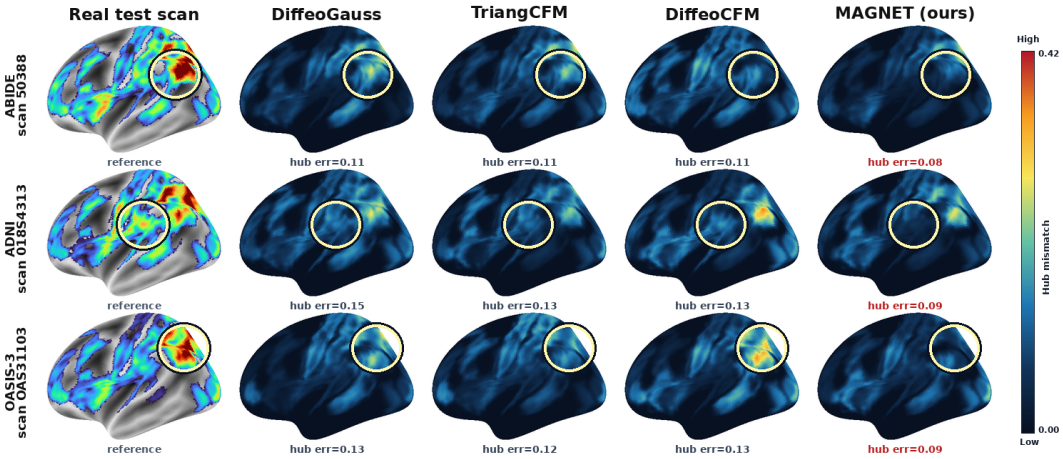

Figure 4: **Spatial mismatch of cortical hubs.** Cortical surface maps illustrating the structural deviation of generated samples relative to a held-out real test scan (left). Brighter regions indicate a larger spatial mismatch compared to the real reference, while darker regions signify higher fidelity. MAGNET consistently produces the darkest surfaces across all cohorts, visually confirming superior support recovery and yielding the lowest quantitative hub errors (shown below each panel).

component of $\hat{v}_\emptyset$ (might point in a conflicting direction to g) and can dilute directional control. By instead using $\hat{v}_\gamma = \hat{v}_\emptyset^\perp + \gamma \tanh(\|g\|_2) g$, the guidance direction is unambiguous and the $\tanh$ factor prevents over-steering for large $\|g\|_2$. The calibration is fit on the training split by matching class-conditional latent moments, and validity is preserved because decoding still proceeds through the normalized-Cholesky correlation chart (see App. F for ablations and sampling sensitivity analyses).

Table 2: **Zero-shot recovery of fine-grained clinical structure.** Generators are trained using only binary labels, but evaluated on resolved four-class clinical spaces: CN/SMC/MCI/AD for ADNI and CDR $0/0.5/1/2+$ for OASIS-3. Real held-out samples retain their metadata subclass labels, while generated samples are mapped into the same subclass space using a classifier trained only on real training data. Macro α-precision, β-recall, and α, β-F1 measure support recovery on the resolved subclass manifolds; multiclass CAS macro-F1 measures downstream utility; class KL divergence and Bhattacharyya distance measure preservation of subclass balance; and λ-MSE measures alignment of subclass-specific covariance structure. Bold and underlining denote the best and second-best generative models for each metric.

	Method	Quality (Macro-Averaged)			CAS	Distributional Preservation ↓		
		α-Prec. ↑	β-Recall ↑	α, β-F1 ↑	Macro F1 ↑	Class D_{KL}	Class D_B	λ-MSE
ADNI	TriangCFM	$\mathbf{0.51 \pm 0.11}$	0.21 ± 0.05	0.25 ± 0.05	0.19 ± 0.04	0.19 ± 0.07	0.05 ± 0.02	$\underline{9.16 \pm 0.16}$
	DiffeoGauss	0.00 ± 0.00	$\underline{0.54 \pm 0.04}$	0.01 ± 0.01	0.26 ± 0.03	$\underline{0.05 \pm 0.06}$	0.01 ± 0.01	11.37 ± 0.64
	DiffeoCFM	$\underline{0.38 \pm 0.16}$	0.48 ± 0.06	$\underline{0.33 \pm 0.07}$	$\underline{0.27 \pm 0.03}$	0.06 ± 0.08	$\underline{0.01 \pm 0.01}$	14.30 ± 0.89
	GDSS-proj	0.27 ± 0.10	0.30 ± 0.06	0.20 ± 0.04	0.19 ± 0.03	0.34 ± 0.16	0.09 ± 0.03	13.47 ± 10.59
	MAGNET (ours)	0.34 ± 0.07	$\mathbf{0.56 \pm 0.14}$	$\mathbf{0.35 \pm 0.04}$	$\mathbf{0.28 \pm 0.02}$	$\mathbf{0.05 \pm 0.03}$	$\mathbf{0.01 \pm 0.01}$	$\mathbf{6.08 \pm 0.53}$
OASIS-3	TriangCFM	0.59 ± 0.05	$\underline{0.19 \pm 0.06}$	$\underline{0.26 \pm 0.07}$	0.25 ± 0.05	0.03 ± 0.03	0.01 ± 0.01	$\underline{1.75 \pm 0.83}$
	DiffeoGauss	$\underline{0.61 \pm 0.06}$	0.15 ± 0.04	0.21 ± 0.05	0.34 ± 0.05	0.03 ± 0.03	0.01 ± 0.01	2.24 ± 0.14
	DiffeoCFM	0.60 ± 0.06	0.18 ± 0.05	0.26 ± 0.06	$\underline{0.34 \pm 0.04}$	$\underline{0.03 \pm 0.03}$	$\underline{0.01 \pm 0.01}$	1.93 ± 0.46
	GDSS-proj	0.00 ± 0.00	0.02 ± 0.01	0.00 ± 0.00	0.22 ± 0.07	0.11 ± 0.09	0.03 ± 0.02	4.00 ± 3.04
	MAGNET (ours)	$\mathbf{0.65 \pm 0.14}$	$\mathbf{0.21 \pm 0.05}$	$\mathbf{0.28 \pm 0.05}$	$\mathbf{0.36 \pm 0.04}$	$\mathbf{0.02 \pm 0.02}$	$\mathbf{0.01 \pm 0.00}$	$\mathbf{1.61 \pm 0.97}$

3 Experimental Results

We evaluate MAGNET on three neuroimaging cohorts: ABIDE [30], ADNI [83], and OASIS-3 [49]. Regional time series are extracted using Nilearn [22] with the MSDL atlas [80] ($d = 39$), z-scored, and converted to correlation matrices via OAS-estimated covariances [25]. The dataset comprises 4989 scans with strict subject-level splits: ABIDE (846 scans; 455 control, 391 positive), ADNI (2343 scans; 980 control, 1363 positive), and OASIS-3 (1800 scans; 1047 control, 753 positive).

In the primary benchmark, we use binary labels (control = 0, all diagnoses = 1), intentionally imposing coarse labels to test recovery of within-class structure (Sec. 3.1). A separate zero-shot multiclass evaluation assesses whether finer clinical subtypes are preserved despite binary training (Sec. 3.2). Baseline details and experimental setup are provided in App. C and App. D respectively.

3.1 Main Benchmark

Under matched sampling budgets, MAGNET improves upon prior baselines, delivering up to a 21% relative increase in class-conditional fidelity (α, β-F1) over the strongest validity-preserving baselines (Tab. 1). These gains are most pronounced on heterogeneous cohorts: 21% on OASIS-3 and 7% on ADNI. This suggests that MAGNET better captures the within-class clinical variation that confounds support recovery in these cohorts. Furthermore, this strict geometric fidelity directly translates to downstream utility: MAGNET achieves the highest or tied-best classifier-based F1 scores across all three cohorts. Crucially, MAGNET establishes this performance while simultaneously accelerating generation. Computational graph reparameterizations have similarly been explored to improve the efficiency of probabilistic deep learning [60]. The primary speedup over flow-matching baselines arises from the graph-denoiser architecture operating in discrete diffusion steps rather than continuous ODE integration; the amortized structural bridge provides an additional approximately $25 - 32\times$ reduction relative to teacher-bridge alternatives (Tab. 5).

Fig. 4 further confirms these gains topologically by visualizing the spatial mismatch of cortical hubs relative to real test scans. While baselines exhibit brighter surface regions indicative of larger spatial deviations, MAGNET generates noticeably darker surfaces that closely resemble the real distributions, yielding the lowest hub errors across all three cohorts. Finally, we note that generic discrete graph models (e.g., GDSS) fail to achieve this level of fidelity. Because these models are natively designed for sparse networks with discrete edge attributes, they cannot inherently enforce the strict positive-definite and unit-diagonal constraints of correlation matrices. Consequently, their reliance on post-hoc projection to force outputs into valid dense, continuous functional connectomes severely disrupts the generative trajectory and destroys subtle structural patterns, rendering them fundamentally insufficient for conditional support recovery (see App. C for detailed analysis).

3.2 Zero-Shot Recovery of Fine-Grained Clinical Subclasses

To test if binary-trained generators natively preserve unseen fine-grained clinical structure, we map generated samples into these subclasses using a classifier trained exclusively on real data. Because no generator is exposed to these labels during training, this zero-shot setting isolates the preservation of classifier-resolved clinical organization rather than validating biological taxonomy. This confound is controlled across methods by applying the same real-data-trained resolver to every generator.

Table 3: **Backbone and controller ablation.** All variants use the same normalized-Cholesky validity chart and a 6-step sampling budget. Controller columns denote class-anchored conditioning (Subtype), amortized bridge conditioning (Bridge), relevance-preserving noise (Noise), and diversity-preserving classifier-free guidance (Guidance). The vector row tests whether the validity chart alone is sufficient with a flat denoiser. The graph row restores the region-level graph denoiser but disables the controller. The individual controller rows activate one mechanism at a time on top of the graph denoiser. The penultimate row uses the integrated controller without diversity preserving classifier-free guidance, and the final row is full MAGNET.

Dataset	Backbone	Controller				Quality			Sampling (s.) ↓
		Subtype	Bridge	Noise	Guidance	α-Prec. ↑	β-Rec. ↑	α, β-F1 ↑	
ABIDE	Vector	✗	✗	✗	✗	0.52 ± 0.05	0.18 ± 0.03	0.26 ± 0.03	$\mathbf{0.04 \pm 0.00}$
	Graph	✗	✗	✗	✗	0.70 ± 0.05	0.46 ± 0.06	0.56 ± 0.05	0.36 ± 0.01
	Graph	✓	✗	✗	✗	0.72 ± 0.05	0.50 ± 0.06	0.59 ± 0.05	0.39 ± 0.01
	Graph	✗	✓	✗	✗	0.73 ± 0.06	0.48 ± 0.06	0.58 ± 0.06	0.38 ± 0.01
	Graph	✗	✗	✓	✗	0.71 ± 0.05	0.50 ± 0.04	0.59 ± 0.03	0.38 ± 0.00
	Graph	✓	✓	✓	✗	0.77 ± 0.06	0.54 ± 0.05	0.63 ± 0.04	0.38 ± 0.06
	Graph	✓	✓	✓	✓	$\mathbf{0.80 \pm 0.04}$	$\mathbf{0.59 \pm 0.04}$	$\mathbf{0.68 \pm 0.03}$	0.38 ± 0.02
ADNI	Vector	✗	✗	✗	✗	0.52 ± 0.01	0.25 ± 0.04	0.34 ± 0.04	$\mathbf{0.07 \pm 0.00}$
	Graph	✗	✗	✗	✗	0.61 ± 0.08	0.54 ± 0.08	0.57 ± 0.06	0.72 ± 0.01
	Graph	✓	✗	✗	✗	0.63 ± 0.06	0.72 ± 0.10	0.67 ± 0.06	0.74 ± 0.04
	Graph	✗	✓	✗	✗	0.62 ± 0.14	0.78 ± 0.07	0.69 ± 0.11	0.76 ± 0.02
	Graph	✗	✗	✓	✗	0.64 ± 0.10	0.76 ± 0.08	0.69 ± 0.07	0.75 ± 0.02
	Graph	✓	✓	✓	✗	0.68 ± 0.12	0.80 ± 0.04	0.74 ± 0.09	0.72 ± 0.01
	Graph	✓	✓	✓	✓	$\mathbf{0.71 \pm 0.13}$	$\mathbf{0.84 \pm 0.06}$	$\mathbf{0.76 \pm 0.06}$	0.86 ± 0.28
OASIS-3	Vector	✗	✗	✗	✗	0.58 ± 0.09	0.26 ± 0.04	0.36 ± 0.05	$\mathbf{0.05 \pm 0.00}$
	Graph	✗	✗	✗	✗	0.67 ± 0.08	0.44 ± 0.07	0.54 ± 0.08	0.60 ± 0.01
	Graph	✓	✗	✗	✗	0.68 ± 0.08	0.44 ± 0.07	0.54 ± 0.06	0.61 ± 0.02
	Graph	✗	✓	✗	✗	0.70 ± 0.06	0.46 ± 0.06	0.55 ± 0.06	0.60 ± 0.02
	Graph	✗	✗	✓	✗	0.70 ± 0.08	0.47 ± 0.05	0.56 ± 0.06	0.61 ± 0.02
	Graph	✓	✓	✓	✗	0.72 ± 0.07	0.46 ± 0.05	0.56 ± 0.06	0.59 ± 0.00
	Graph	✓	✓	✓	✓	$\mathbf{0.75 \pm 0.07}$	$\mathbf{0.47 \pm 0.06}$	$\mathbf{0.57 \pm 0.06}$	0.63 ± 0.00

Results. As detailed in Tab. 2, MAGNET achieves the strongest subclass covariance alignment in this zero-shot setting, reducing λ-MSE on ADNI by $33 - 57\%$ relative to baselines. Also, TriangCFM (a continuous euclidean baseline) over-indexes on precision at the severe expense of coverage. MAGNET also attains the best macro β-recall, α, β-F1, and CAS macro-F1 on ADNI, while sweeping all support-recovery, CAS, and λ-MSE metrics on OASIS-3. Across both datasets, MAGNET preserves subclass balance with the lowest KL divergence and tied Bhattacharyya distance. Distributional alignment has also been studied through optimal transport regularization [55]. Results indicate MAGNET better preserves fine-grained clinical organization under coarse binary supervision. Crucially, after generated samples are assigned to the resolved subclass space, λ-MSE directly compares subclass covariance spectra rather than classifier accuracy, providing complementary evidence of subclass preservation. In stark contrast, the generic discrete graph baseline (GDSS-proj) suffers catastrophic degradation in this fine-grained setting (e.g., dropping to zero F1 on OASIS-3), further emphasizing that post-hoc projection severely disrupts the subtle topological variations necessary for resolving unseen clinical subclasses (see App. C for details).

3.3 Graph Backbone and Controller Necessity

MAGNET relies on two core pillars: a validity-preserving chart for mathematical feasibility [21] and a graph-structured denoiser for relational inductive biases [25, 24, 58]. Tab. 3 isolates these components under a fixed six-step generation budget [75] to evaluate backbone necessity and controller synergy in resolving coarse-label deficits.

The structural gains of the graph backbone. Switching from the vector denoiser to the graph backbone yields substantial gains across all datasets. While the pullback chart guarantees mathematical validity, it lacks a mechanism for modeling ROI relationships; the graph denoiser supplies this crucial inductive bias, bridging mathematical soundness with anatomical relevance [25, 24, 58].

Coupled controller dynamics. The control components: subtype conditioning, amortized bridge, relevance-preserving noise, and classifier-free guidance [39] act as interdependent latent transforms rather than strictly additive modules. Individually, the amortized bridge provides the largest improvement, particularly on the highly diverse ADNI dataset (α, β-F1 improves from 0.57 to 0.69).

Table 4: **Generalization to higher-resolution ABIDE parcellations.** We evaluate MAGNET and the competing generators on AAL116, an anatomically defined 116-region atlas, and CC200, a functionally derived parcellation with approximately 200 regions. Both settings use the same quality-controlled 831-scan ABIDE cohort and 10-split group-aware evaluation protocol. We report class-conditional support fidelity, downstream classifier-based utility, and wall-clock sampling time. Results are mean $\pm$ standard deviation across splits. The Real Data row provides the empirical reference, and bold denotes the best generative model within each atlas.

Atlas	Method	α-Prec.↑	β-Rec.↑	α, β-F1↑	ROC-AUC↑	F1↑	Time (s)↓
AAL116	Real Data	0.78 ± 0.06	0.77 ± 0.09	0.77 ± 0.05	0.72 ± 0.09	0.63 ± 0.07	N/A
	TriangCFM	0.34 ± 0.09	0.02 ± 0.01	0.04 ± 0.02	0.49 ± 0.05	0.42 ± 0.19	3.80 ± 0.21
	TriangDDPM	0.41 ± 0.05	0.12 ± 0.07	0.37 ± 0.08	0.53 ± 0.08	0.48 ± 0.12	3.21 ± 0.32
	DiffeoGauss	0.55 ± 0.05	0.35 ± 0.07	0.43 ± 0.05	0.53 ± 0.07	0.32 ± 0.12	1.94 ± 0.39
	DiffeoCFM	0.58 ± 0.12	0.62 ± 0.08	0.59 ± 0.05	0.65 ± 0.08	0.54 ± 0.13	4.16 ± 0.24
	GDSS-proj	0.05 ± 0.02	0.21 ± 0.01	0.08 ± 0.03	0.51 ± 0.03	0.34 ± 0.09	1.43 ± 0.12
	MAGNET	$\mathbf{0.78 \pm 0.05}$	$\mathbf{0.62 \pm 0.08}$	$\mathbf{0.69 \pm 0.05}$	$\mathbf{0.68 \pm 0.06}$	$\mathbf{0.58 \pm 0.07}$	$\mathbf{1.12 \pm 0.07}$
CC200	Real Data	0.79 ± 0.06	0.66 ± 0.06	0.71 ± 0.02	0.76 ± 0.02	0.68 ± 0.04	N/A
	TriangCFM	0.39 ± 0.02	0.12 ± 0.05	0.18 ± 0.06	0.48 ± 0.03	0.47 ± 0.09	4.21 ± 0.13
	TriangDDPM	0.43 ± 0.02	0.21 ± 0.12	0.28 ± 0.11	0.55 ± 0.02	0.52 ± 0.11	3.71 ± 0.16
	DiffeoGauss	0.57 ± 0.03	0.41 ± 0.02	0.48 ± 0.02	0.58 ± 0.01	0.41 ± 0.08	1.73 ± 0.19
	DiffeoCFM	0.65 ± 0.05	0.53 ± 0.03	0.58 ± 0.03	0.63 ± 0.04	0.54 ± 0.07	3.78 ± 0.13
	GDSS-proj	0.08 ± 0.03	0.26 ± 0.11	0.12 ± 0.04	0.55 ± 0.06	0.41 ± 0.04	1.56 ± 0.07
	MAGNET	$\mathbf{0.77 \pm 0.09}$	$\mathbf{0.58 \pm 0.09}$	$\mathbf{0.66 \pm 0.06}$	$\mathbf{0.68 \pm 0.05}$	$\mathbf{0.59 \pm 0.05}$	$\mathbf{1.42 \pm 0.05}$

Generalization across atlas resolutions. To test whether MAGNET's gains depend on the compact parcellation, we repeat the ABIDE benchmark on two higher-resolution atlases: AAL116, an anatomically defined 116-region parcellation, and CC200, a functionally derived parcellation with approximately 200 regions. Both experiments use a common quality-controlled ABIDE cohort of 831 scans with CPAC preprocessing, band-pass filtering, and no global signal regression, together with the same 10-split group-aware evaluation protocol. These settings increase the dimensionality of the generation problem while varying the construction of the atlas. MAGNET remains the strongest generator at higher atlas resolutions, achieving the best α, β-F1 on AAL116 and CC200, with relative gains of 17% and 14% over DiffeoCFM, respectively (Tab. 4).

Biological plausibility and clinical structure. Beyond distributional fidelity, we evaluate neuroscientific plausibility through cortical hub preservation and network topology, established descriptors of biologically meaningful brain organization [78, 69]. Fig. 4 measures hub preservation, while Figs. 11-12 compare connectivity over fixed, real edge patterns. MAGNET further achieves the lowest subclass covariance error on ADNI and OASIS-3, including a 33–57% reduction in λ-MSE on ADNI (Tab. 2), while its gains persist on the higher-resolution AAL116 and CC200 atlases (Tab. 4).

4 Conclusion

Geometric validity is necessary for conditional connectome synthesis, but it is not sufficient for structural recovery. In this paper, we introduced MAGNET, a structure-aware generator that uses the normalized-Cholesky chart for validity and focuses the modeling problem on recovering class-conditional organization inside the valid correlation manifold. MAGNET combines an ROI-level graph denoiser with class-anchored subtype conditioning, an amortized structural bridge, and relevance-preserving corruption to reduce ambiguity from coarse labels, global structural context, and condition-relevant coordinates. Across three cohorts, MAGNET improves distributional fidelity, support recovery, zero-shot clinical alignment, and few-step sampling efficiency over strong baselines. MAGNET combines an ROI-level graph denoiser with class-anchored subtype conditioning, an amortized structural bridge, and relevance-preserving corruption to address ambiguity from coarse labels, missing global context, and corruption of condition-relevant coordinates. Future directions include multi-modal imaging inputs, longitudinal trajectory generation, and evaluation of synthetic cohort utility in downstream clinical prediction. More broadly, the amortized bridge principle of distilling structural summaries into fast inference-time predictors could be extended to other constrained generative settings beyond connectomes.

Acknowledgment Prof. Lokhande acknowledges support from University at Buffalo startup funds, an Adobe Research Gift, an NVIDIA Academic Grant, and the National Center for Advancing Translational Sciences of the NIH (award UM1TR005296 to the University at Buffalo). Prof. Kim's contribution to this project is supported by RS-2019-II191906 (Artificial Intelligence Graduate School, POSTECH).

References

[1] Hayoung Ahn, Seungjoo Lee, Jaeyoon Sim, Yechan Hwang, Hyuna Cho, Guorong Wu, and Won Hwa Kim. Spectral graph autoregressive modeling for conditional brain network augmentation. In *International Workshop on Reconstruction and Imaging Motion Estimation*, pages 209–218. Springer, 2025.

[2] Ahmed Alaa, Boris Van Breugel, Evgeny S Saveliev, and Mihaela Van Der Schaar. How Faithful is your Synthetic Data? Sample-level Metrics for Evaluating and Auditing Generative Models. In *International conference on machine learning*, pages 290–306. PMLR, 2022.

[3] Ilya Archakov and Peter Reinhard Hansen. A new parametrization of correlation matrices. *Econometrica*, 89(4):1699–1715, 2021.

[4] Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Log-euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 56(2):411–421, 2006.

[5] Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM journal on matrix analysis and applications*, 29(1):328–347, 2007.

[6] Peter L Bartlett, Dylan J Foster, and Matus J Telgarsky. Spectrally-normalized margin bounds for neural networks. *Advances in neural information processing systems*, 30, 2017.

[7] Peter W Battaglia, Jessica B Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, et al. Relational Inductive Biases, Deep Learning, and Graph Networks. *arXiv preprint arXiv:1806.01261*, 2018.

[8] Yaniv Benny, Tomer Galanti, Sagie Benaim, and Lior Wolf. Evaluation metrics for conditional image generation. *International Journal of Computer Vision*, 129(5):1712–1731, 2021.

[9] Marco Bertolini, Tuan Le, and Djork-Arné Clevert. Diffusion generative modeling on lie group representations. *arXiv preprint arXiv:2502.02513*, 2025.

[10] Alaa Bessadok, Mohamed Ali Mahjoub, and Islem Rekik. Graph neural networks in network neuroscience. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):5833–5848, 2022.

[11] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric Deep Learning: Going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

[12] Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Velickovic. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

[13] Mateusz Buda, Atsuto Maki, and Maciej A Mazurowski. A systematic study of the class imbalance problem in convolutional neural networks. *Neural networks*, 106:249–259, 2018.

[14] Jonathon Byrd and Zachary Lipton. What is the effect of importance weighting in deep learning? In *International conference on machine learning*, pages 872–881. PMLR, 2019.

[15] F Xavier Castellanos, Adriana Di Martino, R Cameron Craddock, Ashesh D Mehta, and Michael P Milham. Clinical applications of the functional connectome. *Neuroimage*, 80: 527–540, 2013.

[16] Ting Chen, Ruixiang Zhang, and Geoffrey Hinton. Analog bits: Generating discrete data using diffusion models with self-conditioning. *arXiv preprint arXiv:2208.04202*, 2022.

[17] Hyuna Cho, Jaeyoon Sim, Guorong Wu, and Won Hwa Kim. Neurodegenerative brain network classification via adaptive diffusion with temporal regularization. In *Forty-first International Conference On Machine Learning*, 2024.

[18] Hyuna Cho, Ziquan Wei, Seungjoo Lee, Tingting Dan, Guorong Wu, and Won Hwa Kim. Conditional diffusion with ordinal regression: Longitudinal data generation for neurodegenerative disease studies. In *international conference on learning representations*, volume 2025, pages 33329–33350, 2025.

[19] Hyungjin Chung, Jeongsol Kim, Geon Yeong Park, Hyelin Nam, and Jong Chul Ye. CFG++: Manifold-constrained classifier free guidance for diffusion models. *arXiv preprint arXiv:2406.08070*, 2024.

[20] Sotirios Panagiotis Chytas, Vishnu Lokhande, and Vikas Singh. Pooling image datasets with multiple covariate shift and imbalance. In *International Conference on Learning Representations*, volume 2024, pages 57282–57304, 2024.

[21] Antoine Collas, Ce Ju, Nicolas Salvy, and Bertrand Thirion. Riemannian flow matching for brain connectivity matrices via pullback geometry. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL `https://openreview.net/forum?id=NY3LzmUXl7`.

[22] Nilearn contributors. nilearn. URL `https://github.com/nilearn/nilearn`.

[23] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27, 1967.

[24] Hejie Cui, Wei Dai, Yanqiao Zhu, Xuan Kan, Antonio Aodong Chen Gu, Joshua Lukemire, Liang Zhan, Lifang He, Ying Guo, and Carl Yang. BrainGB: a benchmark for brain network analysis with graph neural networks. *IEEE transactions on medical imaging*, 42(2):493–506, 2022.

[25] Kamalaker Dadi, Mehdi Rahim, Alexandre Abraham, Darya Chyzhyk, Michael Milham, Bertrand Thirion, Gaël Varoquaux, Alzheimer's Disease Neuroimaging Initiative, et al. Benchmarking functional connectome-based predictive models for resting-state fmri. *NeuroImage*, 192:115–134, 2019.

[26] Valentin De Bortoli, Emile Mathieu, Michael Hutchinson, James Thornton, Yee Whye Teh, and Arnaud Doucet. Riemannian Score-based Generative Modelling. *Advances in neural information processing systems*, 35:2406–2422, 2022.

[27] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society: series B (methodological)*, 39 (1):1–22, 1977.

[28] Luc Devroye, László Györfi, and Gábor Lugosi. *A probabilistic theory of pattern recognition*, volume 31. Springer Science & Business Media, 2013.

[29] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

[30] 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.

[31] Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011.

[32] Chris Fraley and Adrian E Raftery. Model-based clustering, discriminant analysis, and density estimation. *Journal of the American statistical Association*, 97(458):611–631, 2002.

[33] Yisen Gao, Xingcheng Fu, Qingyun Sun, Jianxin Li, and Xianxian Li. Toward a unified geometry understanding: Riemannian diffusion framework for graph generation and prediction. *arXiv preprint arXiv:2510.04522*, 2025.

[34] Ian Goodfellow. Nips 2016 tutorial: Generative adversarial networks. *arXiv preprint arXiv:1701.00160*, 2016.

[35] Neil He, Jiahong Liu, Buze Zhang, Ngoc Bui, Ali Maatouk, Menglin Yang, Irwin King, Melanie Weber, and Rex Ying. Position: Beyond Euclidean–Foundation Models Should Embrace Non-Euclidean Geometries. *arXiv preprint arXiv:2504.08896*, 2025.

[36] Yutong He, Naoki Murata, Chieh-Hsin Lai, Yuhta Takida, Toshimitsu Uesaka, Dongjun Kim, Wei-Hsiang Liao, Yuki Mitsufuji, J Zico Kolter, Ruslan Salakhutdinov, et al. Manifold Preserving Guided Diffusion. *arXiv preprint arXiv:2311.16424*, 2023.

[37] Alex Henry, Prudhvi Raj Dachapally, Shubham Shantaram Pawar, and Yuxuan Chen. Query-key normalization for transformers. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4246–4253, 2020.

[38] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

[39] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.

[40] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[41] Chin-Wei Huang, Milad Aghajohari, Joey Bose, Prakash Panangaden, and Aaron C Courville. Riemannian Diffusion Models. *Advances in Neural Information Processing Systems*, 35: 2750–2761, 2022.

[42] Nathalie Japkowicz and Shaju Stephen. The class imbalance problem: A systematic study. *Intelligent data analysis*, 6(5):429–449, 2002.

[43] Jaehyeong Jo, Seul Lee, and Sung Ju Hwang. Score-based generative modeling of graphs via the system of stochastic differential equations. In *International conference on machine learning*, pages 10362–10383. PMLR, 2022.

[44] Xuan Kan, Wei Dai, Hejie Cui, Zilong Zhang, Ying Guo, and Carl Yang. Brain Network Transformer. *Advances in Neural Information Processing Systems*, 35:25586–25599, 2022.

[45] Bingyi Kang, Saining Xie, Marcus Rohrbach, Zhicheng Yan, Albert Gordo, Jiashi Feng, and Yannis Kalantidis. Decoupling representation and classifier for long-tailed recognition. *arXiv preprint arXiv:1910.09217*, 2019.

[46] Leonid V Kantorovich. The mathematical method of production planning and organization. *Management Science*, 6(4):363–422, 1939.

[47] Tauhid Khan. Isokinetic flow matching for pathwise straightening of generative flows. *arXiv preprint arXiv:2604.04491*, 2026.

[48] Tuomas Kynkäänniemi, Miika Aittala, Tero Karras, Samuli Laine, Timo Aila, and Jaakko Lehtinen. Applying guidance in a limited interval improves sample and distribution quality in diffusion models. *Advances in Neural Information Processing Systems*, 37:122458–122483, 2024.

[49] Pamela J LaMontagne, Tammie LS Benzinger, John C Morris, Sarah Keefe, Russ Hornbeck, Chengjie Xiong, Elizabeth Grant, Jason Hassenstab, Krista Moulder, Andrei G Vlassenko, et al. OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease. *medrxiv*, page 12, 2019.

[50] Kyowoon Lee and Jaesik Choi. Local manifold approximation and projection for manifold-aware diffusion planning. *arXiv preprint arXiv:2506.00867*, 2025.

[51] Ziyun Li, Huancheng Hu, Soon Hoe Lim, Xuyu Li, Fei Gao, Enmao Diao, Zezhen Ding, Michalis Vazirgiannis, and Henrik Bostrom. A kinetic-energy perspective of flow matching. *arXiv preprint arXiv:2602.07928*, 2026.

[52] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow Matching for Generative Modeling. *arXiv preprint arXiv:2210.02747*, 2022.

[53] Haggai Maron, Heli Ben-Hamu, Nadav Shamir, and Yaron Lipman. Invariant and equivariant graph networks. *arXiv preprint arXiv:1812.09902*, 2018.

[54] Geoffrey J McLachlan, Sharon X Lee, and Suren I Rathnayake. Finite Mixture Models. *Annual review of statistics and its application*, 6(1):355–378, 2019.

[55] Ronak Mehta, Jeffery Kline, Vishnu Suresh Lokhande, Glenn Fung, and Vikas Singh. Efficient discrete multi marginal optimal transport regularization. In *The Eleventh International Conference on Learning Representations*, 2023. URL `https://openreview.net/forum?id=R98ZfMt-jE`.

[56] Koichi Miyasawa et al. An empirical bayes estimator of the mean of a normal population. *Bull. Inst. Internat. Statist*, 38(181-188):1–2, 1961.

[57] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. *arXiv preprint arXiv:1802.05957*, 2018.

[58] Hamed Mohammadi and Waldemar Karwowski. Graph neural networks in brain connectivity studies: Methods, challenges, and future directions. *Brain Sciences*, 15(1):17, 2024.

[59] Zeinab Navidi, Akshaya Thoutam, Madeline Hughes, Srivatsan Raghavan, Peter S Winter, Lorin Crawford, and Ava P Amini. Adaptive resampling for improved machine learning in imbalanced single-cell datasets. *bioRxiv*, page 11, 2025.

[60] Jurijs Nazarovs, Ronak R Mehta, Vishnu Suresh Lokhande, and Vikas Singh. Graph reparameterizations for enabling 1000+ monte carlo iterations in bayesian deep neural networks. In *Uncertainty in Artificial Intelligence*, pages 118–128. PMLR, 2021.

[61] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pages 8162–8171. PMLR, 2021.

[62] Youngrok Park, Hojung Jung, Sangmin Bae, and Se-Young Yun. Temporal alignment guidance: On-manifold sampling in diffusion models. *arXiv preprint arXiv:2510.11057*, 2025.

[63] William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4195–4205, 2023.

[64] Xavier Pennec, Pierre Fillard, and Nicholas Ayache. A riemannian framework for tensor computing. *International Journal of computer vision*, 66(1):41–66, 2006.

[65] Ethan Perez, Florian Strub, Harm De Vries, Vincent Dumoulin, and Aaron Courville. FiLM: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[66] Sean Pinkney. A short note on a flexible cholesky parameterization of correlation matrices. *arXiv preprint arXiv:2405.07286*, 2024.

[67] Lyle Regenwetter, Akash Srivastava, Dan Gutfreund, and Faez Ahmed. Beyond statistical similarity: Rethinking metrics for deep generative models in engineering design. *Computer-Aided Design*, 165:103609, 2023.

[68] Ayush Roy, Wei-Yang Alex Lee, Rudrasis Chakraborty, and Vishnu Suresh Lokhande. ManifoldGD: Training-free hierarchical manifold guidance for diffusion-based dataset distillation. *arXiv preprint arXiv:2602.23295*, 2026.

[69] Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069, 2010.

[70] Seyedmorteza Sadat, Jakob Buhmann, Derek Bradley, Otmar Hilliges, and Romann M Weber. CADS: Unleashing the diversity of diffusion models through condition-annealed sampling. *arXiv preprint arXiv:2310.17347*, 2023.

[71] Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, pages 461–464, 1978.

[72] David W Scott. Multivariate density estimation and visualization. In *Handbook of computational statistics: Concepts and methods*, pages 549–569. Springer, 2011.

[73] Noam Shazeer. GLU variants improve transformer. *arXiv preprint arXiv:2002.05202*, 2020.

[74] Stephen M Smith, Diego Vidaurre, Christian F Beckmann, Matthew F Glasser, Mark Jenkinson, Karla L Miller, Thomas E Nichols, Emma C Robinson, Gholamreza Salimi-Khorshidi, Mark W Woolrich, et al. Functional connectomics from resting-state fmri. *Trends in cognitive sciences*, 17(12):666–682, 2013.

[75] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.

[76] Olaf Sporns, Giulio Tononi, and Rolf Kötter. The Human Connectome: a structural description of the human brain. *PLoS computational biology*, 1(4):e42, 2005.

[77] Duo Su, Huyu Wu, Huanran Chen, Yiming Shi, Yuzhu Wang, Xi Ye, and Jun Zhu. Diffusion models as dataset distillation priors. *arXiv preprint arXiv:2510.17421*, 2025.

[78] Martijn P Van den Heuvel and Olaf Sporns. Network hubs in the human brain. *Trends in cognitive sciences*, 17(12):683–696, 2013.

[79] Erdem Varol, Aristeidis Sotiras, Christos Davatzikos, Alzheimer's Disease Neuroimaging Initiative, et al. HYDRA: Revealing heterogeneity of imaging and genetic patterns through a multiple max-margin discriminative analysis framework. *Neuroimage*, 145:346–364, 2017.

[80] Gaël Varoquaux, Alexandre Gramfort, Fabian Pedregosa, Vincent Michel, and Bertrand Thirion. Multi-subject dictionary learning to segment an atlas of brain spontaneous activity. In *Biennial International Conference on information processing in medical imaging*, pages 562–573. Springer, 2011.

[81] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[82] Clement Vignac, Igor Krawczuk, Antoine Siraudin, Bohan Wang, Volkan Cevher, and Pascal Frossard. DiGress: Discrete denoising diffusion for graph generation. *arXiv preprint arXiv:2209.14734*, 2022.

[83] Michael W Weiner, Dallas P Veitch, Paul S Aisen, Laurel A Beckett, Nigel J Cairns, Robert C Green, Danielle Harvey, Clifford R Jack Jr, William Jagust, John C Morris, et al. The Alzheimer's Disease Neuroimaging Initiative 3: Continued innovation for clinical trial improvement. *Alzheimer's & Dementia*, 13(5):561–571, 2017.

[84] Junhao Wen, Mathilde Antoniades, Zhijian Yang, Gyujoon Hwang, Ioanna Skampardoni, Rongguang Wang, and Christos Davatzikos. Dimensional neuroimaging endophenotypes: neurobiological representations of disease heterogeneity through machine learning. *Biological psychiatry*, 96(7):564–584, 2024.

[85] Thomas Wolfers, Nhat Trung Doan, Tobias Kaufmann, Dag Alnæs, Torgeir Moberget, Ingrid Agartz, Jan K Buitelaar, Torill Ueland, Ingrid Melle, Barbara Franke, et al. Mapping the heterogeneous phenotype of schizophrenia and bipolar disorder using normative models. *JAMA psychiatry*, 75(11):1146–1155, 2018.

[86] Yu-Han Wu, Quentin Berthet, Gérard Biau, Claire Boyer, Romuald Elie, and Pierre Marion. Optimal stopping in latent diffusion models. *arXiv preprint arXiv:2510.08409*, 2025.

[87] Ruibin Xiong, Yunchang Yang, Di He, Kai Zheng, Shuxin Zheng, Chen Xing, Huishuai Zhang, Yanyan Lan, Liwei Wang, and Tieyan Liu. On layer normalization in the transformer architecture. In *International conference on machine learning*, pages 10524–10533. PMLR, 2020.

[88] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola. Deep sets. *Advances in neural information processing systems*, 30, 2017.

[89] Zhengli Zhao, Sameer Singh, Honglak Lee, Zizhao Zhang, Augustus Odena, and Han Zhang. Improved consistency regularization for gans. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 11033–11041, 2021.

A Notations

We summarize the main notation used.

Symbol	Meaning
$\chi(C)$	Normalized-Cholesky chart coordinate of C
x_t	Noisy latent state at diffusion time t
s	Train-only subtype token within class y
b_t	Amortized bridge state at time t
$\varphi(\cdot)$	Structural feature extractor used by the bridge and spectral loss
$B_\psi(x_t, t, s)$	Amortized bridge predictor
$b^\star(x_0, s, t)$	Teacher bridge target
L_{bridge}	Bridge distillation loss
α, β-F1	Harmonic mean of α-precision and β-recall
$\mathcal{C}_d$	Set of $d \times d$ valid correlation connectomes
$C \in \mathcal{C}_d$	A correlation connectome
$y \in \{0, 1\}$	Class label
P^y_{val}	Held-out real distribution for class y
Q^y_m	Generated distribution for method m, class y
$\Phi(C) = \log C$	Log-Euclidean map to $\text{Sym}(d)$
$d_{\text{LE}}(C_1, C_2)$	Log-Euclidean distance (Eq. 14)
$\mu^{\log,\text{val}}_y$	Mean of $\log C$ under P^y_{val}
$\bar{C}^{\text{val}}_{\text{LE},y}$	Log-Euclidean Fréchet mean
$r_y(C)$	Radial distance from real mean (Eq. 15)
$\nu^{\text{val}}_y, \nu^m_y$	Radial pushforward distributions
$W^{m,y}_{\text{rad}}$	Radial Wasserstein distance (Eq. 16)
$E^{m,y}_{\text{rad}}$	Radial quantile calibration error (Eq. 17)
$z^{\text{val},y}_i$	Log-centered vectorized connectome
W_y	PCA basis (top 2 directions from real data)
$u_y(C)$	2D chart coordinate (Eq. 19)
$\text{ISR}_{m,y}$	Inside-support rate (Eq. 21)
$C^y_{m,t}$	Generated sample at reverse-time t
$\text{ISR}_{m,y}(t)$	Time-dependent inside-support rate (Eq. 22)
$\bar{C}^y_{m,t}$	Generated mean connectome at time t
$\bar{C}^y_{\text{val}}$	Real mean connectome
M_y	Real-defined edge mask
$\text{MAE}^y_{m,t}$	Mean absolute error on mask (Eq. 24)
$\text{Corr}^y_{m,t}$	Correlation on masked edges (Eq. 25)

B Proofs of Formal Results

This appendix collects the proofs of the formal results stated in Section 2. The main text uses these statements to justify the design of class-anchored conditioning, the amortized structural bridge, and relevance-preserving corruption while keeping the method section visually compact.

B.1 Proof of Proposition 1

The result follows from the L^2-projection property of conditional expectation which is a generalization of the Rao-Blackwell theorem to the squared-loss denoising setting. We include the derivation for completeness. The key non-trivial empirical content is that the gain is *strictly positive* whenever S encodes structure that Y collapses. Fig. 2 also confirms visually for all three cohorts.

Proof. For squared loss, the Bayes estimators are the conditional means: $f^\star(X_t, Y) = \mathbb{E}[X_0 \mid X_t, Y]$ and $g^\star(X_t, Y, S) = \mathbb{E}[X_0 \mid X_t, Y, S]$. We decompose the coarser residual as

$$X_0 - \mathbb{E}[X_0 \mid X_t, Y] = \underbrace{X_0 - \mathbb{E}[X_0 \mid X_t, Y, S]}_{A} + \underbrace{\mathbb{E}[X_0 \mid X_t, Y, S] - \mathbb{E}[X_0 \mid X_t, Y]}_{B}.$$

Since A is the residual from the finer predictor, $\mathbb{E}[A \mid X_t, Y, S] = 0$. Because B is $\sigma(X_t, Y, S)$-measurable, the cross term $\mathbb{E}[\langle A, B\rangle] = 0$. Taking squared norms and expectations yields Eq. (4) directly. Strict positivity holds whenever $B \not\equiv 0$ with positive probability, i.e., whenever S resolves within-class structure unresolved by Y. □

Remark (empirical content). The identity converts the entire visual motivation of Fig. 2 into a concrete optimality criterion. The gain is zero iff the subtype provides no additional information beyond the coarse label. The heterogeneity entropy $\mathbf{H}$ analysis in App. F.4 confirms this is non-zero and scales with dataset complexity.

B.2 Proof of Proposition 2

Proof. The risk improvement identity Eq. (7) follows by the same L^2-projection argument as Prop. 1, with (U) in place of (X_t, Y) and $(U, B^\star)$ in place of (X_t, Y, S); we omit the repetition. The result is the amortization bound of Eq. (8), which is proved below.

$$f^\star(U) = \mathbb{E}[X_0 \mid U],$$
$$g^\star(U, B^\star) = \mathbb{E}[X_0 \mid U, B^\star].$$

Substituting these into Eq. 6 we get

$$\mathcal{R}^\star_U = \mathbb{E}\left[\|X_0 - \mathbb{E}[X_0 \mid U]\|_2^2\right],$$
$$\mathcal{R}^\star_{U,B^\star} = \mathbb{E}\left[\|X_0 - \mathbb{E}[X_0 \mid U, B^\star]\|_2^2\right].$$

Now we decompose the coarser residual as

$$X_0 - \mathbb{E}[X_0 \mid U] = \underbrace{X_0 - \mathbb{E}[X_0 \mid U, B^\star]}_{A} + \underbrace{\mathbb{E}[X_0 \mid U, B^\star] - \mathbb{E}[X_0 \mid U]}_{B}.$$

Now, upon squaring we get

$$\|X_0 - \mathbb{E}[X_0 \mid U]\|_2^2 = \|A\|_2^2 + \|B\|_2^2 + 2\langle A, B\rangle.$$

We now condition on $(U, B^\star)$. Since $A = X_0 - \mathbb{E}[X_0 \mid U, B^\star]$, we have

$$\mathbb{E}[A \mid U, B^\star] = 0.$$

Because B is measurable with respect to $(U, B^\star)$, we get

$$\mathbb{E}[\langle A, B\rangle \mid U, B^\star] = \langle \mathbb{E}[A \mid U, B^\star],\, B\rangle = 0.$$

Now, we take the conditional expectation followed by an expectation again, we get

$$\mathcal{R}^\star_U = \mathcal{R}^\star_{U,B^\star} + \mathbb{E}\Big[\|\mathbb{E}[X_0 \mid U, B^\star] - \mathbb{E}[X_0 \mid U]\|_2^2\Big],$$

which proves Eq. 7.

We now prove the amortization bound in Eq. 8. Let $\eta(u, b) = \mathbb{E}[X_0 \mid U = u, B^\star = b]$ and we define the plug-in predictor $\widehat{X}_0(U) = \eta(U, B_\psi(U))$. Inserting and subtracting the bridge-aware Bayes predictor, we get:

$$X_0 - \widehat{X}_0(U) = \underbrace{X_0 - \eta(U, B^\star)}_{A} + \underbrace{\eta(U, B^\star) - \eta(U, B_\psi(U))}_{B}.$$

Now, we take the square and the expectation to yield

$$\mathbb{E}\Big[\|X_0 - \widehat{X}_0(U)\|_2^2\Big] = \mathbb{E}\big[\|A\|_2^2\big] + \mathbb{E}\big[\|B\|_2^2\big] + 2\,\mathbb{E}[\langle A, B\rangle]\,.$$

Further, we condition on $(U, B^\star, B_\psi(U))$. Since $\eta(U, B^\star) = \mathbb{E}[X_0 \mid U, B^\star]$, the residual A has zero conditional mean given $(U, B^\star)$, and therefore also given $(U, B^\star, B_\psi(U))$. Hence

$$\mathbb{E}[\langle A, B\rangle \mid U, B^\star, B_\psi(U)] = \langle \mathbb{E}[A \mid U, B^\star, B_\psi(U)],\, B\rangle = 0.$$

Thus the cross term vanishes, and the first term is exactly $\mathcal{R}^\star_{U,B^\star}$, so

$$\mathbb{E}\Big[\|X_0 - \widehat{X}_0(U)\|_2^2\Big] - \mathcal{R}^\star_{U,B^\star} = \mathbb{E}\big[\|\eta(U, B^\star) - \eta(U, B_\psi(U))\|_2^2\big]\,.$$

By the assumed Lipschitz property of η in its bridge argument, we get

$$\|\eta(U, B^\star) - \eta(U, B_\psi(U))\|_2 \le L\,\|B^\star - B_\psi(U)\|_2.$$

Again, upon squaring and taking the expectation we get,

$$\mathbb{E}\big[\|\eta(U, B^\star) - \eta(U, B_\psi(U))\|_2^2\big] \le L^2\,\mathbb{E}\big[\|B_\psi(U) - B^\star\|_2^2\big]\,.$$

By finally combining the last two equations, we prove Eq. 8. □

Lipschitz condition on the Bayes regressor. The L-Lipschitz condition on $\eta(\cdot, b) = \mathbb{E}[X_0 \mid U, B^\star = b]$ is a regularity assumption on the underlying data distribution: it requires that the population-level conditional expectation varies smoothly with the bridge input. This is a standard smoothness assumption in nonparametric estimation [56]. Separately, to confirm that the error bound in Eq. (8) remains finite in practice, we note that the deployed bridge predictor B_ψ, parameterized as a D-layer MLP with weight matrices $\{W_d\}$ and σ-Lipschitz activations, satisfies $\mathrm{Lip}(B_\psi) \le \sigma^{D-1}\prod_{d=1}^{D}\|W_d\|_{\mathrm{op}}$, which is finite under ℓ_2 weight regularization [6, 57]. This confirms numerical tractability but does not substitute for the distributional regularity assumption on η. Because $b^\star$ is a convex combination of atoms from a compact set $\mathcal{A}_s$ with diameter $\delta_s = \max_{j,k}\|a_{s,j} - a_{s,k}\|_2 < \infty$, the bridge input is bounded: $\|b^\star\|_2 \le \max_k \|a_{s,k}\|_2$. Because $b^\star$ is a convex combination of atoms from the finite set $\mathcal{A}_s$, the bridge input is restricted to the compact domain $\mathrm{conv}(\mathcal{A}_s)$ and is therefore bounded. Compactness alone does not imply Lipschitz continuity of the Bayes regressor; accordingly, Eq. (8) explicitly assumes that $\eta(u, b) = \mathbb{E}[X_0 \mid U = u, B^\star = b]$ is L-Lipschitz in b on the relevant bridge domain. A sufficient condition is that η be continuously differentiable with respect to b on a neighborhood of this domain with uniformly bounded Jacobian. The bounded bridge domain therefore controls where this regularity assumption is required, rather than implying the assumption itself. This is a distributional property and cannot be inferred from the Lipschitz constant of the deployed approximator B_ψ.

Although our amortized bridge is trained against a teacher target *(drawing directly on the principles of knowledge distillation [38])*, our structural objective is entirely different. In traditional distillation, the capacity is shrunk in the same output space, while we use the structure for addressing a lack of context, to provide a macroscopic overview of the conditional probability distribution which cannot be captured by an isolated state X_t via few-step sampling [39, 16]. Equation 8 represents this difference by constraining the amortization tracking error rather than shrinking the capacity. As shown in Appendix F.1, tracking the teacher target effectively leads to increased generative accuracy, allowing us to maintain structural properties of the teacher bridge with up to approximately $25 - 32\times$ savings in sampling cost.

Note. i) Prop. 1 and 1 are *design-justification* results, not end-to-end generalization guarantees with the purpose of showing the information-theoretic direction of MAGNET: adding a subtype variable or a structural bridge can only reduce Bayes denoising risk, and cannot hurt in the oracle limit. This rules out the possibility that the architecture is fundamentally misspecified even under ideal conditions, which is a nontrivial baseline to establish before empirical validation. ii) For the learned regime, *we deliberately rely on empirical evidence*: Tab. 5 shows that lower amortization tracking error directly correlates with higher α, β-F1, empirically validating the amortization bound (Eq. 8) in the regime where it matters. Appendix F.4 shows that the class-anchored conditioning gain scales with dataset heterogeneity entropy $\mathbf{H}$, empirically validating that the learned subtypes capture within-class structure rather than being a regularization artifact. App. B.2 discusses why finiteness of L is guaranteed on the relevant domain by compactness of the atom set $\mathcal{A}_s$, and we clarify there that *this is a distributional regularity condition rather than a claim about the deployed network*.

B.3 A surrogate result for relevance-preserving corruption

The adaptive corruption rule in Section 2.3.3 is data-driven: the relevant directions are estimated from the conditional–unconditional guidance difference introduced in Section 2.3.3. To formalize the underlying design principle, we analyze a stylized Gaussian surrogate.

Lemma 1 (Adaptive corruption preserves class separation under a Gaussian surrogate)**.** *Let the clean latent decompose as $X_0 = (U, V)$, where $U \in \mathbb{R}^r$ is the class-relevant subspace and $V \in \mathbb{R}^q$ is the nuisance subspace. Assume that conditioned on $Y = y$, $U \mid Y = y \sim \mathcal{N}(\mu_y, \Sigma_U)$ and $V \mid Y = y \sim \mathcal{N}(0, \Sigma_V)$, with $\Sigma_U \succ 0$, $\Sigma_V \succ 0$, and $U \perp V \mid Y$. For fixed diffusion time t, consider the anisotropic corruption model*

$$X_t = \sqrt{\bar{\alpha}_t}\, X_0 + \sqrt{1-\bar{\alpha}_t}\, \xi, \qquad \xi \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma_{\mathrm{rel}}^2 I_r & 0 \\ 0 & \sigma_{\mathrm{nui}}^2 I_q \end{bmatrix}\right). \tag{11}$$

Let $\Delta = \mu_1 - \mu_0$, and define the class-separation functional

$$D_t(\sigma_{\mathrm{rel}}^2) = (\mu_{t,1} - \mu_{t,0})^\top \Sigma_t^{-1} (\mu_{t,1} - \mu_{t,0}), \tag{12}$$

where $\mu_{t,y}$ and Σ_t are the class-conditional mean and covariance of X_t. Then

$$D_t(\sigma_{\mathrm{rel}}^2) = \bar{\alpha}_t\, \Delta^\top \left(\bar{\alpha}_t \Sigma_U + (1-\bar{\alpha}_t)\sigma_{\mathrm{rel}}^2 I_r\right)^{-1} \Delta, \tag{13}$$

which is strictly decreasing in σ_{rel}^2 whenever $\Delta \neq 0$.

Proof. Under Eq. 11, the conditional distribution of X_t given $Y = y$ is Gaussian with mean

$$\mu_{t,y} = \sqrt{\bar{\alpha}_t} \begin{bmatrix} \mu_y \\ 0 \end{bmatrix}$$

and covariance

$$\Sigma_t = \begin{bmatrix} \bar{\alpha}_t \Sigma_U + (1-\bar{\alpha}_t)\sigma_{\mathrm{rel}}^2 I_r & 0 \\ 0 & \bar{\alpha}_t \Sigma_V + (1-\bar{\alpha}_t)\sigma_{\mathrm{nui}}^2 I_q \end{bmatrix}.$$

Since only the relevant block carries class-dependent means,

$$\mu_{t,1} - \mu_{t,0} = \sqrt{\bar{\alpha}_t} \begin{bmatrix} \Delta \\ 0 \end{bmatrix}.$$

Substituting these expressions into Eq. 12 yields Eq. 13.

To prove monotonicity, write $A = \bar{\alpha}_t \Sigma_U$, $s = \sigma_{\mathrm{rel}}^2$, and $\beta = 1 - \bar{\alpha}_t$. Then

$$D_t(s) = \bar{\alpha}_t\, \Delta^\top (A + \beta s I_r)^{-1} \Delta.$$

Differentiate with respect to s. Using the matrix derivative identity

$$\frac{d}{ds}(A + \beta s I_r)^{-1} = -\beta (A + \beta s I_r)^{-2},$$

we obtain

$$\frac{d}{ds} D_t(s) = -\bar{\alpha}_t \beta\, \Delta^\top (A + \beta s I_r)^{-2} \Delta.$$

Because $A + \beta s I_r$ is positive definite, $(A + \beta s I_r)^{-2}$ is positive definite as well. Therefore the quadratic form above is strictly positive whenever $\Delta \neq 0$, and hence

$$\frac{d}{ds} D_t(s) < 0 \qquad \text{for } \Delta \neq 0.$$

This proves that $D_t(\sigma_{\mathrm{rel}}^2)$ is strictly decreasing in σ_{rel}^2. □

Remark. Lemma 1 is a population-level surrogate statement that explains why protecting a sparse set of class-relevant directions is preferable to isotropic corruption when the reverse budget is small. In the full model, those directions are estimated from the guidance signal rather than assumed known.

C Baselines

We compare MAGNET against baselines designed to isolate the roles of validity-preserving geometry, learned transport, and structure-aware conditional recovery. All baselines are evaluated on the same splits, preprocessing pipeline, and sampling budgets as MAGNET, so differences reflect model design rather than data handling or integration depth.

DiffeoCFM. DiffeoCFM [21] is the strongest prior validity-preserving baseline in our setting. It performs conditional flow matching in a global pullback chart of the correlation manifold induced by the normalized-Cholesky parameterization. This baseline tests whether validity-preserving transport alone is sufficient for conditional connectome generation.

DiffeoGauss. DiffeoGauss is a structural ablation of DiffeoCFM in which the learned transport dynamics are replaced by a class-conditional Gaussian model fit directly in the same pullback chart. It therefore preserves the geometry of DiffeoCFM while removing learned flow dynamics. Comparing against DiffeoGauss isolates the contribution of learned transport beyond the validity-preserving coordinate system itself.

TriangCFM. TriangCFM is a Euclidean flow-matching baseline defined directly on the strict lower-triangular coordinates, without a validity-preserving correlation-manifold chart. Generated samples must therefore be mapped back to the correlation manifold through post-hoc projection. This baseline isolates the importance of validity-preserving geometry by retaining continuous transport while removing the pullback construction.

Real Data. The real-data row is not a generative model, but an empirical oracle computed by comparing disjoint real-data splits under the same evaluation pipeline. It quantifies the finite-sample ceiling induced by split-wise estimation and provides a reference point for interpreting generative fidelity and downstream utility.

Why not generic graph generative baselines? We do not include graph generative models such as GDSS [43] or DiGress [82] because they are designed for sparse graphs with discrete edge labels, while our task requires the generation of dense, continuous-valued functional connectomes. In functional connectomes, every ROI pair carries a real-valued dependency, and the output must satisfy symmetry and unit-diagonal structure. These requirements fall outside the design assumptions of such baselines. Adapting them to this setting would require a nontrivial redesign of their representation, denoising objective, and output parameterization, along with additional procedures to force a valid connectome structure. This would make the comparison hard to interpret, since any performance degradation could come either from the underlying model or from the adaptation stack needed to fit a fundamentally different data type. We therefore focus on baselines that are inherently designed for continuous connectome generation under comparable structural assumptions.

To make this limitation concrete, we also implemented an adaptation of GDSS. We trained the official GDSS adjacency score network on dense signed ROI adjacency matrices and applied post-hoc projection to force each generated output back to the correlation-matrix domain. We denote this adapted baseline as GDSS-proj. This is not a native connectome generator: the projection step is required because generic graph diffusion does not enforce positive definiteness or unit-diagonal correlation structure. The results in Tables 1–2 support the design choice behind MAGNET. Although GDSS-proj is fast, it substantially underperforms validity-aware connectome generators on support fidelity and zero-shot subclass recovery, indicating that generic graph diffusion plus post-hoc projection is insufficient for dense functional connectome synthesis.

D Experimental Setup

Evaluation metrics. We assess class-conditional support recovery using α-precision (fidelity to held-out support) and β-recall (coverage of that support), summarizing the trade-off with their harmonic mean, α, β-F1 [2]. Because generative fidelity alone does not guarantee the preservation of task-relevant structures, we also evaluate downstream utility by training a logistic regression classifier on Riemannian tangent-space features extracted from generated samples, reporting ROC-AUC and F1 on real validation connectomes.

Zero-shot subclass covariance metric. To evaluate whether generated samples reproduce the dominant covariance spectrum of resolved clinical subclasses (e.g., in ADNI [83] or OASIS-3 [49]), we report λ-MSE. For each subclass $\ell \in \mathcal{L}$, we collect held-out real connectomes $\mathcal{R}_\ell$ and generated connectomes $\mathcal{G}_\ell$ assigned to ℓ by the real-data-trained resolver. We project each matrix to a valid correlation manifold and vectorize its strictly upper-triangular entries via $\psi(C) = \mathrm{vech}_0(\Pi_{\mathcal{C}_d}(C)) \in \mathbb{R}^p$, where $p = d(d-1)/2$. After mean-centering these features to form X_ℓ and $\widehat{X}_\ell$, we compute their top covariance eigenvalues from singular values $s_{\ell,k}$ and $\widehat{s}_{\ell,k}$ as $\lambda^{\mathrm{real}}_{\ell,k} = s^2_{\ell,k}/(n_\ell - 1)$ and $\lambda^{\mathrm{gen}}_{\ell,k} = \widehat{s}^2_{\ell,k}/(\widehat{n}_\ell - 1)$. For $K_\ell = \min(5, n_\ell - 1, \widehat{n}_\ell - 1, p)$, the metric is defined as $\lambda\text{-MSE} = |\mathcal{L}_{\mathrm{valid}}|^{-1} \sum_{\ell \in \mathcal{L}_{\mathrm{valid}}} K_\ell^{-1} \sum_{k=1}^{K_\ell} (\lambda^{\mathrm{real}}_{\ell,k} - \lambda^{\mathrm{gen}}_{\ell,k})^2$, where $\mathcal{L}_{\mathrm{valid}}$ contains subclasses with ≥ 5 real and generated samples. Unlike CAS macro-F1, this strictly isolates the alignment of subclass-specific covariance spread rather than evaluating classifier accuracy.

Training and sampling protocol. All methods are evaluated across 10 disjoint train/validation splits (10% validation). Experiments were conducted on a single NVIDIA A6000 GPU. Wall-clock sampling times in Table 1 are measured as the mean time to generate one full split's worth of samples, averaged over 10 splits on the same hardware. For ADNI and OASIS-3, train/test disjointness is strictly enforced at the subject level [21]. MAGNET utilizes balanced class sampling, inverse subject-frequency weighting on repeated-subject cohorts, and min-SNR loss reweighting. To isolate model design rather than integration depth, we fix the computational budget: MAGNET samples via a 6-step DDIM chain [75], while DiffeoCFM and its ablations use matched 6-point ODE grids.

E Architectural Details

MAGNET couples the validity-preserving normalized-Cholesky chart with a region-aware denoising backbone. In the MSDL setting, each connectome has 39 regions, giving 741 chart coordinates. Rather than processing the noisy latent state x_t as a flat vector, the denoiser first lifts it into a symmetric ROI-level matrix $A(x_t)$. The chart coordinates are written into the strict lower triangle, reflected across the diagonal, and the diagonal is set to zero. The ith row, denoted $a_i(x_t)$, is then interpreted as the current noisy connectivity profile of region i.

Each region is represented as one token. The initial token combines the row-wise connectivity profile with anatomical identity and global conditioning: $h_i^{(0)} = W_{\mathrm{in}} a_i(x_t) + p_i + c_t$. Here, W_{in} is a learned row projection, p_i is a learned ROI positional embedding, and c_t is a shared conditioning broadcast to all regions. The conditioning vector combines the diffusion timestep embedding, the class-anchored subtype embedding, and the amortized structural bridge embedding. In the implementation, the bridge embedding is passed through a learned gate before being injected, allowing the denoiser to modulate how strongly the global structural summary affects each reverse step.

The token sequence is processed by a pre-norm Transformer encoder with hidden size 128, four layers, four QK-normalized attention heads, GeGLU feed-forward blocks of width 512, and dropout 0.1. Because attention is applied across all ROI tokens, each region updates its representation using information from the full current connectome. This gives the denoiser a relational inductive bias without hard-coding an anatomical edge graph. The learned positional embeddings preserve ROI identity, so the model remains tied to the atlas-defined ordering of brain regions.

After the transformer stack, MAGNET maps the final token matrix $H^{(L)}$ back to chart coordinates. A row-wise output projection produces a square matrix R, which is symmetrized and vectorized over the strict lower triangle: $E = \frac{1}{2}(R + R^\top)$, $\widehat{v}_\theta = \mathrm{vech}_0(E)$. Thus, the denoiser predicts a diffusion target in the same normalized-Cholesky chart used by the sampler. The chart itself, rather than the denoiser output alone, is what guarantees that decoded samples are valid correlation matrices.

The same graph denoiser is used for conditional and unconditional predictions during classifier-free guidance. Let $\widehat{v}_c$ and $\widehat{v}_\varnothing$ denote the conditional and unconditional denoiser outputs, and let $g = \widehat{v}_c - \widehat{v}_\varnothing$ be the guidance direction. MAGNET retains the component of the unconditional prediction that is orthogonal to g: $\widehat{v}_\varnothing^{\perp} = \widehat{v}_\varnothing - \frac{\langle \widehat{v}_\varnothing, g \rangle}{\|g\|_2^2 + \varepsilon} g$. The guided prediction is then $\widehat{v}_\gamma = \widehat{v}_\varnothing^{\perp} + \gamma \tanh(\|g\|_2)\, g$. The $\tanh(\|g\|_2)$ factor limits over-steering when the conditional and unconditional predictions are close, while the orthogonal residual preserves variation not aligned with the class-discriminative direction.

Table 5: **Bridge ablations across ABIDE, ADNI, and OASIS-3.** We compare six structural-conditioning variants: no bridge, self-conditioned bridge, teacher bridge at inference, amortized bridge without L_{bridge}, amortized bridge with L_{bridge}, and the full MAGNET objective. We report α, β-F1, ROC-AUC, F1, bridge distillation error where applicable, and sampling time. Lower bridge error tracks stronger structural fidelity, while removing L_{bridge} causes severe bridge-tracking failure. Gray rows denote the real-data reference. Bold highlights the best generative model for each metric.

Dataset	Method	Quality	CAS Metrics		Bridge Distillation Error ↓	Time (s.)
		α, β-F1 ↑	ROC-AUC ↑	F1 ↑		Sampling ↓
ABIDE	Real Data	0.80 ± 0.01	0.66 ± 0.04	0.58 ± 0.06	N/A	N/A
	No bridge	0.58 ± 0.05	0.66 ± 0.06	0.57 ± 0.09	-	**0.10 ± 0.00**
	Self-conditioning	0.51 ± 0.06	0.66 ± 0.06	0.56 ± 0.11	-	3.01 ± 0.01
	Teacher bridge	0.64 ± 0.06	0.66 ± 0.06	0.57 ± 0.09	-	3.00 ± 0.01
	Amortized bridge (L_{bridge})	0.65 ± 0.05	0.66 ± 0.05	0.57 ± 0.08	0.48 ± 0.04	0.12 ± 0.00
	Amortized bridge w/o bridge loss	0.59 ± 0.07	0.66 ± 0.06	0.56 ± 0.09	14.81 ± 0.53	0.13 ± 0.00
	Proposed ($L_{bridge} + L_{spec}$)	**0.68 ± 0.04**	**0.67 ± 0.06**	**0.58 ± 0.09**	**0.46 ± 0.04**	0.12 ± 0.00
ADNI	Real Data	0.87 ± 0.05	0.58 ± 0.03	0.59 ± 0.03	N/A	N/A
	No bridge	0.59 ± 0.12	0.58 ± 0.06	0.66 ± 0.05	-	**0.22 ± 0.00**
	Self-conditioning	0.41 ± 0.15	0.58 ± 0.04	0.55 ± 0.22	-	8.99 ± 0.04
	Teacher bridge	0.67 ± 0.11	0.58 ± 0.05	0.65 ± 0.05	-	8.09 ± 0.08
	Amortized bridge (L_{bridge})	0.68 ± 0.12	0.58 ± 0.06	0.65 ± 0.06	0.05 ± 0.00	0.25 ± 0.00
	Amortized bridge w/o bridge loss	0.48 ± 0.11	0.58 ± 0.05	**0.67 ± 0.05**	14.67 ± 1.16	0.25 ± 0.00
	Proposed ($L_{bridge} + L_{spec}$)	**0.74 ± 0.08**	**0.58 ± 0.07**	**0.67 ± 0.06**	**0.05 ± 0.00**	0.25 ± 0.00
OASIS-3	Real Data	0.88 ± 0.01	0.71 ± 0.04	0.61 ± 0.06	N/A	N/A
	No bridge	0.49 ± 0.07	0.69 ± 0.05	0.58 ± 0.05	-	**0.18 ± 0.00**
	Self-conditioning	0.46 ± 0.06	0.69 ± 0.04	0.57 ± 0.04	-	6.38 ± 0.05
	Teacher bridge	0.54 ± 0.06	0.69 ± 0.04	**0.58 ± 0.04**	-	6.28 ± 0.02
	Amortized bridge (L_{bridge})	0.55 ± 0.05	0.69 ± 0.05	0.57 ± 0.05	0.09 ± 0.00	0.20 ± 0.00
	Amortized bridge w/o bridge loss	0.49 ± 0.05	**0.69 ± 0.04**	0.57 ± 0.05	14.90 ± 1.06	0.21 ± 0.00
	Proposed ($L_{bridge} + L_{spec}$)	**0.57 ± 0.05**	0.68 ± 0.04	**0.58 ± 0.05**	**0.08 ± 0.00**	0.20 ± 0.00

Finally, MAGNET applies class-wise coordinate-affine calibration to the final sampled latent before decoding. This calibration is fitted on the training split by matching class-conditional latent means and variances between uncalibrated generated samples and real training latents. The calibrated latent is clipped and then decoded through the normalized-Cholesky inverse map, so validity is still enforced by the chart-based decoder.

F Framework Analysis

F.1 Amortized Structural Bridge

To evaluate the role of the amortized structural bridge, we compare six variants: no bridge, self-conditioned bridge, teacher bridge at inference, amortized bridge without L_{bridge}, amortized bridge with L_{bridge}, and the full MAGNET objective. For bridge-enabled variants, we report the Bridge Distillation Error $\|B_\psi(x_t, t, s) - b^\star(x_0, s, t)\|_2^2$, which measures how well the amortized bridge tracks the time-dependent teacher target.

Table 5 shows a clear bridge-quality versus structural-fidelity pattern. Lower distillation error is consistently associated with stronger α, β-F1, while removing L_{bridge} causes a severe bridge-tracking failure without yielding a corresponding fidelity gain. This is the regime the bridge is designed for, and the ablation matches the Remark of proposition 2: the amortized bridge is useful to the extent that it tracks the teacher target well.

The efficiency gap is equally important. Teacher-bridge and self-conditioned alternatives are substantially slower because they require a second structural pass at inference, whereas the amortized bridge preserves the few-step sampling regime. In practice, the bridge-enabled MAGNET variants retain roughly $30 - 60\times$ lower sampling cost than the teacher and self-conditioned alternatives while delivering the strongest overall fidelity profile. Taken together, these results justify the amortized bridge as the primary structural-conditioning mechanism in MAGNET. It provides the gains of explicit structural guidance without sacrificing the speed advantages of few-step generation.

Table 6: **Relevance-preserving corruption ablation across ABIDE, ADNI, and OASIS-3.** We compare standard isotropic diffusion against three temporal variations of our adaptive protection schedule (Low, Adaptive, and High density decay). Performance remains stable across all adaptive regimes, demonstrating that structural recovery is invariant to the exact decay curve provided early-stage topological protection is active.

Dataset	Configuration	Decay (p)	Quality	CAS Metrics		Sampling Time (s) ↓
			α, β-F1 ↑	ROC-AUC ↑	F1 ↑	
ABIDE	Isotropic (Baseline)	N/A	0.62 ± 0.03	0.60 ± 0.07	0.50 ± 0.09	0.36 ± 0.00
	Low Density	1	0.68 ± 0.03	0.64 ± 0.07	0.55 ± 0.06	0.44 ± 0.12
	Adaptive (Ours)	**3**	$\mathbf{0.68 \pm 0.03}$	$\mathbf{0.64 \pm 0.07}$	$\mathbf{0.56 \pm 0.06}$	$\mathbf{0.38 \pm 0.02}$
	High Density	5	0.69 ± 0.03	0.64 ± 0.07	0.55 ± 0.05	0.39 ± 0.03
ADNI	Isotropic (Baseline)	N/A	0.69 ± 0.09	0.52 ± 0.07	0.55 ± 0.05	0.73 ± 0.01
	Low Density	1	0.75 ± 0.07	0.57 ± 0.06	0.61 ± 0.06	0.76 ± 0.02
	Adaptive (Ours)	**3**	$\mathbf{0.76 \pm 0.06}$	$\mathbf{0.57 \pm 0.07}$	$\mathbf{0.61 \pm 0.05}$	$\mathbf{0.86 \pm 0.28}$
	High Density	5	0.76 ± 0.06	0.56 ± 0.06	0.61 ± 0.05	0.76 ± 0.02
OASIS-3	Isotropic (Baseline)	N/A	0.53 ± 0.06	0.64 ± 0.04	0.49 ± 0.05	0.60 ± 0.00
	Low Density	1	0.58 ± 0.06	0.68 ± 0.03	0.55 ± 0.04	0.63 ± 0.01
	Adaptive (Ours)	**3**	$\mathbf{0.59 \pm 0.06}$	$\mathbf{0.68 \pm 0.03}$	$\mathbf{0.56 \pm 0.04}$	$\mathbf{0.63 \pm 0.00}$
	High Density	5	0.57 ± 0.06	0.68 ± 0.03	0.56 ± 0.04	0.63 ± 0.01

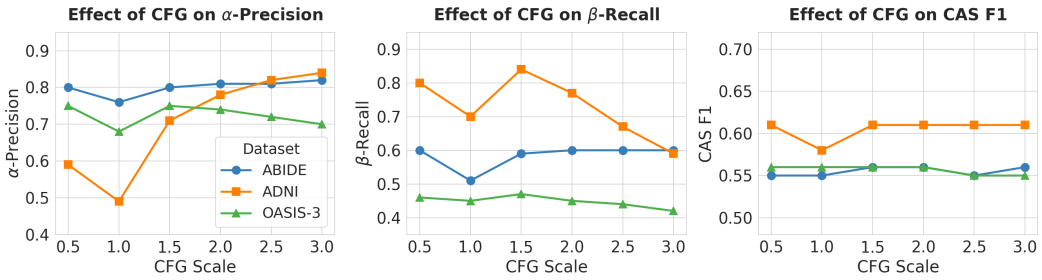

Figure 5: **Classifier-free guidance scale ablation for MAGNET.** We vary the guidance scale γ from 0.5 to 3.0 and plot class-conditional fidelity (α-Precision, β-Recall) together with downstream utility (CAS F1) on ABIDE, ADNI, and OASIS-3. Across datasets, performance is strongest in the moderate-guidance regime, with $\gamma = 1.5$ giving the most stable fidelity–utility trade-off. Larger guidance scales tend to increase precision at the expense of recall, consistent with overspecialization to a narrower region of the class manifold.

F.2 Relevance-Preserving Corruption Ablation

Our relevance-preserving corruption relies on a monotonically decreasing protection schedule, $\rho(t)$, to shield condition-defining coordinates during early forward diffusion. To evaluate the framework's sensitivity to this mechanism, we parameterize the decay curve as $\rho(t) = (1 - \tau_t)^p$, where $\tau_t \in [0, 1]$ is the normalized continuous diffusion time (i.e., $\tau_t = t/T$). The hyperparameter p strictly controls the temporal density of the protection, dictating how quickly structural coordinates yield to Gaussian noise, rather than the total magnitude of protection applied.

We ablate this schedule across four distinct regimes: **(i) Isotropic (Baseline):** No adaptive manifold preservation; **(ii) Low Density ($p = 1$):** Gradual, linear relaxation of constraints; **(iii) Adaptive ($p = 3$, Ours):** Our primary, balanced decay schedule; and **(iv) High Density ($p = 5$):** Aggressive, early release of structural constraints.

Table 6 highlights a clear trend across ABIDE, ADNI, and OASIS-3: MAGNET is robust to the exact shape of the decay curve. Generative fidelity (α, β-F1) and downstream utility (CAS) remain consistently stable across the Low, Adaptive, and High density regimes. The adaptive schedule provides consistent improvement over isotropic diffusion across all three cohorts. Crucially, the robustness to the exact decay shape ($p = 1, 3, 5$) confirms that the mechanism is not sensitive to hyperparameter tuning — any protection schedule that shields condition-relevant coordinates during the high-noise phase suffices. The primary contribution of relevance-preserving corruption is providing a cleaner latent trajectory for the bridge and guidance components to operate on, as evidenced by the full-system results in Table 3.

F.3 Classifier-Free Guidance Scale

The classifier-free guidance scale γ controls how strongly reverse-time sampling is steered toward the conditional branch. In MAGNET, guidance is not implemented as plain linear extrapolation. Instead,

the sampler scales the conditional steering direction while retaining an orthogonal residual from the unconditional prediction. Figure 5 shows that performance is strongest in the moderate-guidance regime.

On ADNI, $\gamma = 1.5$ gives the best overall fidelity–utility trade-off, improving clearly over both weak guidance ($\gamma = 0.5$) and overly strong guidance ($\gamma \geq 2.5$). The main failure mode at high guidance is not precision but coverage: α-precision increases, but β-recall drops sharply, indicating overspecialization to a narrower subset of the class manifold. OASIS-3 follows the same pattern, though less sharply, with $\gamma = 1.5$ again near-optimal. ABIDE is less sensitive once guidance is enabled, consistent with its simpler binary structure.

A useful implementation detail is that $\gamma = 1.0$ in Fig. 5 is implemented as a discrete branch skip (the unconditional forward pass is omitted entirely), rather than as a continuous evaluation of the guidance formula at that scale. Consequently, the transition from $\gamma = 1.0$ to $\gamma = 0.5$ represents a qualitative mode change, not a smooth interpolation, which accounts for the performance gap observed at that boundary. This explains both its lower runtime and its weaker fidelity relative to the moderately guided regime. We therefore use $\gamma = 1.5$ throughout the paper.

F.4 Class-Anchored Conditioning

To evaluate the necessity of Class-Anchored Conditioning, we quantify using the complexity of the datasets. The heterogeneity entropy $\mathbf{H}$ of a dataset is the Shannon Entropy of the discovered structural subtypes S conditioned on coarse labels C as $\mathbf{H} = -\sum_{c \in \mathcal{C}} P(c) \sum_{k=1}^{K_c} P(s_k|c) \log P(s_k|c)$ where s_k represents the latent sub-manifolds within a given class c. As $\mathbf{H}$ increases, the variance in a single label expands, leading to denoising ambiguity.

We ablate against three paradigms (Fig. 6) to isolate performance gains. **Class-only** utilizes the class coarse labels, **Random Subtypes** reassigns tokens randomly within a class to measure if gains are a regularizing artifact of the noise injection, and **Global Clustering** to discover M latent structural clusters across the dataset. On highly heterogeneous cohort like ADNI and OASIS-3, the global clustering fails to differentiate between structurally similar but clinically distinct states. Our proposed Class-Anchored Conditioning enforces a stricter hierarchy to find K distinct clusters only within the bounds of a specific clinical label ensuring the diffusion process is guided by latent substructures, not violating the broader class condition. The performance gain in generative fidelity confirms high-quality connectome generation in complex clinical datasets requires sub-manifold conditioning that remains strictly anchored to the coarse class constraint.

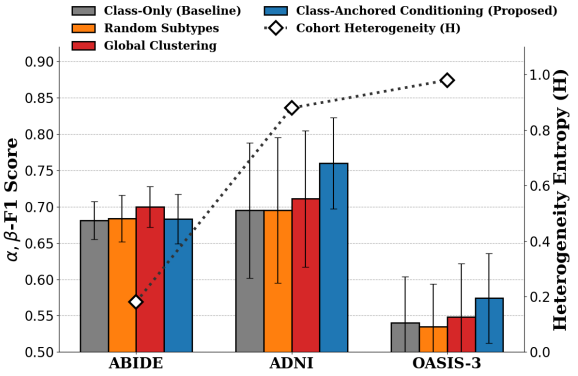

Figure 6: **Efficacy of Class-Anchored Conditioning** against class-only, random-subtype, and global clustering baselines. The performance gain of Class-Anchored Conditioning strictly scales with structural heterogeneity $\mathbf{H}$.

F.5 Self-conditioning probability

We also tested the fallback self-conditioning probability p_{sc}, which in earlier non-amortized variants controls how often a first-pass denoiser estimate is reused as auxiliary structural context during training. In the final MAGNET configuration, however, structural context is supplied primarily by the amortized bridge, so this fallback mechanism is no longer central to the active model.

Empirically, varying p_{sc} from 0.2 to 0.8 produces no meaningful change in generative fidelity or downstream performance across ABIDE, ADNI, and OASIS-3: all three datasets remain effectively invariant at the precision reported in the main benchmark, with only minor variation in runtime. We therefore treat p_{sc} as a secondary implementation parameter and use $p_{\text{sc}} = 0.5$ throughout the paper. This confirms that, once the amortized bridge is active, structural recovery is largely insensitive to the scheduling of the fallback self-conditioning branch.

F.6 Training and Sampling Efficiency

MAGNET trades additional training cost for substantially faster repeated sampling. On a single NVIDIA A6000 GPU, training requires 532.08 ± 16.66 s per split on ABIDE, 1593.63 ± 27.99 s on ADNI, and 1242.53 ± 18.49 s on OASIS-3, corresponding to approximately 9–27 minutes per trained split. This is roughly an order of magnitude more training time than the DiffeoCFM and TriangCFM baselines under the same hardware setting.

The inference regime is substantially different. MAGNET uses a six-step DDIM sampler and requires 0.38 s, 0.86 s, and 0.63 s per generated split on ABIDE, ADNI, and OASIS-3, respectively, compared with 9.04 s, 36.08 s, and 23.24 s for DiffeoCFM (Table 1). This corresponds to approximately $24\times$, $42\times$, and $37\times$ faster sampling, respectively. The additional training cost is therefore incurred once when fitting a checkpoint, whereas the sampling advantage is retained whenever that checkpoint is subsequently used for synthetic cohort generation.

Importantly, this efficiency does not arise from sacrificing conditional fidelity: MAGNET simultaneously improves α, β-F1 by 7–21% over the strongest validity-preserving baseline across the primary cohorts and achieves the lowest zero-shot subclass covariance error in Table 2. These results characterize the computational trade-off of MAGNET as higher one-time training cost in exchange for improved support recovery and substantially lower repeated inference cost.

F.7 Sensitivity to λ_{bridge} and λ_{spec}

The full MAGNET objective combines the diffusion loss with bridge supervision and structural regularization, $L = L_{\text{diff}} + \lambda_{\text{bridge}} L_{\text{bridge}} + \lambda_{\text{spec}} L_{\text{spec}}$. We therefore evaluate whether the gains of the proposed framework depend on narrowly tuned auxiliary-loss weights. For each loss, we vary its coefficient over $\{0, 0.025, 0.05, 0.1, 0.2, 0.4\}$ while holding the other coefficient fixed at its default value of 0.1. The nonzero sweep spans a $16\times$ range around the default operating regime. All remaining training, sampling, and evaluation settings are unchanged.

Table 7 reveals two complementary properties. First, both auxiliary objectives make a material contribution to support recovery. Removing the spectral term decreases α, β-F1 from 0.68 to 0.51 on ABIDE, from 0.76 to 0.68 on ADNI, and from 0.57 to 0.49 on OASIS-3, corresponding to relative reductions of approximately 11–25%. Removing bridge supervision produces similarly consistent degradation, reducing α, β-F1 to 0.53, 0.65, and 0.42, respectively. These ablations are consistent with the intended roles of the two losses: L_{bridge} constrains the amortized predictor to track its structural target, while L_{spec} encourages the denoised latent to preserve global structural statistics.

Second, once either loss is enabled, performance is insensitive to its precise weighting. Across the nonzero λ_{spec} sweep, α, β-F1 varies by at most 0.02 on each cohort; across the nonzero λ_{bridge} sweep, the maximum variation is 0.03. These changes are smaller than, or comparable to, the split-to-split variability reported in the main benchmark. ROC-AUC and downstream F1 exhibit the same qualitative stability. This behavior is also consistent with the construction of the objective: L_{bridge} is restricted to conditioned examples and weighted by $\bar{\alpha}_t$, while L_{spec} is mean-reduced over the minibatch.

Thus, the auxiliary objectives are important through their presence rather than through a narrowly selected coefficient. The default choice $\lambda_{\text{bridge}} = \lambda_{\text{spec}} = 0.1$ lies within a broad stable region across all three cohorts, indicating that MAGNET's performance is not dependent on fine-grained tuning of either loss weight.

G Geometric Analysis of Conditional Connectome Generation

In this section, we analyze whether the generated connectomes are in the same class-conditional geometric space through a series of experiments to compare against the held-out real validation data. In order to keep *occupancy, overlap, and support recovery* non self-referential, we fix all the coordinate system, support contours, and local scales using the real data and ensure valid comparison between Q_m^y and P_{val}^y.

Table 7: **Sensitivity to bridge-supervision and spectral-regularization weights.** We vary λ_{spec} and λ_{bridge} independently while fixing the other coefficient at its default value of 0.1. All remaining training and evaluation settings are unchanged. Results are mean $\pm$ standard deviation across the same splits used in the main benchmark. Setting either coefficient to zero measures removal of the corresponding objective, while the nonzero values test sensitivity over a $16\times$ range.

Dataset	λ_{spec} sensitivity				λ_{bridge} sensitivity			
	λ_{spec}	α, β-F1 ↑	ROC-AUC ↑	F1 ↑	λ_{bridge}	α, β-F1 ↑	ROC-AUC ↑	F1 ↑
ABIDE	0	0.51 ± 0.04	0.52 ± 0.06	0.43 ± 0.01	0	0.53 ± 0.03	0.56 ± 0.06	0.41 ± 0.09
	0.025	0.66 ± 0.06	0.62 ± 0.09	0.55 ± 0.03	0.025	0.66 ± 0.05	0.63 ± 0.05	0.55 ± 0.06
	0.05	0.67 ± 0.01	0.62 ± 0.06	0.55 ± 0.09	0.05	0.66 ± 0.03	0.64 ± 0.08	0.55 ± 0.05
	0.1 (default)	0.68 ± 0.03	0.64 ± 0.07	0.56 ± 0.06	0.1 (default)	0.68 ± 0.03	0.64 ± 0.07	0.56 ± 0.06
	0.2	0.67 ± 0.02	0.63 ± 0.02	0.56 ± 0.12	0.2	0.66 ± 0.03	0.64 ± 0.02	0.54 ± 0.01
	0.4	0.68 ± 0.07	0.64 ± 0.01	0.55 ± 0.07	0.4	0.66 ± 0.04	0.65 ± 0.01	0.55 ± 0.04
ADNI	0	0.68 ± 0.03	0.49 ± 0.01	0.53 ± 0.08	0	0.65 ± 0.06	0.45 ± 0.03	0.47 ± 0.07
	0.025	0.75 ± 0.09	0.57 ± 0.03	0.61 ± 0.09	0.025	0.73 ± 0.01	0.56 ± 0.01	0.61 ± 0.03
	0.05	0.75 ± 0.02	0.55 ± 0.08	0.61 ± 0.03	0.05	0.75 ± 0.09	0.57 ± 0.02	0.60 ± 0.02
	0.1 (default)	0.76 ± 0.06	0.57 ± 0.07	0.61 ± 0.05	0.1 (default)	0.76 ± 0.06	0.57 ± 0.07	0.61 ± 0.05
	0.2	0.75 ± 0.03	0.56 ± 0.01	0.60 ± 0.08	0.2	0.75 ± 0.04	0.57 ± 0.06	0.60 ± 0.09
	0.4	0.76 ± 0.09	0.56 ± 0.03	0.60 ± 0.01	0.4	0.74 ± 0.02	0.55 ± 0.04	0.59 ± 0.04
OASIS-3	0	0.49 ± 0.09	0.59 ± 0.07	0.47 ± 0.01	0	0.42 ± 0.01	0.53 ± 0.04	0.42 ± 0.06
	0.025	0.57 ± 0.08	0.67 ± 0.05	0.56 ± 0.09	0.025	0.55 ± 0.04	0.68 ± 0.08	0.56 ± 0.03
	0.05	0.56 ± 0.06	0.68 ± 0.08	0.55 ± 0.04	0.05	0.55 ± 0.09	0.65 ± 0.06	0.57 ± 0.01
	0.1 (default)	0.57 ± 0.06	0.68 ± 0.03	0.56 ± 0.04	0.1 (default)	0.57 ± 0.06	0.68 ± 0.03	0.56 ± 0.04
	0.2	0.57 ± 0.07	0.67 ± 0.01	0.55 ± 0.08	0.2	0.56 ± 0.03	0.67 ± 0.09	0.56 ± 0.09
	0.4	0.58 ± 0.01	0.65 ± 0.09	0.54 ± 0.07	0.4	0.57 ± 0.06	0.66 ± 0.01	0.55 ± 0.08

G.1 Visualization Setup

Let $\mathcal{C}_d = \left\{C \in \mathbb{R}^{d\times d} \,:\, C = C^\top,\ C \succ 0,\ \operatorname{diag}(C) = \mathbf{1}\right\}$ denote the symmetric positive definite correlation matrix manifold on the MSDL atlas [80]. Let P^y_{val} denote the held-out real validation distribution and Q^y_m denote the class-conditional distribution generated by method m where $y \in \{0, 1\}$ is the dataset specific class label. We compare $Q_m(\cdot \mid Y = y)$ against $P_{\text{val}}(\cdot \mid Y = y)$ to make class-conditional comparisons. This is essential because *a conditional generator can appear globally plausible while still failing to reproduce the support or geometry of a particular class* [8, 67].

G.2 Geometry, Assumptions, and why this analysis

Log-Euclidean representation. Since $\mathcal{C}_\lceil$ is not a vector space, euclidean operations in matrix space do not capture the geometry of connectomes faithfully. Standard operations on $C_1, C_2 \in \mathcal{C}_\lceil$ treat the objects as points in a flat vector space, ignoring the intrinsic geometric structure of the set. Hence, we adopt the log-euclidean representation [4, 5] and define $\Phi : \mathcal{C}_\lceil \to \operatorname{Sym}(d), \quad \Phi(C) = \log C$ where $\operatorname{Sym}(d)$ is the vector space of real symmetric matrices and $\log C$ is the matrix logarithm. The log-euclidean distance is

$$d_{\text{LE}}(C_1, C_2) = \|\Phi(C_1) - \Phi(C_2)\|_F. \tag{14}$$

Moreover, every $C \in \mathcal{C}_d$ has strictly positive eigenvalues and $\log C$ is well defined via the spectral decomposition. If, $C = U \operatorname{diag}(\lambda_1, \ldots, \lambda_d) U^\top$ with $\lambda_i > 0$, then $\log C = U \operatorname{diag}(\log \lambda_1, \ldots, \log \lambda_d) U^\top$. Because $\log C$ is symmetric, the image of Φ lies in $\operatorname{Sym}(d)$. Thus, instead of comparing the connectomes in euclidean matrix space, we first map them to $\operatorname{Sym}(d)$ and then utilize euclidean geometry via the Frobenius norm.

Assumptions. All the visual analysis is based on the following assumptions.

- **A1. SPD Validity.** The basic structural assumption required by the log-euclidean model is that all connectomes lie in $\mathcal{C}_d$, making $\Phi(C)$ well defined.
- **A2. Class-conditional evaluation.** We evaluate Q^y_m against P^y_{val} for each class y. Enforcing this allows us to differentiate between models that mixes structures while still appearing globally close to the pooled real distribution.
- **A3. Real-derived reference geometry.** All references for evaluations are derived from the real data only to prevent the reference from adapting to the generated samples.

A4. Local regularity of held-out support. The density of $\Phi_{\#}P^y_{\text{val}}$ is assumed to be locally bounded away from 0 and ∞ to support interior of P_{val}. We adapted the regime from [23, 28] where nearest-neighbor radii consistently estimate local spacing.

Why these assumptions enter the analysis. **A1, A2, and A3** ensure that every visual object remains fixed across methods and is well defined in a valid geometry. **A4** is the standard condition to estimate negative local spacing in the nearest neighbor radii. Each assumption allows to interpret the visualization in regards to the class-conditional distributional agreement in addition to summary statistics.

G.3 Global geometric fidelity through radial calibration

We want to identify if the generated samples have the correct global geometric understanding as per the held-out real center. For class y, the log-domain mean is $\mu^y_{\log,\text{val}} = \mathbb{E}_{C\sim P^y_{\text{val}}}[\Phi(C)]$, and the corresponding log-Euclidean Fréchet mean is $\bar{C}^y_{\text{LE,val}} = \exp(\mu^y_{\log,\text{val}})$. The radial coordinate of a connectome is

$$r_y(C) = d_{\text{LE}}(C, \bar{C}^y_{\text{LE,val}}) = \|\Phi(C) - \mu^y_{\log,\text{val}}\|_F. \tag{15}$$

Utilizing Eq. 15, we map the full manifold-valued distribution to a one-dimensional law to summarize the mass as a function of distance from the held-out real center. This is a push-forward construction, $\nu^y_{\text{val}} = (r_y)_{\#}P^y_{\text{val}},\ \nu^y_m = (r_y)_{\#}Q^y_m$ where $(r_y)_{\#}\mu$ denotes the pushforward of the distribution μ under the map r_y. This allows us to sample a matrix from the distribution, compute its radial distance $r_y(C)$, and study the resulting distribution on $\mathbb{R}_{\geq 0}$. The radial curves in Fig. 7 highlight these pushforward laws. To account for radial mismatch, we use the 1-Wasserstein distance [46]:

$$W^{\text{rad}}_{m,y} = W_1\big(\nu^m_y, \nu^{\text{val}}_y\big), \tag{16}$$

and the log-quantile calibration error

$$E^{\text{rad}}_{m,y} = \frac{1}{|Q|}\sum_{q\in Q}\left|\log\frac{Q_{\nu^m_y}(q)+\varepsilon}{Q_{\nu^{\text{val}}_y}(q)+\varepsilon}\right|, \tag{17}$$

where $Q \subset (0,1)$ is a fixed quantile grid and $Q_\nu(q)$ denotes the quantile function of a distribution ν.

Proposition 3. $\bar{C}^{\text{LE,val}}_y = \exp(\mu^{\log,\text{val}}_y)$ *minimizes the expected squared log-Euclidean distance to the held-out real distribution:*

$$\bar{C}^{\text{LE,val}}_y \in \arg\min_{C\in\mathcal{C}_d} \mathbb{E}_{X\sim P^{\text{val}}_y}\big[d_{\text{LE}}(X,C)^2\big].$$

Proof. For any fixed $C \in \mathcal{C}_d$, by Eq. 14, $d_{\text{LE}}(X,C)^2 = \|\log X - \log C\|^2_F$. Therefore the optimization problem is

$$\min_{C\in\mathcal{C}_d} \mathbb{E}\big[\|\log X - \log C\|^2_F\big].$$

Now let $M = \log C$. Because every $C \in \mathcal{C}_d$ is SPD, the matrix logarithm $\log C$ is well defined and maps $\mathcal{C}_d$ injectively into $\text{Sym}(d)$; the image is a proper submanifold of $\text{Sym}(d)$. We relax the constraint and minimize over the ambient space $\text{Sym}(d)$, so the unconstrained minimizer $\bar{C}^{\text{LE,val}}_y = \exp(\mu^{\log,\text{val}}_y)$ serves as a reference point in the broader SPD manifold rather than a minimizer strictly within $\mathcal{C}_d$. Hence the problem becomes: $\min_{M\in\text{Sym}(d)} \mathbb{E}\big[\|\log X - M\|^2_F\big]$.

Eq. 16 measures the minimal transport cost required to transform the generated radial space into the held-out real radial space. Eq. 17 provides a scale-sensitive comparison of the two distributions across their entire support.

Let $Z = \log X$. Then Z is a random element of the Euclidean vector space $\text{Sym}(d)$, and the objective becomes $f(M) = \mathbb{E}\big[\|Z - M\|^2_F\big]$. We now expand this expression. Recall that for any matrices A, B, $\|A-B\|^2_F = \langle A-B, A-B\rangle_F$, where the Frobenius inner product is $\langle A,B\rangle_F = \text{tr}(A^\top B)$. Since all matrices here are symmetric, this is just the standard Euclidean inner product on $\text{Sym}(d)$.

Expanding the square gives $\|Z - M\|_F^2 = \|Z\|_F^2 - 2\langle Z, M\rangle_F + \|M\|_F^2$. Taking expectation on both sides, $f(M) = \mathbb{E}[\|Z\|_F^2] - 2\langle \mathbb{E}[Z], M\rangle_F + \|M\|_F^2$. The first term, $\mathbb{E}[\|Z\|_F^2]$, does not depend on M, so it is a constant for the optimization. Thus minimizing $f(M)$ is equivalent to minimizing

$$g(M) = \|M\|_F^2 - 2\langle \mathbb{E}[Z], M\rangle_F.$$

Now we explain why this objective is convex, and in fact *strictly convex*. The term $\|M\|_F^2$ is a strictly convex quadratic function of M, because for any distinct $M_1, M_2 \in \mathrm{Sym}(d)$ and any $t \in (0, 1)$,

$$\|tM_1 + (1-t)M_2\|_F^2 < t\|M_1\|_F^2 + (1-t)\|M_2\|_F^2.$$

The linear term $-2\langle \mathbb{E}[Z], M\rangle_F$ is both convex and concave, since linear functions do not affect strict convexity. Therefore $g(M)$, and hence $f(M)$, is strictly convex. A strictly convex function has at most one minimizer.

To find the minimizer explicitly, we complete the square:

$$\|M\|_F^2 - 2\langle \mathbb{E}[Z], M\rangle_F = \|M - \mathbb{E}[Z]\|_F^2 - \|\mathbb{E}[Z]\|_F^2.$$

Therefore $f(M) = \mathbb{E}[\|Z\|_F^2] - \|\mathbb{E}[Z]\|_F^2 + \|M - \mathbb{E}[Z]\|_F^2$. Again, the first two terms are constants with respect to M. So $f(M)$ is minimized exactly when $\|M - \mathbb{E}[Z]\|_F^2 = 0$, that is, $M = \mathbb{E}[Z]$. Recalling that $Z = \log X$, we obtain $M = \mathbb{E}[\log X] = \mu_y^{\log,\mathrm{val}}$. Since $M = \log C$, this gives $C = \exp(\mu_y^{\log,\mathrm{val}}) = \bar{C}_y^{\mathrm{LE,val}}$. By strict convexity, this minimizer is unique. □

Proposition 4. *If the generated class-conditional distribution matches the held-out real distribution, i.e., $Q_y^m = P_y^{\mathrm{val}}$, then the induced radial laws are identical: $\nu_y^m = \nu_y^{\mathrm{val}}$. Consequently, $W_{m,y}^{\mathrm{rad}} = 0$ and $E_{m,y}^{\mathrm{rad}} = 0$.*

Proof. Recall Eq. 14, the matrix logarithm is continuous on $\mathcal{C}_d$, subtraction by a fixed matrix is continuous, and the Frobenius norm is also continuous. Therefore r_y is continuous, hence measurable. By definition of pushforward, for any measurable set $A \subseteq \mathbb{R}_{\geq 0}$, $\nu_y^m(A) = Q_y^m\big(r_y^{-1}(A)\big)$, $\quad \nu_y^{\mathrm{val}}(A) = P_y^{\mathrm{val}}\big(r_y^{-1}(A)\big)$. If $Q_y^m = P_y^{\mathrm{val}}$ as probability measures on $\mathcal{C}_d$, then for every measurable A, $Q_y^m\big(r_y^{-1}(A)\big) = P_y^{\mathrm{val}}\big(r_y^{-1}(A)\big)$. Hence $\nu_y^m(A) = \nu_y^{\mathrm{val}}(A)$, which implies $\nu_y^m = \nu_y^{\mathrm{val}}$.

If two probability distributions are identical, then any proper distance or discrepancy between them must vanish. The 1-Wasserstein distance $W_1(\nu_y^m, \nu_y^{\mathrm{val}})$ is zero if and only if the two measures are identical. The quantile functions also coincide, i.e., $Q_{\nu_y^m}(q) = Q_{\nu_y^{\mathrm{val}}}(q) \quad$ for all $q \in (0, 1)$, which implies every term in Eq. 17 is zero. Therefore, $W_{m,y}^{\mathrm{rad}} = 0$ and $E_{m,y}^{\mathrm{rad}} = 0$. □

Remark. Proposition 4 establishes a one-directional necessity: matching radial laws is required, but fundamentally insufficient, for full distribution agreement. Intuitively, the radial mapping r_y acts as an isotropic aggregator, collapsing the $\frac{d(d-1)}{2}$-dimensional tangent space of $\mathrm{Sym}(d)$ into a single non-negative scalar. Because this operation strips away all angular information, two distributions can easily share identical radial profiles while occupying completely disjoint directional regions. Geometrically, this is analogous to comparing a uniform spherical shell to a narrow, dense equatorial ring; both share the same radial distance from the center, yet their structures differ entirely. In the context of connectomes, this implies that a generator might perfectly replicate the correct *magnitude* of overall network deviation from the mean class state, but allocate those topological changes to entirely incorrect functional edges or subnetworks. Therefore, while vanishing radial discrepancy ($W_{m,y}^{\mathrm{rad}} = 0$) provides a rigorous, necessary confirmation that global variance is calibrated, it must be paired with directional or topological evaluations to guarantee true conditional support recovery.

Empirical Takeaway. Figure 7 visualizes the full radial pushforward distributions and their Wasserstein distances to the real held-out center. Table 8 reports the corresponding *median* log-Euclidean radial distances for MAGNET, DiffeoCFM, and DiffeoGauss. TriangCFM is excluded from the table because its severe overdispersion (median radial distances exceeding 40 units, visible as inset annotations in Fig. 7) places it far outside the axis range of the remaining methods, making a shared table misleading. Across all three datasets, MAGNET achieves the smallest deviation from the real center, confirming tighter global geometric alignment. If a generator matches the true distribution, then this radial law must match exactly, and both Wasserstein and quantile discrepancies vanish. However, radial agreement alone does not guarantee full geometric fidelity.

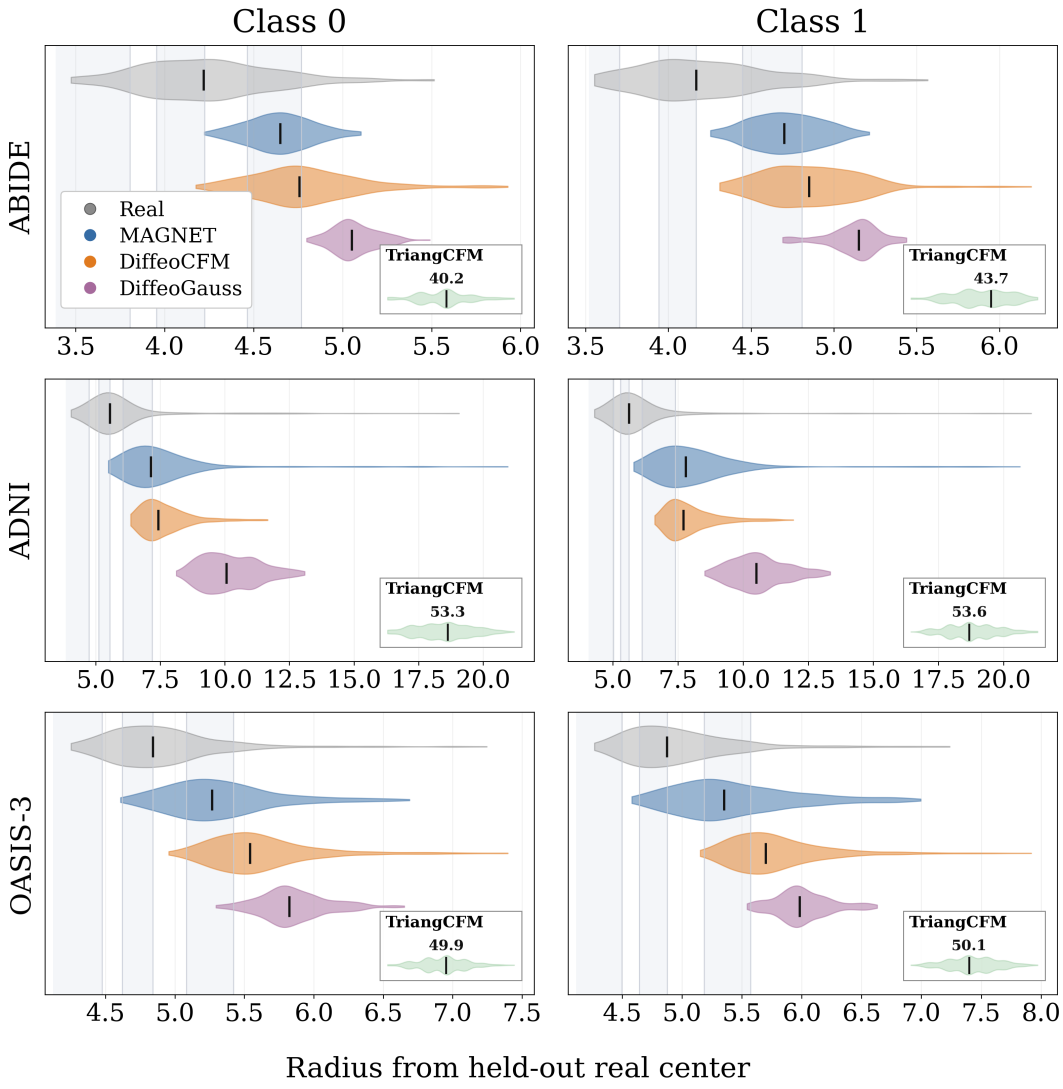

Figure 7: **Global radial calibration in log-Euclidean geometry.** We plot the distribution of distances from the held-out real data center for each class and dataset. A systematic left shift relative to Real *(grey)* indicates manifold contraction, while a right shift indicates overdispersion. MAGNET *(blue)* consistently demonstrates the closest alignment to the real distribution across all datasets, while baselines like TriangCFM *(insets)* exhibit significant overdispersion.

G.4 Local Support Occupancy via Tangent Charts

While radial analysis verifies that the generated samples exhibit the correct global scale around the real class center (Appendix G.3), it does not reveal whether the generated probability mass correctly aligns with the true support of the data distribution. To investigate this local occupancy, we construct a low-dimensional tangent chart using held-out real data and evaluate how the generated samples project onto this localized coordinate space.

Log-domain centering and vectorization. For each held-out real connectome $C_i^{\mathrm{val},y}$, we define

$$z_i^{\mathrm{val},y} = \mathrm{vech}_0\big(\log C_i^{\mathrm{val},y} - \mu_y^{\log,\mathrm{val}}\big) \in \mathbb{R}^p, \quad p = \frac{d(d-1)}{2} \tag{18}$$

Here $\mathrm{vech}_0(M)$ denotes vectorization of the nonredundant off-diagonal entries of M. This step converts each connectome into a vector in a Euclidean space, centered at the held-out real mean.

A real-defined tangent chart. Let $W_y \in \mathbb{R}^{p\times 2}$ denote the first two principal directions (PCA) of the held-out real vectors $\{z_i^{\mathrm{val},y}\}$. We define the two-dimensional chart coordinate of any connectome C

by

$$u_y(C) = W_y^\top \operatorname{vec}_\triangle\big(\log C - \mu_y^{\log,\mathrm{val}}\big) \tag{19}$$

Importantly, the basis W_y is computed using *held-out real data only*. Generated samples are projected into a chart they did not define.

Real support region in the chart. Let $\widehat{\rho}_y^{\mathrm{val}}$ be a kernel density estimate of the distribution of $\{u_y(C_i^{\mathrm{val},y})\}$. For a chosen threshold τ_y, the support region:

$$S_y = \{\, u \in \mathbb{R}^2 : \widehat{\rho}_y^{\mathrm{val}}(u) \geq \tau_y \,\} \tag{20}$$

The KDE uses a Gaussian kernel with Scott's bandwidth [72] $h_y = n_y^{-1/6}$ and empirical covariance $\widehat{\Sigma}_y$ of the real chart points [28]. The threshold $\tau_y = \mathrm{Quantile}_{0.10}\Big(\{\widehat{\rho}_y^{\mathrm{val}}(u_y(C_i^{\mathrm{val},y}))\}_i\Big)$ is the 10th percentile of real KDE scores, so S_y approximates a central 90% real-support region; all parameters are derived from held-out real data only. Intuitively, S_y captures the region where held-out real data is concentrated in the chart.

Inside-support rate. We measure how much generated mass falls inside this real-defined region via

$$\mathrm{ISR}_{m,y} = \mathbb{P}_{C\sim Q_y^m}\big(u_y(C) \in S_y\big) \tag{21}$$

This is a simple and interpretable quantity: the fraction of generated samples that lie inside the held-out real support in the tangent chart.

Proposition 5. *Suppose the chart distributions agree, i.e., $(u_y)_\# Q_y^m = (u_y)_\# P_y^{\mathrm{val}}$, and assume that the boundary of S_y has zero probability under the held-out real chart distribution. Then* $\mathrm{ISR}_{m,y} = \mathbb{P}_{C\sim P_y^{\mathrm{val}}}\big(u_y(C) \in S_y\big)$.

Table 8: **Geometric Fidelity Analysis.** Median log-Euclidean radial distances to the held-out real class center (Eq. 15). Red deltas (Δ) indicate the shift relative to the ground-truth median. TriangCFM is excluded due to severe overdispersion (see Fig. 7 insets). D-CFM is DiffeoCFM; D-Gauss is DiffeoGauss. CI indicates the class index.

Data	CI	Real	MAGNET	D-CFM	D-Gauss
ABIDE	C0	4.22	**4.65 (+.43)**	4.76 (+.53)	5.05 (+.83)
	C1	4.17	**4.70 (+.54)**	4.85 (+.68)	5.15 (+.98)
ADNI	C0	5.54	**7.15 (+1.6)**	7.43 (+1.9)	10.1 (+4.5)
	C1	5.62	7.80 (+2.2)	**7.71 (+2.1)**	10.5 (+4.9)
OASIS	C0	4.84	**5.27 (+.43)**	5.54 (+.70)	5.83 (+.98)
	C1	4.88	**5.35 (+.48)**	5.70 (+.82)	5.98 (+1.1)

Proof. The map $u_y(C) = W_y^\top \operatorname{vec}_\triangle(\log C - \mu_y^{\log,\mathrm{val}})$ is a composition of continuous operations (logarithm, subtraction, linear projection), hence it is continuous and therefore measurable. The set S_y is defined through a density threshold and is therefore measurable. By assumption, $(u_y)_\# Q_y^m = (u_y)_\# P_y^{\mathrm{val}}$. This means that for any measurable set $A \subseteq \mathbb{R}^2$,

$$\mathbb{P}_{C\sim Q_y^m}(u_y(C) \in A) = \mathbb{P}_{C\sim P_y^{\mathrm{val}}}(u_y(C) \in A).$$

Taking $A = S_y$, we obtain

$$\mathbb{P}_{C\sim Q_y^m}(u_y(C) \in S_y) = \mathbb{P}_{C\sim P_y^{\mathrm{val}}}(u_y(C) \in S_y),$$

provided that S_y is a continuity set of the distribution. This holds if the boundary of S_y has zero probability, which is assumed. Thus, $\mathrm{ISR}_{m,y} = \mathbb{P}_{C\sim P_y^{\mathrm{val}}}(u_y(C) \in S_y)$. □

Remark. According to proposition 5, samples generated by a generator that perfectly aligns with the distribution will also align with the validation support data. Thus, the empirical results of Inside-Support Rate, defined as $\mathrm{ISR}_{m,y}$ supported by Eq. 21 directly helps us to identify the two geometric failures: *over-dispersion* where the generator decides to place distribution mass outside the high-density real regions denoted by S_y and *under-coverage* where the generator fails to cover the distribution of the true support set.

Empirical Takeaway. Figure 8 and Table 9 quantify the spatial alignment of the generated density using the inside-support rate ($\mathrm{ISR}_{m,y}$) defined in Equation 21. It can be observed that the atlas (see Figure 8) shows distinct behavioral structures across the baselines. TriangCFM (*T-CFM in Table 9*) shows severe mode collapse and fails to populate the interior real support contour (S_y), generating $\mathrm{ISR}_{m,y}$ near zero across all clinical cohorts.

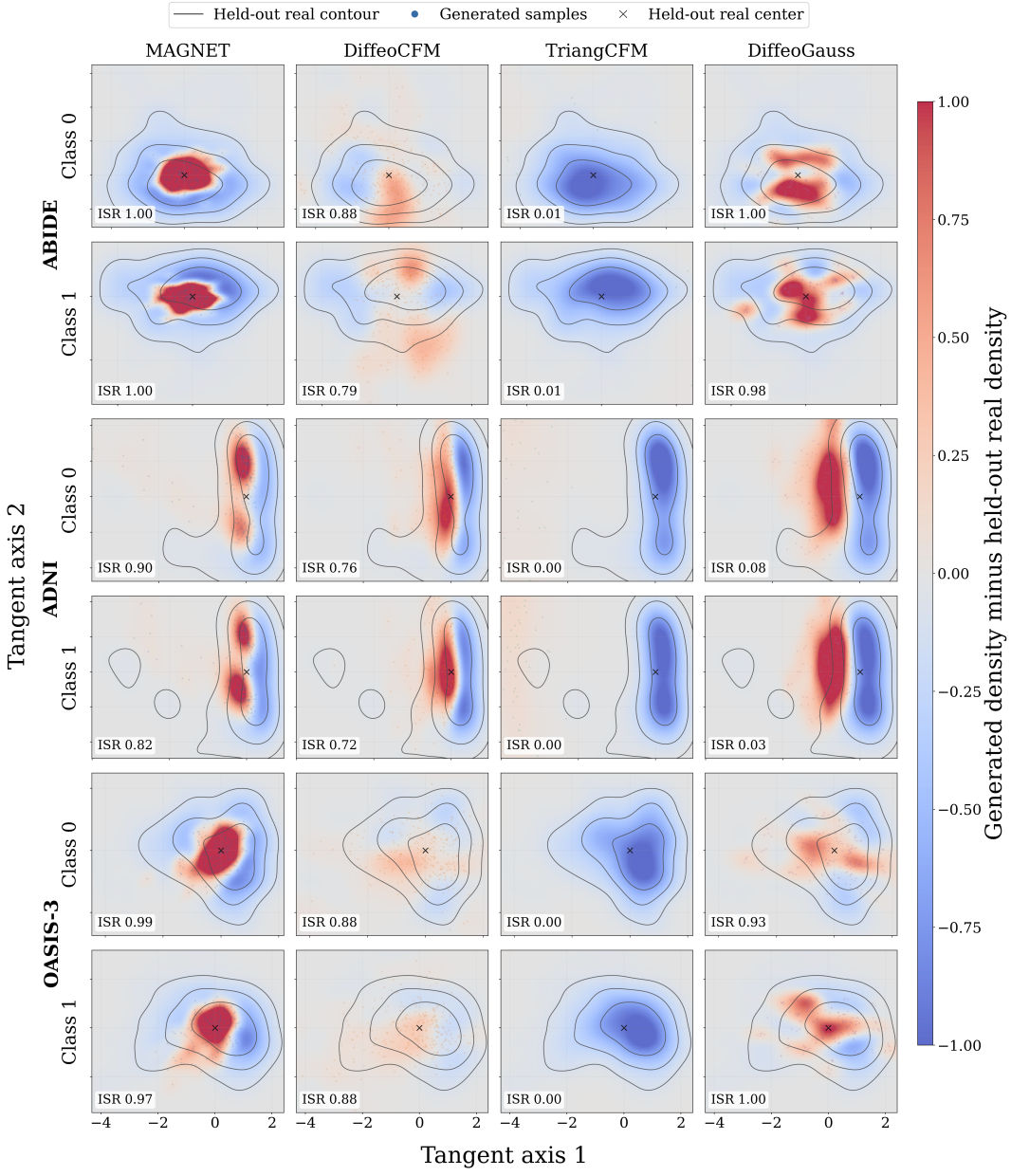

Figure 8: **Held-out real tangent-manifold occupancy atlas.** We project generated samples into a 2D tangent chart defined strictly by held-out real data to evaluate local support occupancy. **Visual guide:** Gray contours represent the true data distribution (real support, S_y). The colormap indicates the density difference (generated minus real), where red highlights areas of over-generation (excess mass) and blue indicates under-generation (missing mass). **Takeaway:** MAGNET tightly bounds its generated samples within the valid real contours, consistently achieving the highest Inside-Support Rate (ISR, Eq. 21). Conversely, DiffeoCFM leaks mass beyond the true boundaries, TriangCFM exhibits severe mode collapse, and DiffeoGauss hallucinates out-of-distribution samples under the high heterogeneity of the ADNI cohort.

DiffeoGauss (*D-Gauss in Table 9*) populates the real support contour very well on ABIDE and OASIS-3, but completely hallucinates on ADNI due to the severe heterogeneity present inside the dataset and ends up projecting beyond the true contours. On the other hand, DiffeoCFM (*D-CFM in Table 9*) tracks the correct macroscopic structure but leaks the probability mass just outside the real boundaries. In contrast, our proposed method, MAGNET, localizes the generated density inside the dominant contour of held-out real support which allows us to achieve the highest ISR across all the datasets.

Table 9: **Quantitative tangent-manifold support recovery.** Inside-support rate ($\mathrm{ISR}_{m,y}$ ↑) measures the fraction of generated samples falling within the held-out real support region (Eq. 21).This quantifies whether a model accurately allocates probability mass to validation-supported regions of the tangent-manifold atlas (Fig. 8) without relying on regions it defined itself.

Dataset	Class	MAGNET	D-CFM	T-CFM	D-Gauss
ABIDE	Class 0	**1.000**	0.879	0.008	0.996
	Class 1	**1.000**	0.787	0.008	0.983
	Avg.	**1.000**	0.833	0.008	0.990
ADNI	Class 0	**0.896**	0.762	0.000	0.083
	Class 1	**0.821**	0.721	0.000	0.029
	Avg.	**0.859**	0.742	0.000	0.056
OASIS-3	Class 0	**0.988**	0.879	0.000	0.925
	Class 1	0.967	0.875	0.000	**0.996**
	Avg.	**0.978**	0.877	0.000	0.961
Overall avg.		**0.945**	0.817	0.003	0.669

H Reverse-time refinement

In the preceding analysis, we established that generated samples from the different baselines showcase distinct behaviors. The radial calibration (Appendix G.3) helped to identify the global scale mismatch and the tangent occupancy (Appendix G.4) revealed the local issues of *over-dispersion* and *under-coverage* with respect to true support S_y. However, we can only assess the terminal samples ($t = 0$) of the generative process using those analysis. We also need to understand *how* the probability mass navigates towards the held-out real manifold during the denoising. Hence, we now move beyond the static endpoints and inspect the dynamics of the reverse-time path $C^y_{m,t}$ as $t \to 0$.

H.1 Reverse-time tangent-manifold trajectory

Let $C^y_{m,t}$ denote the connectome generated by method m at reverse-time checkpoint $t \in \{5, 4, 3, 2, 1, 0\}$ for class y. We project each checkpoint through the same held-out real chart (Eq. 19): $u^y_{m,t} = u_y(C^y_{m,t})$. The reverse-time trajectory plots in Fig. 9 and 10 therefore answer a sharply defined question: *as reverse time progresses, does the generated cloud move into the same held-out real support region identified in the final-sample occupancy atlas?*

The associated checkpoint-wise inside-support rate is

$$\mathrm{ISR}_{m,y}(t) = \mathbb{P}\big(u^y_{m,t} \in S_y\big)\,. \tag{22}$$

Remark. Note, we do not assume that $t \mapsto \mathrm{ISR}_{m,y}(t)$ is monotone. The reverse-time quality metrics in diffusion and flow-based generative trajectories need not to evolve monotonically with time [86, 51]. The regime the path follows is much more important.

Empirical Takeaway. Trajectory manifolds are shown for split 0 using all held-out validation subjects of the selected class: ABIDE (n_0=52, n_1=33), ADNI (n_0=103, n_1=145), and OASIS-3 (n_0=109, n_1=65). DiffeoCFM enters the vicinity of the real contour support earlier but MAGNET finishes with tighter support alignment in the second half of the reverse process to ensure high alignment to the real support contour. On OASIS-3, we see a modest gap between the two methods with MAGNET remaining near the real support contour and DiffeoCFM decaying mildly from 1.00 to 0.91. However, ABIDE highlights different dynamics. Both the methods remain inside the support contour throughout the denoising trajectory, indicating easier adaptability to the data as compared to the previous two datasets.

H.2 Stepwise residual refinement on a fixed held-out real edge mask

The reverse trajectory is geometric but it is not able to show edge-specific generation by the methods. We examine the residual mean connectomes on a fixed real-defined edge mask to understand how the first order structure is refined by each method over the denoising timesteps.

For class y, let $\bar{C}^y_{\mathrm{val}} = \mathbb{E}_{C \sim P^y_{\mathrm{val}}}[C]$ be the held-out real class mean. Let M_y be the binary mask containing the top held-out real edges according to $|\bar{C}^y_{\mathrm{val}}(a, b)|$. For method m at checkpoint t, we

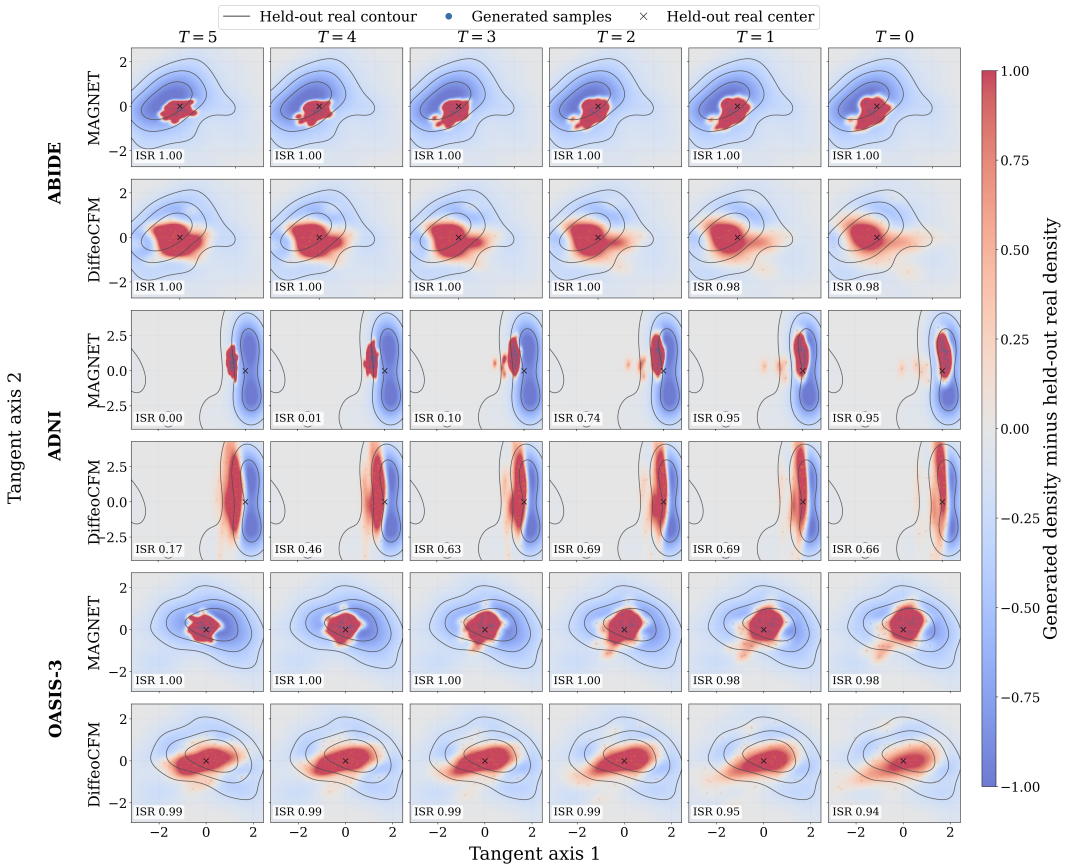

Figure 9: **Dynamic support alignment for healthy controls (Class 0).** We track the generated probability mass over the reverse-time denoising trajectory ($T = 5 \to 0$). Each checkpoint is projected into the fixed, real-data-defined tangent chart established in our occupancy atlas. **Visual Guide:** The solid gray contour demarcates the true held-out data support (S_y); the colormap tracks localized density error. **Takeaway:** MAGNET strictly bounds its trajectory within the interior of the true held-out data contour, achieving near-perfect inside-support rates without the boundary leakage typical of transport baselines.

define the generated class mean $\bar{C}^y_{m,t} = \mathbb{E}[C^y_{m,t}]$. We compare the fixed-mask residual

$$R^y_{m,t} = \left(\bar{C}^y_{m,t} - \bar{C}^y_{\text{val}}\right) \odot M_y. \tag{23}$$

First column of Fig. 11 and 12 shows the real held-out mean on the mask itself. Each subsequent column shows the "generated - real" residual using the same fixed mask. Since we keep the mask and the color scale constant across columns, we can directly compare across denoising checkpoints, avoiding any threshold drift that can arise if the support definition changes. We report two scalar statistic on the masked edges:

$$\text{MAE}^y_{m,t} = \frac{1}{|M_y|} \sum_{(a,b)\in M_y} \left|\bar{C}^y_{m,t}(a,b) - \bar{C}^y_{\text{val}}(a,b)\right|, \tag{24}$$

and

$$\text{Corr}^y_{m,t} = \text{Corr}\left(\{\bar{C}^y_{m,t}(a,b)\}_{(a,b)\in M_y}, \{\bar{C}^y_{\text{val}}(a,b)\}_{(a,b)\in M_y}\right). \tag{25}$$

Proposition 6. *For fixed m, t, and y, the following statements are equivalent:* ❶ $R^y_{m,t} = 0$, ❷ $\text{MAE}^y_{m,t} = 0$, ❸ $\bar{C}^y_{m,t}(a,b) = \bar{C}^y_{\text{val}}(a,b)$ *for all* $(a,b) \in M_y$.

Proof. We reutilize the definitions discussed in Eq. 23 and 24.

❶ $\Rightarrow$ ❸ If $R^y_{m,t} = 0$, then every masked entry of the difference is zero (Eq. 23). Hence $\bar{C}^y_{m,t}(a,b) = \bar{C}^y_{\text{val}}(a,b)$ for all $(a,b) \in M_y$.

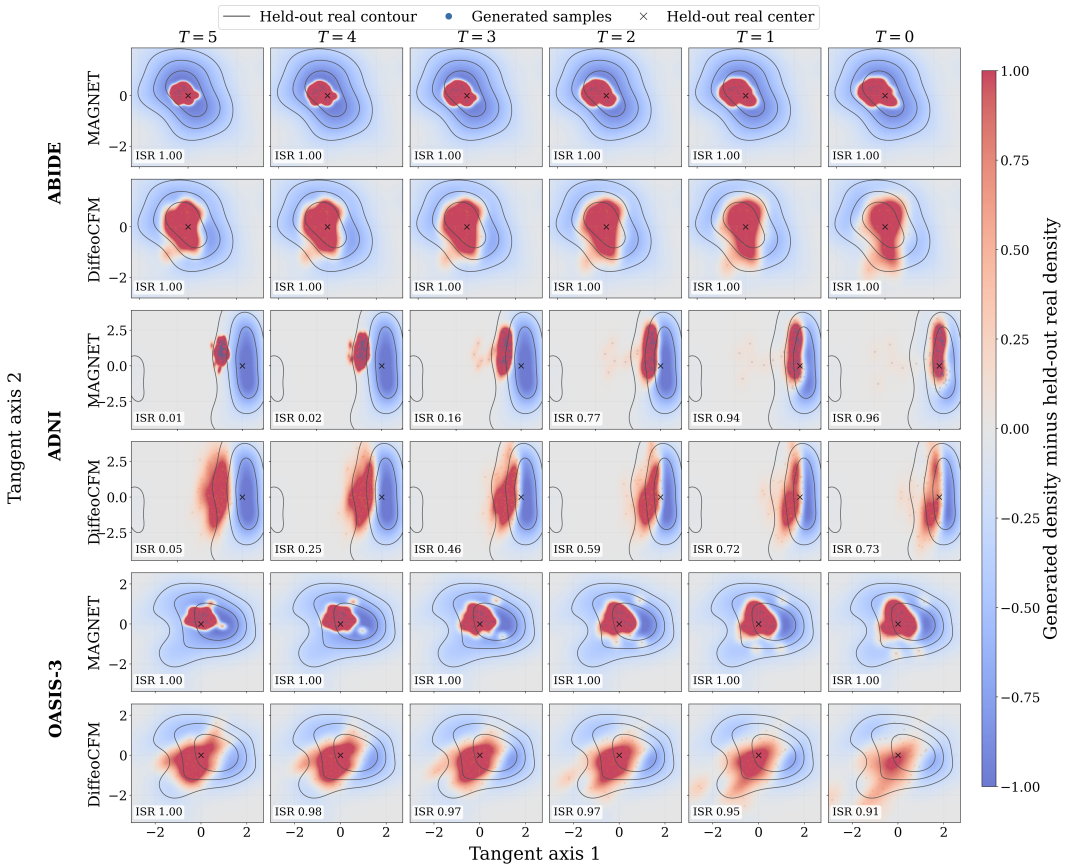

Figure 10: **Dynamic support alignment for heterogeneous disease cohorts (Class 1).** Tracking the reverse-time trajectory ($T = 5 \to 0$) against the fixed real-data tangent chart. **Visual Guide:** The solid gray contour demarcates the true held-out data support (S_y); the colormap tracks localized density error. **Takeaway:** The high structural heterogeneity of diseased subjects causes baselines like DiffeoCFM to stagnate early and bleed mass outside valid clinical regions. Conversely, MAGNET actively tightens its alignment during the critical final denoising steps, successfully navigating complex, multi-modal topologies to terminate securely within the true clinical support.

❸ $\Rightarrow$ ❷ If all masked entries match exactly, then every absolute difference in the MAE (Eq. 24) is zero, so $\mathrm{MAE}^y_{m,t} = 0$.

❷ $\Rightarrow$ ❶ The MAE is an average of nonnegative terms. If the average is zero, then every term must be zero, implying that all masked residuals vanish. Therefore $R^y_{m,t} = 0$.

Thus all three statements are equivalent. □

Remark. Using proposition 6, it is easily understandable what the connectome strip measures and what it does not. The connectome strips are first order structural refinement view on a held-out real edge set. *Note, this is not a substitute for full-distribution fidelity.* By anchoring the evaluation to a fixed, real-defined mask M_y, we guarantee that $\mathrm{MAE}^y_{m,t}$ serves as a faithful, unconfounded proxy for the residual heatmaps $R^y_{m,t}$.

Empirical Takeaway. Figures 11 and 12 extend the main-text qualitative evidence from one scan per class to three additional held-out scans per class. The same pattern persists. ADNI remains the clearest separation: DiffeoGauss oversmooths, TriangCFM often collapses toward a narrow core, and DiffeoCFM preserves some dominant edges but more frequently distorts their global arrangement. MAGNET is more consistent at preserving the leading cross-hemispheric and frontal-posterior structure while maintaining lower or competitive Δ values. ABIDE is visually easier, with all valid baselines closer to the target, but MAGNET remains competitive under the same held-out mask. OASIS-3 is more mixed, yet MAGNET still preserves the most stable global edge geometry across

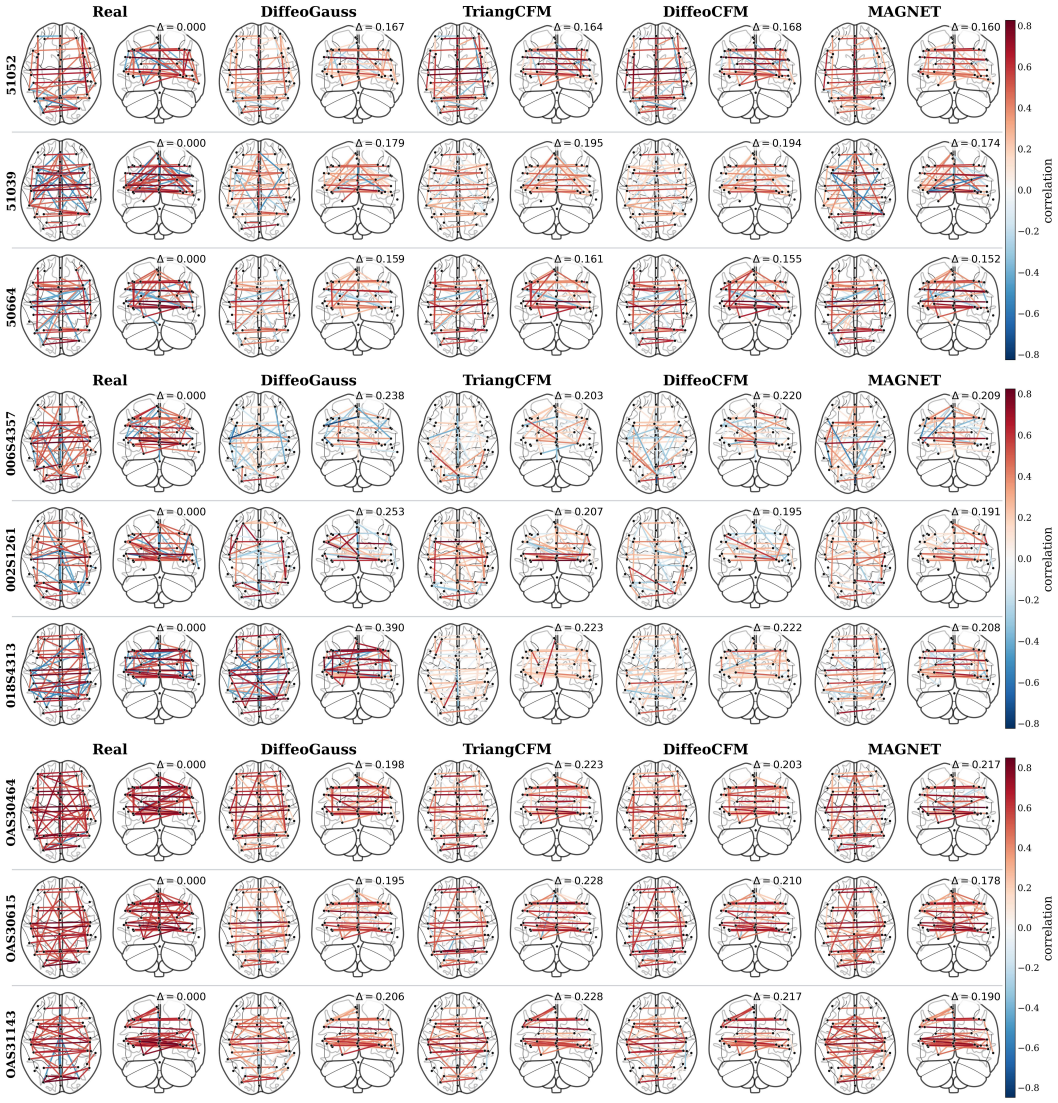

Figure 11: **Held-out connectome comparisons for class 0.** Top to bottom: ABIDE, ADNI, and OASIS-3. For each dataset, we show three held-out scans, comparing the real connectome against DiffeoGauss, TriangCFM, DiffeoCFM, and MAGNET. Each row uses the top-45 edges of the corresponding real scan and reuses the same real-defined color scale across methods. Δ denotes the upper-triangular MAE to the real correlation matrix, with smaller values indicating closer agreement.

scans. These additional examples reduce the risk that the main-text qualitative result is driven by a single curated scan and support the same structural-recovery conclusion.

I Broader Impacts

Conditional connectome synthesis lies at the intersection of two broader challenges: generating valid samples on constrained manifolds and recovering class-conditional structure under coarse supervision. MAGNET addresses both, with implications beyond neuroimaging.

On the positive side, generating geometrically valid, class-conditional connectomes supports data-scarce clinical research by augmenting underrepresented groups, enabling multi-site studies, and facilitating privacy-preserving data sharing. By modeling within-class heterogeneity via subtype conditioning, MAGNET avoids a key failure of binary-supervised generators, producing unrealistic averaged samples, thereby improving the reliability of synthetic augmentation. More broadly, the work highlights a general problem: *class-conditional manifold awareness*. In domains where

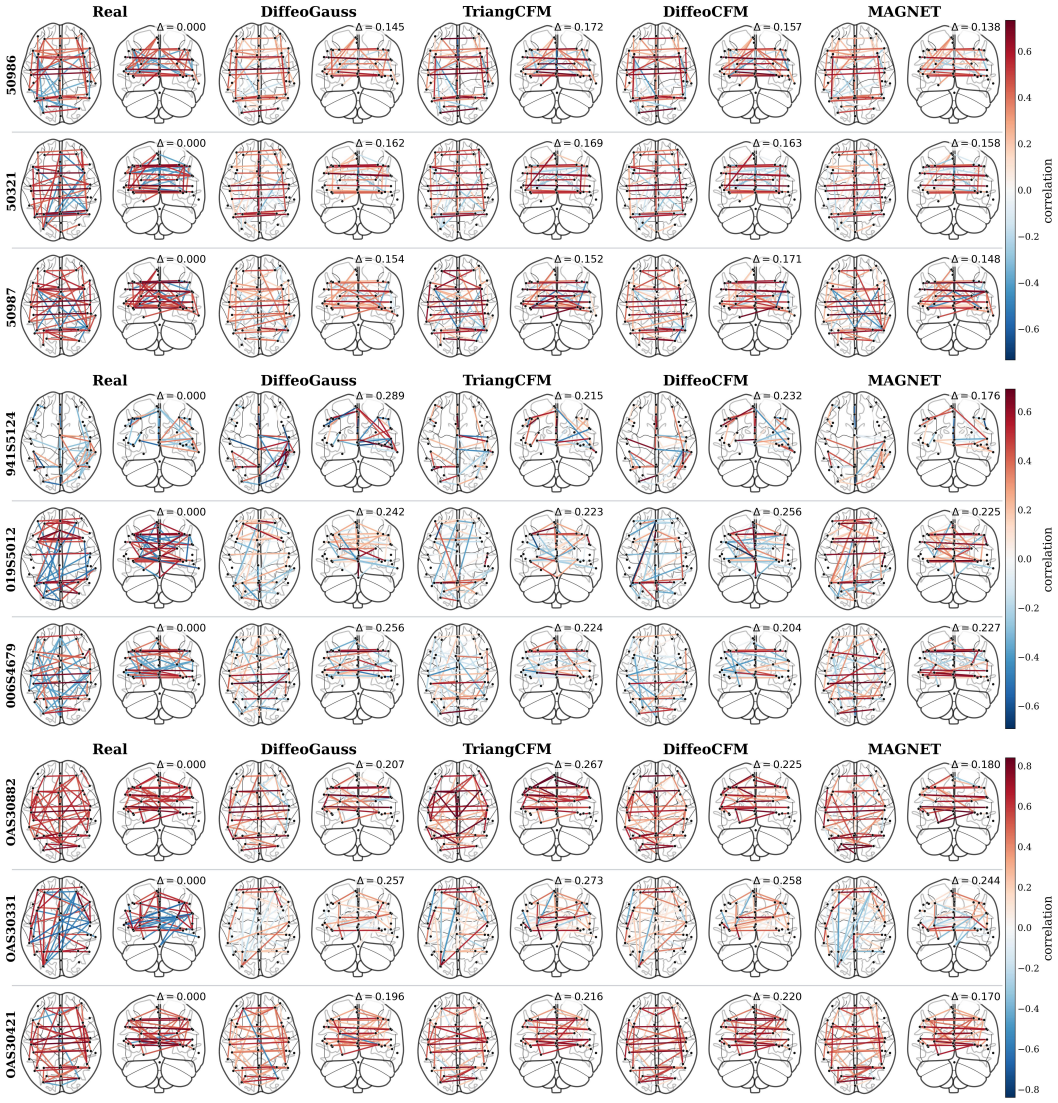

Figure 12: **Held-out connectome comparisons for class 1.** Top to bottom: ABIDE, ADNI, and OASIS-3. For each dataset, we show three held-out scans, comparing the real connectome against DiffeoGauss, TriangCFM, DiffeoCFM, and MAGNET. Each row uses the top-45 edges of the corresponding real scan and reuses the same real-defined color scale across methods. Δ denotes the upper-triangular MAE to the real correlation matrix, with smaller values indicating closer agreement.

outputs must lie on nonlinear manifolds and labels collapse diverse subpopulations (e.g., molecular generation, materials science, medical imaging), validity alone is insufficient as models may remain manifold-valid while producing implausible samples due to support averaging. MAGNET provides a transferable framework for such settings. In particular, relevance-preserving corruption ensures class-relevant structure persists during generation, with implications for diffusion-based distillation and continual learning, and potential extensions to mitigating hallucinations and poor-quality synthesis.

Limitations and risks remain. Synthetic connectomes may amplify demographic biases present in source datasets, and discovered subtypes are statistical rather than clinically validated. Additionally, highly realistic synthetic data closely mimicking real data samples (memorization) could be misused (e.g., fabricating cohorts), necessitating careful validation before deployment in sensitive applications concerning privacy.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and main text state the central claims on validity-preserving conditional support recovery, few-step sampling, and preservation of fine-grained clinical structure. The main benchmark is supported by three cohorts, while Table 4 additionally evaluates robustness to higher-resolution anatomical and functional parcellations.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The limitations of the proposed MAGNET framework is present in the Conclusion (sec. 4).

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: The paper formally states propositions in Section 2 and provides complete mathematical proofs alongside structural assumptions in Appendices B, G, and F.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The paper thoroughly details the experimental setup, dataset splits, preprocessing steps (MSDL atlas, z-scoring), sampling budgets, and network hyperparameters in Section D and Appendix E.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: Section D and Appendix E specify the data splits, preprocessing, architecture, training and sampling procedures, evaluation protocol, and principal hyperparameters. The accompanying code repository provides the implementation and scripts used for the reported benchmarks and ablations.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Experimental details, including the 10 disjoint train/test splits, balanced class sampling protocols, SNR loss reweighting, and precise neural architecture dimensions are documented in Section D and Appendix E.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

Justification: All empirical tables (e.g., Tables 1, 2, 5, etc.) report standard deviations computed across the 10 disjoint train/validation splits for every metric.

Guidelines:

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Wall-clock sampling times are reported in Table 1. Hardware details (single NVIDIA A6000 GPU, 48GB VRAM) are specified in Section D.

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research relies strictly on standard, pre-anonymized public datasets (ABIDE, ADNI, OASIS-3) and complies fully with ethical guidelines regarding human data and research conduct.

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: Broader impacts including potential negative societal risks (demographic bias amplification, misuse of synthetic cohorts in clinical settings) and mitigation strategies are discussed in the Broader Impacts in Appendix I.

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The models generated in this research (synthetic functional connectomes) are highly domain-specific and do not pose the generalized high-risk for malicious misuse associated with large language models or deepfake image generators.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The creators of the utilized datasets (ABIDE, ADNI, OASIS-3) and software libraries (e.g., Nilearn) are properly credited with citations throughout the text.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The research utilizes existing datasets and does not introduce new primary data assets.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: This research relies entirely on retrospective analysis of existing, de-identified databases and does not involve primary crowdsourcing or new human subject interaction.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: Because the study strictly utilizes pre-collected, fully anonymized public neuroimaging datasets, new Institutional Review Board (IRB) approval is not required.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: Large Language Models were not utilized as a component of the core generative methodology or scientific evaluation in this research.