Orthogonal Sparse Subgraph Alignment for Structure-Function Coupling in Brain Networks

Haonan Gao[1] **Daeyoung Ham**[2] **Yifei Zhang**[1] **Xinyuan Tian**[1]
Shengxian Ding[1] **Yize Zhao**[1,†] **Tianxi Li**[3,†]

[1]Department of Biostatistics, Yale School of Public Health, Yale University
[2]Department of Statistics and Data Science, University of Texas at San Antonio
[3]School of Statistics, University of Minnesota, Twin Cities

Abstract

Characterizing the relationship between brain anatomical wiring and functional coordination remains a fundamental challenge in computational neuroscience. This difficulty primarily arises because structure-function coupling in the human brain is spatially heterogeneous, often localized to subnetworks, and inherently difficult to align across disparate representational modalities. To address this methodological gap, we introduce **O**rthogonal **S**parse **S**ubgraph **A**lignment (OSSA), a principled and interpretable framework designed to identify compact structural cliques that exhibit maximal concordance with functional connectivity (FC) across subjects. Within each FC-derived community, OSSA optimizes an alignment objective over an orthogonal transformation, which absorbs coordinate mismatch between the two connectivity modalities, and a sparse node-selection vector, which identifies the induced structural connectivity (SC) subgraph. Furthermore, for canonical OSSA within a fixed population community, we establish local statistical guarantees under the required assumptions. With high probability, an empirical local maximizer exists in a specified neighborhood of the population target, and every local maximizer in that neighborhood exactly recovers the active support and satisfies an nonasymptotic error bound for the aligned functional projector. Extensive empirical evaluations on simulated data, alongside analyses of the Adolescent Brain Cognitive Development (ABCD) baseline and longitudinal cohorts, demonstrate that OSSA successfully recovers aligned structural cliques under latent cross-modal rotations. Ultimately, the proposed methodology significantly improves out-of-sample SC–FC alignment over size-matched baselines and identifies compact structural supports with longitudinal consistency in node selection, while multi-support diagnostics suggest a dominant aligned core within the estimated FC communities.

1 Introduction

Understanding how anatomical wiring governs coordinated functional activity in the brain remains a fundamental problem in computational neuroscience. This objective naturally motivates the study of two complementary representations of the brain connectome. Specifically, structural connectivity (SC), inferred from diffusion imaging, charts the physical pathways along which regions communicate, whereas functional connectivity (FC), derived from correlated activity fluctuations, captures the statistical dependence between regional signals [16, 40]. Characterizing the relationship between SC and FC has emerged as a salient analytical task: it provides a principled framework to investigate

[†]Corresponding authors: Yize Zhao (yize.zhao@yale.edu) and Tianxi Li (tianxili@umn.edu).

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

how anatomy constrains function, how individual variability manifests across modalities, and which subnetworks offer the most robust structural support for observed functional coordination [22, 15].

Meanwhile, the SC-FC relationship is **intrinsically indirect and spatially heterogeneous** [41, 15]. An extensive body of literature demonstrates that functional connectivity is only partially constrained by direct anatomical pathways, and the cross-modal agreement varies systematically across cortical hierarchies, individuals, and developmental stages [22, 42, 4, 51, 17]. Specifically, while certain functional communities are anchored by a dense set of anatomical connections, others rely on diffuse or polysynaptic interactions [49, 19, 38, 51]. Consequently, the most pronounced cross-modal signals are concentrated within localized subnetworks rather than distributed uniformly across the connectome [42, 51, 17]. Existing analytical approaches often summarize SC–FC coupling using a single connectome-wide correlation or region-wise scalar scores [22, 42, 4]. Global correlations obscure spatial variation, whereas region-wise scores characterize such variation without directly identifying a compact node set whose induced structural subgraph maximizes cross-modal concordance. To move beyond global and region-wise scalar summaries, we therefore seek a framework that identifies structural subnetworks with maximal SC–FC concordance within functional communities.

Extensive prior research has sought to characterize the relationship between structural and functional connectivity (SC-FC). Global correlation approaches [22, 41, 15] reduce the connectome to a single summary statistic, thereby failing to identify localized subnetworks. Predictive SC-to-FC models [22, 1, 41] aim to explain functional connectivity from structural connectivity, while graph-theoretic summaries [5, 36] characterize network topology. These formulations do not directly target the joint estimation of a cross-modal alignment transformation and a sparse node set defining an induced structural subgraph. Subnetwork inference [50], functional parcellation [49, 10], and regional SC–FC analysis [51] provide complementary tools for characterizing localized network organization. Here, SC and FC share anatomical node labels, and alignment relates latent functional geometry to a selected structural subgraph while preserving the anatomical identities of its nodes. In practice, leading functional gradients are often rotated or otherwise transformed relative to the structural connectome [29, 42, 46, 17]. Consequently, a localized search that ignores this mismatch tends to recover the structurally densest subset rather than the subnetwork genuinely aligned with functional organization.

Our primary objective in this paper is to address this fundamental challenge: given paired SC and FC matrices across subjects, we aim to identify a sparse, interpretable set of brain regions whose induced structural subgraph is maximally concordant with the dominant functional organization. Specifically, the proposed methodology is designed to overcome the two primary limitations of existing SC-FC analyses. First, the relevant signal manifests at the subnetwork level. Concordance is an intrinsic property of a node set rather than of individual nodes. Consequently, a valid estimation procedure must jointly model the connection patterns of node sets instead of evaluating individual regions sequentially. Second, as FC and SC are typically embedded in disparate coordinate systems, the two modalities may differ by a nontrivial transformation that a purely localized search cannot adequately capture. Therefore, a rigorous methodological solution must explicitly account for this cross-modal mismatch to localize the target subnetwork, while simultaneously ensuring the statistical interpretability of the identified brain regions.

To address this methodological gap, we introduce **O**rthogonal **S**parse **S**ubgraph **A**lignment (**OSSA**), a novel framework that integrates these dual requirements into a unified variational problem. Specifically, OSSA couples an orthogonal transformation component with a sparse subgraph selection component (Figure 1). The method jointly estimates an orthogonal transformation of the functional connectivity and a sparse node-selection vector, thereby inducing a structural subgraph that maximizes cross-modal concordance. The transformation component serves as a structure-aware alignment mechanism, absorbing the global coordinate mismatch between modalities while preserving the intrinsic geometry of the functional data. Concurrently, the sparse selection component directly encodes the combinatorial subgraph structure within the continuous optimization space. Together, these components extend the classical orthogonal Procrustes problem [37] to a setting where the alignment target is itself an unknown, induced subgraph of the structural connectome. More generally, OSSA addresses the joint alignment and localization of paired similarity representations on a common index set when both the transformation and the informative subset are unknown. Related settings include comparing learned representations with localized shared information and aligning embedding spaces with only a subset of reliable anchors.

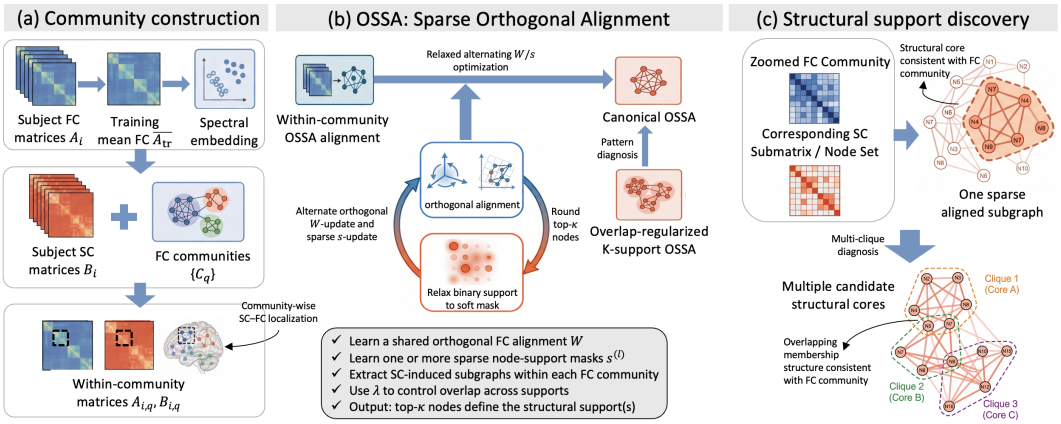


Figure 1: Overview of the OSSA. **(a)** SC and FC are restricted to functional communities obtained from a spectral clustering of population-averaged FC. **(b)** Within a community, OSSA jointly learns an orthogonal FC alignment W and K sparse node-support masks. **(c)** Rounding each mask gives an induced SC subgraph. An overlap penalty controls whether the K supports collapse or separate.

To the best of our knowledge, OSSA represents the first principled statistical formulation for the cross-modal subnetwork identification problem in SC-FC analysis. The problem can be solved by a computationally efficient algorithm that seamlessly integrates the orthogonal alignment and sparse selection components. Furthermore, we establish theoretical guarantees for canonical OSSA within a fixed population community. Under the required conditions, a local maximizer exists near the population target, and every local maximizer in the specified neighborhood achieves exact support recovery and the finite-sample aligned functional projector error bound with high probability. The empirical utility of the proposed methodology is evaluated through extensive simulation experiments, which highlight its superior accuracy and robustness compared to conventional baseline methods. Finally, we apply OSSA to analyze multi-modal brain connectome data from the Adolescent Brain Cognitive Development (ABCD) study. Our analysis successfully extracts highly interpretable structural modules that tightly couple with functional organization, yielding novel scientific insights into the anatomical constraints of the developing human brain. The identified supports also remain highly consistent across the baseline and Year-2 follow-up releases, indicating that the dominant aligned cores are stable under longitudinal neurodevelopmental variation.

2 Orthogonal sparse subgraph alignment

Consider a study comprising n subjects, indexed by $i \in \{1, \ldots, n\}$. For each subject i, we observe a paired set of $P \times P$ symmetric matrices, A_i and B_i, representing the functional connectivity (FC) and structural connectivity (SC) networks, respectively, defined on a common atlas of P brain regions. We assume that A_i and B_i have been preprocessed according to standard pipelines in brain connectome studies.

Communities. Let $\{\mathcal{C}_q\}_{q=1}^{Q}$ denote a partition of the atlas nodes derived from the training FC data. Each set $\mathcal{C}_q$ represents a module of brain regions, hereafter referred to as a functional community, which is assumed to be shared across all subjects. Such community structures are commonly studied in the context of functional connectivity patterns [18, 7]. Let $p_q = |\mathcal{C}_q|$ denote the size of community q. Under the assumption of a multiple degree-corrected block structure, the algorithm proposed by [3] exactly identifies the communities $\mathcal{C}_q$ with high probability. Consequently, we condition our subsequent analysis on the successful recovery of these communities, and we formulate the OSSA objective within each FC-derived community. In practice, the community detection algorithm of [3] is applied prior to OSSA. Define $A_{i,q}$ and $B_{i,q}$ as the principal submatrices of A_i and B_i, respectively, induced by the node set $\mathcal{C}_q$. For notational conciseness, when the context restricts focus to a single community, we suppress the subscript q and overload the notation, denoting the submatrices as A_i and B_i, and the community size as p.

Algorithm 1 Canonical OSSA

Require: Symmetric matrices $\{A_i, B_i\}_{i=1}^n$ in one community; budget κ; initial steps η_W, η_s; backtracking factor $\beta \in (0,1)$

Ensure: Alignment W, relaxed support s, and hard support $\widehat{\mathcal{V}}$

Initialize $W \leftarrow I_p$ or a random orthogonal matrix; initialize $s \in \mathcal{S}_\kappa$
repeat
  $S \leftarrow \mathrm{diag}(s)$ and $M_i \leftarrow SB_iS$ for $i = 1, \ldots, n$
  $G_W \leftarrow 2\sum_{i=1}^n A_i W M_i$ and $W^+(\eta) \leftarrow \Pi_{O(p)}(W + \eta G_W)$
  Choose the largest $\eta \in \{\eta_W \beta^m : m \geq 0\}$ satisfying $J(W^+(\eta), s) \geq J(W, s)$, and set $W \leftarrow W^+(\eta)$
  $D \leftarrow \mathrm{sym}\{\sum_{i=1}^n (W^\top A_i W) \circ B_i\}$
  $g_s \leftarrow 2Ds$ and $s^+(\eta) \leftarrow \Pi_{\mathcal{S}_\kappa}(s + \eta g_s)$
  Choose the largest $\eta \in \{\eta_s \beta^m : m \geq 0\}$ satisfying $J(W, s^+(\eta)) \geq J(W, s)$, and set $s \leftarrow s^+(\eta)$
until the objective and iterate changes satisfy the stopping tolerances
$\widehat{\mathcal{V}} \leftarrow \mathrm{Top}_\kappa(s)$
return W, s, $\widehat{\mathcal{V}}$

Fixing each community, our target is to recover a shared subset of brain regions $\mathcal{V} \subseteq \{1, \ldots, p\}$, where $|\mathcal{V}| \leq \kappa$, such that the $\mathcal{V}$-induced SC subgraph has maximal agreement with the FC geometry across subjects. This $\mathcal{V}$-induced SC subgraph thus serves as a structural backbone for the FC patterns.

2.1 The canonical OSSA

Let $O(p)$ be the set of all $p \times p$ orthogonal matrices. Fixing one community, the OSSA problem is defined as the following optimization problem

$$\begin{aligned} \max_{W,s} \quad & J(W,s) = \sum_{i=1}^n \mathrm{Tr}\big[\big(W^\top A_i W\big)\,(SB_iS)\big] \\ \text{s.t.} \quad & W^\top W = I_p, \qquad s \in \{0,1\}^p, \qquad \mathbf{1}^\top s = \kappa. \end{aligned} \tag{1}$$

where Tr is the trace operator. Two parameters are introduced in the OSSA problem. The selection vector s identifies a subset of nodes from the SC, such that SB_iS represents the induced structural subgraph of B_i. Concurrently, the rotation matrix W captures a shared cross-modal transformation mapping. The objective function $J(W, s)$ explicitly evaluates the alignment between the transformed FC matrix and the corresponding induced SC subgraph. Because the induced matrix SB_iS is generally expected to form a locally dense structural module within the selected node set, it can be conceptualized as a structural clique of the SC network, a topological connectivity pattern widely documented in empirical connectome studies [34, 39, 33]. Both components are indispensable to this formulation: without the orthogonal transformation W, the model would impose an exact coordinate-level comparability between FC and SC, a restrictive assumption rarely supported by empirical connectome data. Conversely, without the sparse selection vector s, the objective degenerates into a global alignment metric, failing to achieve precise anatomical localization. As OSSA is formulated within individual functional communities, its integration with functional community detection establishes a hierarchical refinement of the cross-modal SC-FC relationship.

Directly optimizing the objective in (1) is NP-hard due to the combinatorial nature of the binary constraints. To circumvent these difficulties, we employ a continuous relaxation of the domain for s, formulating the following surrogate optimization problem.

$$\max_{W \in O(p),\ s \in \mathcal{S}_\kappa} J(W,s), \qquad \mathcal{S}_\kappa = \{s \in [0,1]^p : \mathbf{1}^\top s \leq \kappa\}. \tag{2}$$

Upon solving the relaxed formulation in (2), the final induced subnetwork is recovered via a hard thresholding procedure, $\widehat{\mathcal{V}} = \mathrm{Top}_\kappa(\widehat{s})$, where Top_κ extracts the indices corresponding to the κ largest entries. This continuous relaxation preserves the natural interpretation of s as a soft subgraph-membership indicator while simultaneously rendering the optimization computationally tractable. We solve (2) by a projected block-coordinate ascent procedure, outlined in Algorithm 1.

Theoretical properties. We study (2) within a fixed population community. The matrices restricted to that community satisfy the sampling conditions in Assumptions A1–A2. Define

$$\bar{J}_n(W,s) = \frac{J(W,s)}{n}, \qquad \mathcal{J}(W,s) = \mathbb{E}\bar{J}_n(W,s). \tag{3}$$

For $V \subseteq \{1,\dots,p\}$, let $\mathbf{1}_V$ be its indicator vector and $P_V = \mathrm{diag}(\mathbf{1}_V)$. Fix $V^\star$ with $|V^\star| = \kappa$ and write $u^\star = \mathbf{1}_{V^\star}$. Assumption A3 specifies a compact smooth embedded submanifold $\mathcal{W}^\star \subset O(p)$ and a rank-κ orthogonal projector $\Pi^\star$ such that $\mathcal{J}(U,u^\star) = J^\star$, $\mathrm{grad}_W\, \mathcal{J}(U,u^\star) = 0$, and $UP_{V^\star}U^\top = \Pi^\star$ for every $U \in \mathcal{W}^\star$. Thus, the matrices in $\mathcal{W}^\star$ are stationary alignments with the same population objective value, and their columns indexed by $V^\star$ span the same subspace. Here grad_W denotes the gradient for the Frobenius metric on $O(p)$. Let $d_O(W,\mathcal{W}^\star)$ denote the minimum geodesic distance from W to $\mathcal{W}^\star$, namely, the shortest path length needed to reach $\mathcal{W}^\star$ while maintaining orthogonality, with path length measured using the Frobenius norm.

Theorem 1 establishes the existence of local maximizers satisfying (133) and proves $\widehat{s} = u^\star$ and the bound on $\|\widehat{W}P_{V^\star}\widehat{W}^\top - \Pi^\star\|_F$ in (4) for every such maximizer.

Theorem 1. *Suppose Assumptions A1–A5 hold. In addition, let $x > 0$ and $0 < r < r_0$ satisfy Assumption A6. With probability at least $1-2e^{-x}$, (2) has a local maximizer $(\widehat{W},\widehat{s})$ in a neighborhood of $(\mathcal{W}^\star,u^\star)$ satisfying*

$$\widehat{s} = u^\star, \qquad \|\widehat{W}P_{V^\star}\widehat{W}^\top - \Pi^\star\|_F \le \frac{8(\kappa-1)\sqrt{\kappa}\,\eta_n(x)}{\alpha_n}. \tag{4}$$

Here $\eta_n(x)$, ν_n, α_n, and m_n are specified in Section B.

Proof is given in Section B. Theorem 1 establishes the existence of local maximizers in the neighborhood of the truth and gives two recovery guarantees for every such maximizer. First, the relaxed support is binary and exactly identifies the true SC node set. Second, the inequality in (4) bounds the error in recovering the corresponding FC subspace. Specifically, $\widehat{W}P_{V^\star}\widehat{W}^\top$ is the orthogonal projector onto the span of the columns of $\widehat{W}$ indexed by the recovered SC support. Its distance from $\Pi^\star$ measures how closely this estimated subspace matches the population target, regardless of how the basis within each subspace are chosen. Along any sequence satisfying the assumptions of Theorem 1 with $x = x_n \to \infty$, the estimated projectors of these localized maximizers are consistent for $\Pi^\star$ whenever the right-hand side of the final inequality in (4) tends to zero.

2.2 Extension to multiple cliques

The canonical OSSA framework identifies a single dominant clique within each functional community. This configuration is practically well-justified, as communities in functional connectivity networks are typically already restricted in scale (comprising approximately 10-30 nodes). Further partitioning these modules into multiple cliques is rarely necessary from a dimensionality reduction perspective. Indeed, as demonstrated in our analysis of the ABCD dataset (Section 4), empirical evidence strongly corroborates the prevalence of a single-clique structural topology within these localized functional modules. Nevertheless, for broader applicability, the OSSA formulation can be naturally extended to extract multiple cliques per community. Specifically, we generalize the method by introducing K to denote the predefined number of target cliques to be identified within a given community, as OSSA-K. Define

$$s^{(\ell)} \in \mathcal{S}_\kappa, \qquad S^{(\ell)} = \mathrm{diag}(s^{(\ell)}), \qquad \mathcal{S}_\kappa = \{s \in [0,1]^p \colon \mathbf{1}^\top s \le \kappa\}, \quad 1 \le \ell \le K.$$

The K-clique OSSA problem solves

$$\max_{W\in O(p),\ s^{(1)},\dots,s^{(K)}\in\mathcal{S}_\kappa} \sum_{\ell=1}^{K}\sum_{i=1}^{n} \mathrm{Tr}\Big[\big(W^\top A_i W\big)\Big(S^{(\ell)}B_iS^{(\ell)}\Big)\Big] - \lambda \sum_{\ell<\ell'} \Big\langle s^{(\ell)},\, s^{(\ell')}\Big\rangle. \tag{5}$$

This objective is a direct generalization of the canonical formulation, incorporating an explicit regularization term to penalize redundant or highly overlapping node memberships. Crucially, the orthogonal transformation matrix W is shared across all structural subgraphs. This shared rotation ensures that all candidate subnetworks are evaluated within a unified functional coordinate system, thereby preventing clique-specific transformations from artificially absorbing the underlying separation structure. We employ a similar projected block-coordinate ascent procedure. The complete optimization routine is detailed as Algorithm 2 in Appendix C.

Remark. We employ the K-clique extension primarily as a diagnostic tool. If multiple distinct cliques adequately explain the functional connectivity, the estimated supports should diverge at moderate values of the regularization parameter λ, while maintaining robust cross-modal alignment on held-out data. Conversely, if a given FC community contains only a single aligned clique, the supports would collapse into a unified structure at small values of λ. When artificially forced apart by increased penalization, the extraneous supports may capture generic correlation structures but will cease to remain competitive under the primary alignment objective.

3 Simulation experiments

We first evaluate OSSA in simulations with known target supports. These experiments isolate support recovery, the role of the orthogonal block, and whether the K-support extension separates repeated recovery from distinct cores. All simulation experiments use fixed train/validation/test splits. We use the $W = I_p$ ablation, OSSA-NOW, to isolate the learned rotation. Uncertainty is reported as standard error over simulation replicates.

3.1 Evaluation of the canonical OSSA

Experiment setup. We first test recovery of one planted clique under FC-SC coordinate mismatch. Each replicate contains $n = 30$ subjects in one community of size $p = 40$, with a planted support $\mathcal{V}^\star$ of size $\kappa^\star = 10$. For each subject i,

$$B_i = C_i + \zeta_i^B + \sigma_B E_i^B, \qquad A_i = W^\star C_i {W^\star}^\top + \zeta_i^A + \sigma_A E_i^A,$$

with all matrices symmetrized and diagonal entries set to zero. The shared signal is $C_i = \mathrm{LR}_3(\mathcal{V}^\star, 1.5)$, where $\mathrm{LR}_r(D, c)$ denotes a symmetric mixed-sign block supported on $D \times D$, generated as $UU^\top \circ \Sigma$ with entries $U_{ab} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, c^2/r)$ and a symmetric random sign matrix Σ. The additional terms ζ_i^A, ζ_i^B collect subject-level background and modality-specific nuisance variation, their exact construction is given in Appendix D.1. These terms are included to stress-test recovery under extra single-modality variation, not as part of the OSSA model itself. We control mismatch by $W^\star = \mathrm{polar}((1-\alpha)I_p + \alpha Q)$, where $Q \sim \mathrm{Haar}(O(p))$ and $\mathrm{polar}(X) = UV^\top$ for $X = U\Sigma V^\top$ is the orthogonal factor in the polar decomposition. Thus $\alpha = 0$ gives matched coordinates, and $\alpha = 1$ gives a random rotation.

We compare OSSA with size-matched baselines that select κ nodes by simple marginal or graph summaries: (1) Marginal degree: SC-DEGREE and FC-DEGREE take the top-κ nodes by row-sum of $\overline{|B|}$ and $\overline{|A|}$, respectively, where the average is over subjects. (2) Edgewise modality fusion: JOINT-SCXFC takes the top-κ nodes by row-sum of $\overline{|A_i \circ B_i|}$. (3) SC graph summary: COMMUNICABILITY [14] takes the top-κ nodes by column-sum of $\exp(\overline{B})$. (4) RANDOM samples a uniform κ-node subset as a control floor. We also include two OSSA controls: OSSA-NOW fixes $W = I_p$, and OSSA-ORACLE fixes $W = W_\star$ as an upper bound. All OSSA variants use eight restarts, and we keep the largest training objective. We report Jaccard similarity with the planted support and exact recovery rate, averaged over 10 independent replicates for each (α, σ, n) setting. Detailed setup is defined in appendix D.1.

Results. Among the tested methods, OSSA remains accurate under latent FC-SC rotation (Table 1; full rotation and noise sweeps are reported in Appendix D.2). In the full sweep, fixed-coordinate methods work at $\alpha = 0$ but fail as rotation increases. At full rotation, OSSA reaches Jaccard 0.861 ± 0.042 and exactly recovers $\mathcal{V}^\star$ in 40% of replicates, compared with 0.205 ± 0.045 for OSSA-NOW. The oracle run with $W = W^\star$ is nearly perfect, suggesting that the remaining gap comes from estimating W. OSSA also remains stable at $\sigma = 0.5$ (0.827 ± 0.043) and reaches 0.85-0.89 for $n \geq 30$, where baselines show little gain with larger n.

3.2 Multi-clique simulation

We further evaluate OSSA-K in a multi-clique setting. We use one community ($p = 50$, $n = 30$), a shared latent rotation $\alpha = 0.6$, and noise $\sigma = 0.4$, with either two disjoint supports, two supports sharing four nodes, or three disjoint supports. We compare against random, SC-spectral, communicability, and SEQUENTIAL-OSSA K-support baselines. SEQUENTIAL-OSSA is a greedy

Table 1: Canonical OSSA clique support recovery on three diagnostic slices (mean $\pm$ s.e.). EXACT reports $\Pr(\widehat{\mathcal{V}} = \mathcal{V}^\star)$ at $\alpha = 1$. OSSA-ORACLE is reported separately as an upper bound.

Method	Rot. α=1		Noise σ=0.5	n=80 (α=1)
	Jaccard	Exact	Jaccard	Jaccard
OSSA	$\mathbf{0.861 \pm 0.042}$	0.40	$\mathbf{0.827 \pm 0.043}$	$\mathbf{0.894 \pm 0.038}$
RANDOM	0.162 ± 0.028	0.00	0.154 ± 0.025	0.219 ± 0.031
SC-DEGREE	0.203 ± 0.027	0.00	0.207 ± 0.034	0.174 ± 0.024
FC-DEGREE	0.037 ± 0.012	0.00	0.050 ± 0.018	0.045 ± 0.019
JOINT-SCXFC	0.157 ± 0.031	0.00	0.190 ± 0.028	0.263 ± 0.028
COMMUNICABILITY	0.179 ± 0.035	0.00	0.186 ± 0.035	0.162 ± 0.027
OSSA-NOW	0.205 ± 0.045	0.00	0.199 ± 0.033	0.240 ± 0.040
OSSA-ORACLE	1.000 ± 0.000	1.00	1.000 ± 0.000	0.982 ± 0.018

Table 2: Multi-support simulation summary (mean $\pm$ s.e.). The OSSA-K and best-baseline columns report best-match Jaccard. Fitted overlap is the mean pairwise support overlap. The full baseline table is in Appendix D.6.

Regime	Planted ovl.	OSSA-K Jacc.	Best base. Jacc. (SEQ.-OSSA)	OSSA-K ovl.
Disjoint $K = 2$	0	$\mathbf{0.945 \pm 0.023}$	0.710 ± 0.040	0.00 ± 0.00
Overlap $r = 4, K = 2$	4	$\mathbf{0.910 \pm 0.066}$	0.448 ± 0.026	$\mathbf{3.75 \pm 0.56}$
Disjoint $K = 3$	0	$\mathbf{0.975 \pm 0.016}$	0.744 ± 0.063	0.00 ± 0.00

baseline that fits canonical OSSA, masks the selected nodes, and refits. Details of the generator, baselines, λ grid, and full results are in Appendices D.6-D.7.

Table 2 summarizes the diagnostic outcome. OSSA-K recovers all three regimes. In the overlap case, it estimates 3.75 ± 0.56 shared nodes for a planted overlap of four. The strongest baseline, SEQUENTIAL-OSSA, works mainly in disjoint regimes and degrades under planted overlap because its masking step enforces disjoint supports. Thus the shared-W objective in (5) recovers multi-clique structure while allowing controlled overlap.

3.3 Robustness

We run three robustness checks for choices that are usually unknown in practice, full results are in Appendix D. **(i) Sparsity-budget misspecification** (Appendix D.4): When κ is swept from 4 to 16 around $\kappa^\star = 10$ at $\alpha = 0.6$ and $\sigma = 0.4$, OSSA stays above all non-OSSA baselines. Smaller budgets favor precision and larger budgets favor recall. **(ii) Selection stability** (Appendix D.5): Bootstrap selection frequency is higher on planted than non-planted nodes (0.46 vs. 0.18), and per-restart pairwise Jaccard has median 0.43. The observed variation in selected supports across initializations indicates sensitivity to initialization in the joint nonconvex optimization problem. We therefore use multiple restarts in the main experiments and keep the run with the largest training objective. **(iii) Continuous λ-trajectory** (Appendix D.7): the dense λ grid spans four decades, and fitted overlap decreases from collapse, through the planted value, to full disjointness. These checks motivate the fixed-budget, multi-restart, and λ-diagnostic choices used for ABCD.

4 Cross-modal analysis of adolescent brain connectome

We next evaluate OSSA on Adolescent Brain Cognitive Development (ABCD) study [6, 20], a large multi-site longitudinal neuroimaging study of adolescent brain development. We use ABCD structural and functional connectomes on the 68-region Desikan-Killiany atlas [11]. The main evaluation uses the baseline release: FC is signed Pearson correlation and SC is normalized streamline density. We symmetrize both modalities, ignore diagonals, Fisher-z transform FC, clip negative FC values, and apply $\log(1 + \cdot)$ to SC. A cohort-median 5σ QC rule on off-diagonal means removes 10 subjects, leaving $n_{\mathrm{ABCD}} = 2{,}350$. We use deterministic 60/20/20 train/validation/test splits over

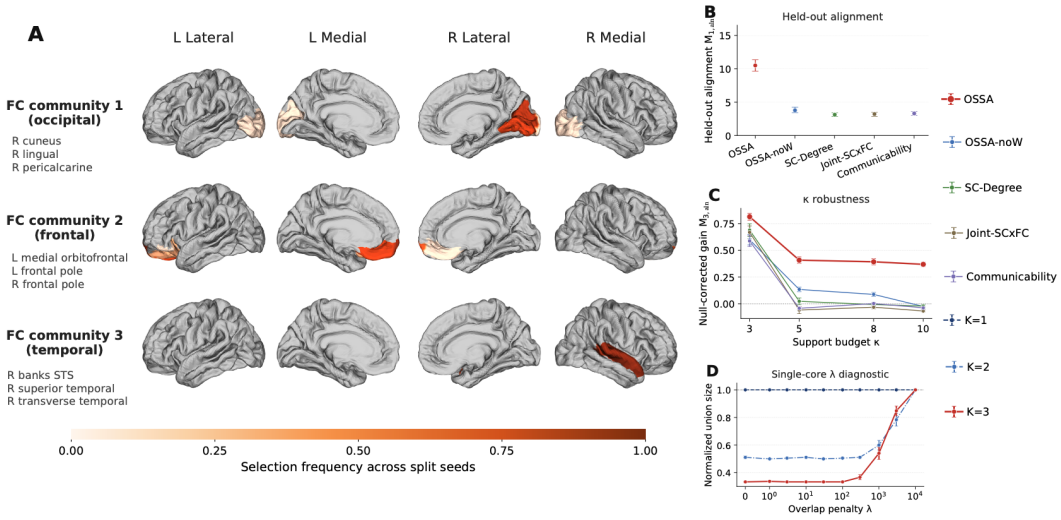


Figure 2: ABCD localization and diagnostics. (A) OSSA selection frequency for the three FC communities at $\kappa = 3$. Labels denote consensus supports. (B) Mean $\pm$ s.e. of held-out rotation-aware alignment $M_{1,\mathrm{aln}}$ for the five methods shown (Table 3). (C) Budget robustness: OSSA maintains positive SC-degree-quartile-matched gain $M_{3,\mathrm{aln}}$. (D) K-support λ diagnostic: supports collapse at small λ and separate only under large penalties.

10 seeds. Community count, sparsity budget, and baseline-side rotations are selected using only training/validation data. The test set is evaluated once after these choices are fixed.

Community partition and budget. Network cross-validation offers an alternative framework for selecting the number of communities under network models [8, 27]. Across all 10 split seeds, we select $Q = 3$ communities by jointly considering eigengap [45], Newman-Girvan modularity [30], and silhouette [35], with partitions stable up to label permutation. This stability is assessed under the same community-construction procedure; the identities of the regions selected by OSSA remain conditional on that procedure (see Appendix D.14). The training-only diagnostic curves are reported in Appendix D.3. The pooled community sizes are $(30, 23, 15)$. We set $\kappa = 3$ in each community, the smallest nontrivial induced SC subgraph after zeroing SC diagonals: $\kappa = 1$ gives zero objective, and $\kappa = 2$ selects one edge. This choice is also supported by the baseline validation curves: averaged across split seeds, alignment per selected node, $J_{\mathrm{val}}(\kappa)/\kappa$, is highest at $\kappa = 3$ among the tested budgets $\kappa \geq 3$ in each community. We also check $\kappa \in \{3, 5, 8, 10\}$ to test sensitivity to this scale (Appendix D.8).

Baselines and metrics. We compare OSSA with fourteen size-matched support-selection baselines and the OSSA-NOW ablation. (1) *Marginal degree* (SC-DEGREE, FC-DEGREE) rank nodes by row sums of the training means of $|B_i|$ and $|A_i|$, respectively. (2) *Cross-modal fusion* (JOINT-SCXFC) ranks row sums of the averaged $|A_i \circ B_i|$. (3) *SC graph summary* (COMMUNICABILITY [14]) uses column sums of $\exp(\bar{B})$. (4) SDI [32] ranks nodes by the high/low-frequency FC energy ratio on the normalized SC Laplacian. (5) *Deep learning* (BRAINGNN [28], BRAINNETCNN [23]) both provide learned node saliencies, with hyperparameters selected by validation loss under the same compute budget as one OSSA run. (6) RANDOM samples a uniform κ-node support. The ablation OSSA-NOW runs Algorithm 1 with $W = I_p$ fixed. We additionally evaluate nonnegative matrix factorization (NMF) [24], sparse partial least squares (PLS) [9], path efficiency, mean first passage time, FAQ graph matching [44], and SC-to-FC ridge prediction as size-matched node-selection baselines. For fairness, each baseline support is refit with its own rotation. For a selected support $\widehat{\mathcal{V}}_{m,q}$ from method m in community q, we hold $s = \mathbf{1}_{\widehat{\mathcal{V}}_{m,q}}$ fixed (protocol detail in Appendix D.9) and fit a method-specific $\widehat{W}_{m,q}$ by running the OSSA W-block on training subjects. Thus the evaluator is the same across methods and only the support changes. Our primary metric is the out-of-sample OSSA objective

evaluated on held-out test data, given by,

$$M_{1,\mathrm{aln}}(\widehat{\mathcal{V}}, W) = \frac{1}{n_{\mathrm{te}}} \sum_{i=1}^{n_{\mathrm{te}}} \mathrm{Tr}\left[\left(W^\top A_i^{\mathrm{te}} W\right)\left(S B_i^{\mathrm{te}} S\right)\right], \qquad S = \mathrm{diag}(\mathbf{1}_{\widehat{\mathcal{V}}}),$$

which directly quantifies the generalization of the alignment objective (1) to unseen subjects. Furthermore, let $\rho(\cdot,\cdot)$ denote the Pearson correlation computed on the upper triangular elements, and let $\mathrm{vec}_>(\cdot)$ represent the corresponding strictly off-diagonal vectorization. We define the scale-free, rotation-aware concordance metric as

$$M_{2,\mathrm{aln}}(\widehat{\mathcal{V}}, W) = \frac{1}{n_{\mathrm{te}}} \sum_i \rho\left(\mathrm{vec}_>\left((W^\top A_i^{\mathrm{te}} W)[\widehat{\mathcal{V}}, \widehat{\mathcal{V}}]\right),\ \mathrm{vec}_>(B_i^{\mathrm{te}}[\widehat{\mathcal{V}}, \widehat{\mathcal{V}}])\right),$$

with $M_{2,\mathrm{raw}} = M_{2,\mathrm{aln}}(\widehat{\mathcal{V}}, I_p)$ as a rotation-unaware control. To test whether a support merely selects structurally dense nodes, we use an SC-degree-quartile-matched null:

$$M_{3,\mathrm{aln}}(\widehat{\mathcal{V}}, W) = M_{2,\mathrm{aln}}(\widehat{\mathcal{V}}, W) - \frac{1}{N} \sum_{r=1}^{N} M_{2,\mathrm{aln}}(\widehat{\mathcal{V}}_r, W_r),$$

where each $\widehat{\mathcal{V}}^{(r)}$ matches the SC-degree-quartile profile of $\widehat{\mathcal{V}}$ within the same community and $\widehat{W}^{(r)}$ is refit by the same training-side W-block. We use $N = 1000$ null supports and report the one-sided permutation p-value as the fraction whose $M_{2,\mathrm{aln}}$ exceeds the observed value.

Canonical OSSA evaluation. At $\kappa = 3$, OSSA outperforms all size-matched baselines on held-out alignment (Table 3; Figure 2b). Across 30 seed-community units, OSSA attains $M_{1,\mathrm{aln}} = 10.55 \pm 0.85$; the closest competitors are NMF (9.96 ± 0.82) and sparse PLS (9.22 ± 1.10) (Tables 3 and 8). OSSA's improvement over OSSA-NOW (3.82 ± 0.42), which fixes $W = I_p$, points to the contribution of the learned orthogonal block. OSSA achieves higher held-out alignment in all 30 paired seed-community comparisons against each baseline in Table 3. These comparisons are descriptive, because the repeated splits reuse subjects from the same cohort. For held-out subjects from one split, OSSA's mean $M_{1,\mathrm{aln}}$ exceeds NMF's by 0.468 (paired subject-bootstrap 95% CI: $[0.400, 0.534]$; Appendix D.9). The recovered supports are compact and stable across split seeds. After aligning community labels to the pooled $Q = 3$ partition, the consensus supports localize to right occipital, fronto-orbitofrontal/frontal-pole, and right temporal regions, respectively (Figure 2a; see Table 4 for SC edge weights). Scale-free metrics suggest that the gain is not just a structural-density effect. In raw coordinates, SC-DEGREE slightly exceeds OSSA on $M_{2,\mathrm{raw}}$ ($+0.674$ vs. $+0.635$), consistent with its preference for dense SC nodes. After each support is evaluated with its own fitted rotation, OSSA leads on $M_{2,\mathrm{aln}}$ ($+0.821$ vs. $+0.695$) and on the SC-degree-quartile-matched gain $M_{3,\mathrm{aln}}$ ($+0.81$ vs. next-best $+0.69$). These aligned metrics quantify the concordance of the selected structural subgraph with the latent functional geometry represented by the fitted transformation (see Appendix D.13 for full empirical diagnostics); the selected SC nodes retain their anatomical identities. The κ sweep in Figure 2c keeps positive null-corrected gain for $\kappa \in \{3, 5, 8, 10\}$, full values are in Appendix D.8.

Structural and alignment interpretation. The occipital and temporal supports include primary visual and auditory regions, consistent with stronger SC–FC correspondence in unimodal cortex [42]. Table 4 reports subject-averaged $\log(1 + x)$ SC edge weights and mean-weight ratios to SC-degree-matched nulls.

For $\widehat{\Pi} = \widehat{W} P_{\widehat{\mathcal{V}}} \widehat{W}^\top$, the community-mean normalized squared distance from $P_{\widehat{\mathcal{V}}}$ is 0.54–0.67, versus 0.80–0.90 for dimension-matched Haar references. Loading participation ratios indicate 5–10 effective FC regions versus 10–19 under Haar. These characterize latent functional geometry; SC node identities remain unchanged. The diagnostics alone do not establish biological specificity (Appendix D.13).

Multi-clique sensitivity (λ-diagnostic). We use the K-support extension as a diagnostic for additional aligned cores. At $\kappa = 3$, the ABCD path shows collapse for $\lambda \le 10^2$ (pairwise Jaccard 0.98, $|U|/\kappa \approx 1$) and separation only under large penalties. Forced separation at $\lambda = 10^4$ reduces the primary held-out objective from 10.42 to 6.42 ($K = 2$) and 4.60 ($K = 3$). Although $M_{2,\mathrm{aln}}$ and $M_{3,\mathrm{aln}}$ gain increase at high λ, these appear to reflect secondary correlation motifs rather than additional high-M_1 cores. These results support using $K > 1$ as a diagnostic rather than the default

Table 3: ABCD canonical OSSA evaluation, mean $\pm$ s.e. across 30 (seed, community) units, $\kappa = 3$. $M_{3,\mathrm{aln}}$ gain is the difference between the method's $M_{2,\mathrm{aln}}$ and the mean of $N = 1000$ SC-degree-quartile-matched random nulls. M_3 p is mean (median) over the 30 units.

Method	$M_{1,\mathrm{aln}}$	$M_{2,\mathrm{raw}}$	$M_{2,\mathrm{aln}}$	$M_{3,\mathrm{aln}}$ gain	M_3 p-value
OSSA	$\mathbf{10.55 \pm 0.85}$	$+0.635$	$\mathbf{+0.821}$	$\mathbf{+0.81}$	$\mathbf{0.080}\,(\mathbf{0.059})$
OSSA-NOW	3.82 ± 0.42	$+0.605$	$+0.605$	$+0.60$	$0.209\,(0.180)$
COMMUNICABILITY	3.31 ± 0.30	$+0.572$	$+0.594$	$+0.59$	$0.190\,(0.139)$
JOINT-SCXFC	3.19 ± 0.32	$+0.653$	$+0.675$	$+0.67$	$0.147\,(0.059)$
SC-DEGREE	3.14 ± 0.32	$\mathbf{+0.674}$	$+0.695$	$+0.69$	$0.138\,(0.028)$
FC-DEGREE	2.72 ± 0.32	$+0.066$	$+0.066$	$+0.06$	$0.433\,(0.598)$
BRAINNETCNN	0.56 ± 0.22	$+0.192$	$+0.283$	$+0.28$	$0.345\,(0.390)$
BRAINGNN	0.49 ± 0.08	$+0.336$	$+0.427$	$+0.42$	$0.275\,(0.196)$
SDI	0.31 ± 0.04	$+0.299$	$+0.389$	$+0.38$	$0.286\,(0.343)$
RANDOM	0.18 ± 0.04	$+0.397$	$+0.386$	$+0.38$	$0.300\,(0.218)$

Table 4: SC connectivity of the consensus supports. Ratios use SC-degree-matched null means.

Community	Three SC edge weights	Mean	SC/null ratio
Occipital	1.75, 2.62, 3.46	2.612	$2.09\times$
Frontal	2.34, 3.11, 3.42	2.958	$1.84\times$
Temporal	0.42, 1.24, 2.37	1.343	$5.95\times$

estimator. Because the FC-derived communities are small, forcing several disjoint κ-node supports is hard to justify unless the added supports retain $M_{1,\mathrm{aln}}$, which is not observed in ABCD. Full trajectories and baseline-extension checks are reported in Appendices D.10-D.11.

Second-visit replication. We apply the same protocol to the ABCD Year-2 follow-up release, without retuning splits, thresholds, budgets, baseline rotations, or the λ grid. After the same QC rule, the follow-up cohort contains $n_{\mathrm{ABCD2}} = 1{,}897$ subjects. The three-community structure is preserved, with cross-seed ARI 0.99 ± 0.01 and pooled sizes $(31, 23, 14)$, close to baseline $(30, 23, 15)$. The canonical ranking is also preserved: OSSA attains $M_{1,\mathrm{aln}} = 10.16 \pm 0.87$, compared with 3.23 ± 0.30 for OSSA-NOW and at most 3.19 for classical baselines. OSSA again achieves higher held-out alignment in all 30 paired seed-community comparisons against each baseline in Table 11. These comparisons are descriptive, because repeated splits reuse subjects from the same follow-up cohort. The λ diagnostic gives the same single-core pattern: collapse at small λ and forced separation only at large λ. The selected supports are also highly reproducible across visits. OSSA attains cross-visit support replication score $\mathrm{CVSR} = 0.928 \pm 0.033$ based on support-set Jaccard overlap, with mean overlap $2.86/3$ nodes across visits. Graph-summary baselines also show high support consistency, but deep-learning baselines replicate less well. We therefore interpret CVSR as evidence of longitudinal support stability, which alone does not establish biological specificity. Appendix E gives per-method results and the follow-up λ sweep.

5 Conclusion and discussion

Overall, OSSA improves the concordance of structural subnetworks with latent functional geometry by jointly estimating localization and cross-modal alignment. On the ABCD dataset, the selected structural supports are compact, stable, and longitudinally reproducible. Furthermore, the multi-support λ solution path reveals that additional supports either collapse onto the primary core or lose cross-modal alignment when forced apart, indicating a single, dominant structural core within these functional communities. Our theoretical guarantees cover canonical OSSA; extending them to the multi-clique setting remains future work. Empirically, the analyses are restricted to the Desikan–Killiany parcellation and the ABCD cohort; the identities of the selected regions are conditional on the functional-community construction and the sparsity budget. Rigorous replication across parcellations, acquisition pipelines, and clinical cohorts is needed before assigning definitive biological significance to these supports.

Acknowledgments and Disclosure of Funding

This work was supported by the National Institutes of Health (NIH) under grants R01AG081413, R01EB034720, and R01AG068191 (to Y. Zhao). T. Li was also supported in part by the National Science Foundation (NSF) under grant DMS2515367. We thank the anonymous reviewers for their constructive feedback, which substantially improved the paper.

References

[1] Farras Abdelnour, Henning U. Voss, and Ashish Raj. Network diffusion accurately models the relationship between structural and functional brain connectivity networks. *NeuroImage*, 90: 335–347, 2014. doi: 10.1016/j.neuroimage.2013.12.039.

[2] P-A Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2008.

[3] Joshua Agterberg, Zachary Lubberts, and Jesús Arroyo. Joint spectral clustering in multilayer degree-corrected stochastic blockmodels. *Journal of the American Statistical Association*, 120 (551):1607–1620, 2025. doi: 10.1080/01621459.2025.2516201.

[4] Graham L. Baum, Zaixu Cui, David R. Roalf, Rastko Ciric, Richard F. Betzel, Bart Larsen, Matthew Cieslak, Philip A. Cook, Cedric H. Xia, Tyler M. Moore, Kosha Ruparel, Desmond J. Oathes, Aaron F. Alexander-Bloch, Russell T. Shinohara, Armin Raznahan, Raquel E. Gur, Ruben C. Gur, Danielle S. Bassett, and Theodore D. Satterthwaite. Development of structure–function coupling in human brain networks during youth. *Proceedings of the National Academy of Sciences of the United States of America*, 117(1):771–778, 2020. doi: 10.1073/pnas.1912034117.

[5] Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, 2009. doi: 10.1038/nrn2575.

[6] B.J. Casey, Tariq Cannonier, May I. Conley, et al. The adolescent brain cognitive development (abcd) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience*, 32: 43–54, 2018. doi: 10.1016/j.dcn.2018.03.001.

[7] Andrew A. Chen, Dhivya Srinivasan, Raymond Pomponio, Yong Fan, Ilya M. Nasrallah, Susan M. Resnick, Lori L. Beason-Held, Christos Davatzikos, Theodore D. Satterthwaite, Dani S. Bassett, Russell T. Shinohara, and Haochang Shou. Harmonizing functional connectivity reduces scanner effects in community detection. *NeuroImage*, 256:119198, 2022. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2022.119198.

[8] Kehui Chen and Jing Lei. Network cross-validation for determining the number of communities in network data. *Journal of the American Statistical Association*, 113(521):241–251, 2018. doi: 10.1080/01621459.2016.1246365.

[9] Hyonho Chun and Sündüz Keleş. Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(1):3–25, 2010. doi: 10.1111/j.1467-9868.2009.00723.x.

[10] R. Cameron Craddock, G. Andrew James, Paul E. Holtzheimer III, Xiaoping P. Hu, and Helen S. Mayberg. A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Human Brain Mapping*, 33(8):1914–1928, 2012. doi: 10.1002/hbm.21333.

[11] Rahul S. Desikan et al. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *NeuroImage*, 31(3):968–980, 2006. doi: 10.1016/j.neuroimage.2006.01.021.

[12] John Duchi, Shai Shalev-Shwartz, Yoram Singer, and Tushar Chandra. Efficient projections onto the ℓ_1-ball for learning in high dimensions. In *Proceedings of the 25th International Conference on Machine Learning*, pages 272–279. ACM, 2008. doi: 10.1145/1390156.1390191.

[13] Alan Edelman, Tomás A Arias, and Steven T Smith. The geometry of algorithms with orthogonality constraints. *SIAM journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.

[14] Ernesto Estrada and Naomichi Hatano. Communicability in complex networks. *Physical Review E*, 77(3):036111, 2008. doi: 10.1103/PhysRevE.77.036111.

[15] Panagiotis Fotiadis, Linden Parkes, Kathryn A. Davis, Theodore D. Satterthwaite, Russell T. Shinohara, and Dani S. Bassett. Structure-function coupling in macroscale human brain networks. *Nature Reviews Neuroscience*, 25(10):688–704, 2024. doi: 10.1038/s41583-024-00846-6.

[16] Karl J. Friston. Functional and effective connectivity in neuroimaging: a synthesis. *Human Brain Mapping*, 2(1-2):56–78, 1994. doi: 10.1002/hbm.460020107.

[17] Simiao Gao, Zhiling Gu, Shengxian Ding, Gefei Wang, Zhengwu Zhang, Hongyu Zhao, and Yize Zhao. Brain functional–structural gradient coupling reflects development, behavior and genetic influences. *Nature Communications*, 2026. doi: 10.1038/s41467-026-71719-y.

[18] Javier O. Garcia, Arian Ashourvan, Sarah Muldoon, Jean M. Vettel, and Danielle S. Bassett. Applications of community detection techniques to brain graphs: Algorithmic considerations and implications for neural function. *Proceedings of the IEEE*, 106(5):846–867, 2018. doi: 10.1109/JPROC.2017.2786710.

[19] Joaquín Goñi, Martijn P. van den Heuvel, Andrea Avena-Koenigsberger, Nieves Vélez de Mendizábal, Richard F. Betzel, Alessandra Griffa, Patric Hagmann, Bernat Corominas-Murtra, Jean-Philippe Thiran, and Olaf Sporns. Resting-brain functional connectivity predicted by analytic measures of network communication. *Proceedings of the National Academy of Sciences of the United States of America*, 111(2):833–838, 2014. doi: 10.1073/pnas.1315529111.

[20] Donald J. Hagler, Jr., Sean N. Hatton, M. Daniela Cornejo, Carolina Makowski, Damien A. Fair, Anthony Steven Dick, Matthew T. Sutherland, B. J. Casey, Deanna M. Barch, Michael P. Harms, Richard Watts, James M. Bjork, Hugh P. Garavan, Anders M. Dale, et al. Image processing and analysis methods for the adolescent brain cognitive development study. *NeuroImage*, 202: 116091, 2019. doi: 10.1016/j.neuroimage.2019.116091.

[21] Nicholas J. Higham. Computing the polar decomposition—with applications. *SIAM Journal on Scientific and Statistical Computing*, 7(4):1160–1174, 1986. doi: 10.1137/0907079.

[22] Christopher J. Honey, Olaf Sporns, Leila Cammoun, Xavier Gigandet, Jean-Philippe Thiran, Reto Meuli, and Patric Hagmann. Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences*, 106(6):2035–2040, 2009. doi: 10.1073/pnas.0811168106.

[23] Jeremy Kawahara, Colin J. Brown, Steven P. Miller, Brian G. Booth, Vann Chau, Ruth E. Grunau, Jill G. Zwicker, and Ghassan Hamarneh. Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*, 146:1038–1049, 2017. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2016.09.046.

[24] Daniel D. Lee and H. Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999. doi: 10.1038/44565.

[25] John M. Lee. *Introduction to Smooth Manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer, New York, 2 edition, 2013. doi: 10.1007/978-1-4419-9982-5.

[26] John M. Lee. *Introduction to Riemannian Manifolds*, volume 176 of *Graduate Texts in Mathematics*. Springer, 2 edition, 2018. doi: 10.1007/978-3-319-91755-9.

[27] Tianxi Li, Elizaveta Levina, and Ji Zhu. Network cross-validation by edge sampling. *Biometrika*, 107(2):257–276, 2020. doi: 10.1093/biomet/asaa006.

[28] Xiaoxiao Li, Yuan Zhou, Nicha Dvornek, Muhan Zhang, Siyuan Gao, Juntang Zhuang, Dustin Scheinost, Lawrence H. Staib, Pamela Ventola, and James S. Duncan. BrainGNN: Interpretable brain graph neural network for fMRI analysis. *Medical Image Analysis*, 74:102233, 2021. doi: 10.1016/j.media.2021.102233.

[29] Daniel S. Margulies, Satrajit S. Ghosh, Alexandros Goulas, Marcel Falkiewicz, Julia M. Huntenburg, Georg Langs, Gleb Bezgin, Simon B. Eickhoff, F. Xavier Castellanos, Michael Petrides, Elizabeth Jefferies, and Jonathan Smallwood. Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences of the United States of America*, 113(44):12574–12579, 2016. doi: 10.1073/pnas.1608282113.

[30] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69:026113, Feb 2004. doi: 10.1103/PhysRevE.69.026113.

[31] Yang Ning and Han Liu. High-dimensional semiparametric bigraphical models. *Biometrika*, 100(3):655–670, 2013.

[32] Maria Giulia Preti and Dimitri Van De Ville. Decoupling of brain function from structure reveals regional behavioral specialization in humans. *Nature Communications*, 10(1):4747, 2019. doi: 10.1038/s41467-019-12765-7.

[33] Reddy Rani Vangimalla and Jaya Sreevalsan-Nair. Communities and cliques in functional brain network using multiscale consensus approach. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30:1951–1960, 2022.

[34] Michael W. Reimann, Max Nolte, Martina Scolamiero, Katharine Turner, Rodrigo Perin, Giuseppe Chindemi, Paweł Dłotko, Ran Levi, Kathryn Hess, and Henry Markram. Cliques of neurons bound into cavities provide a missing link between structure and function. *Frontiers in Computational Neuroscience*, Volume 11 - 2017, 2017. ISSN 1662-5188. doi: 10.3389/fncom.2017.00048.

[35] Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987. doi: 10.1016/0377-0427(87)90125-7.

[36] Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: uses and interpretations. *NeuroImage*, 52(3):1059–1069, 2010. doi: 10.1016/j.neuroimage.2009.10.003.

[37] Peter H. Schönemann. A generalized solution of the orthogonal Procrustes problem. *Psychometrika*, 31(1):1–10, 1966. doi: 10.1007/BF02289451.

[38] Caio Seguin, Ye Tian, and Andrew Zalesky. Network communication models improve the behavioral and functional predictive utility of the human structural connectome. *Network Neuroscience*, 4(4):980–1006, 2020. doi: 10.1162/netn_a_00161.

[39] Ann E. Sizemore, Chad Giusti, Ari Kahn, Jean M. Vettel, Richard F. Betzel, and Danielle S. Bassett. Cliques and cavities in the human connectome. *J. Comput. Neurosci.*, 44(1):115–145, February 2018. ISSN 0929-5313. doi: 10.1007/s10827-017-0672-6.

[40] O. Sporns. Structure and function of complex brain networks. *Dialogues in Clinical Neuroscience*, 15(3):247–262, 2013. doi: 10.31887/DCNS.2013.15.3/osporns.

[41] Laura E. Suárez, Ross D. Markello, Richard F. Betzel, and Bratislav Mišić. Linking structure and function in macroscale brain networks. *Trends in Cognitive Sciences*, 24(4):302–315, 2020. doi: 10.1016/j.tics.2020.01.008.

[42] Bertha Vázquez-Rodríguez, Laura E. Suárez, Ross D. Markello, Golia Shafiei, Casey Paquola, Patric Hagmann, Martijn P. van den Heuvel, Boris C. Bernhardt, R. Nathan Spreng, and Bratislav Mišić. Gradients of structure–function tethering across neocortex. *Proceedings of the National Academy of Sciences of the United States of America*, 116(42):21219–21227, 2019. doi: 10.1073/pnas.1903403116.

[43] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2018. ISBN 9781108415194. doi: 10.1017/9781108231596.

[44] Joshua T. Vogelstein, John M. Conroy, Vince Lyzinski, Louis J. Podrazik, Steven G. Kratzer, Eric T. Harley, Donniell E. Fishkind, R. Jacob Vogelstein, and Carey E. Priebe. Fast approximate quadratic programming for graph matching. *PLOS ONE*, 10(4):e0121002, 2015. doi: 10.1371/journal.pone.0121002.

[45] Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007. doi: 10.1007/s11222-007-9033-z.

[46] Reinder Vos de Wael, Oualid Benkarim, Casey Paquola, Sara Lariviere, Jessica Royer, Shahin Tavakol, Ting Xu, Seok-Jun Hong, Georg Langs, Sofie Valk, Bratislav Misic, Michael Milham, Daniel Margulies, Jonathan Smallwood, and Boris C. Bernhardt. Brainspace: A toolbox for the analysis of macroscale gradients in neuroimaging and connectomics datasets. *Communications Biology*, 3(1):103, 2020. doi: 10.1038/s42003-020-0794-7.

[47] Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge university press, 2019.

[48] Weiran Wang and Canyi Lu. Projection onto the capped simplex. *arXiv preprint arXiv:1503.01002*, 2015. doi: 10.48550/arXiv.1503.01002.

[49] B. T. Thomas Yeo, Fenna M. Krienen, Jorge Sepulcre, Mert R. Sabuncu, Danial Lashkari, Marisa Hollinshead, Joshua L. Roffman, Jordan W. Smoller, Lilla Zöllei, Jonathan R. Polimeni, Bruce Fischl, Hesheng Liu, and Randy L. Buckner. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 106(3): 1125–1165, 2011. doi: 10.1152/jn.00338.2011.

[50] Andrew Zalesky, Alex Fornito, and Edward T. Bullmore. Network-based statistic: identifying differences in brain networks. *NeuroImage*, 53(4):1197–1207, 2010. doi: 10.1016/j.neuroimage.2010.06.041.

[51] Farnaz Zamani Esfahlani, Joshua Faskowitz, Jonah Slack, Bratislav Mišić, and Richard F. Betzel. Local structure–function relationships in human brain networks across the lifespan. *Nature Communications*, 13(1):2053, 2022. doi: 10.1038/s41467-022-29770-y.

A Optimization details and derivations

This appendix gives the first-order quantities used in Algorithms 1 and 2. All derivations assume the matrices have been symmetrized. In implementation we use $X \leftarrow (X + X^\top)/2$ for numerical stability and ignore diagonal entries when evaluating inter-regional connectivity.

A.1 Gradient for the canonical OSSA clique W-block

Fix s and define $M_i = SB_iS$. For one subject,

$$f_i(W) = \mathrm{Tr}(W^\top A_i W M_i).$$

Taking differentials,

$$\begin{aligned} df_i &= \mathrm{Tr}(dW^\top A_i W M_i) + \mathrm{Tr}(W^\top A_i dW M_i) \\ &= \mathrm{Tr}\big((A_i W M_i)^\top dW\big) + \mathrm{Tr}\big((A_i^\top W M_i^\top)^\top dW\big). \end{aligned}$$

Therefore

$$\nabla_W f_i(W) = A_i W M_i + A_i^\top W M_i^\top.$$

When A_i and M_i are symmetric, this simplifies to

$$\nabla_W J(W, s) = 2 \sum_{i=1}^{n} A_i W (SB_iS). \tag{6}$$

A Euclidean ascent proposal $\widetilde{W} = W + \eta \nabla_W J$ is mapped back to the orthogonal group by the polar projection [21]

$$\Pi_{O(p)}(\widetilde{W}) = UV^\top, \qquad \widetilde{W} = U\Sigma V^\top, \tag{7}$$

which is the closest orthogonal matrix to $\widetilde{W}$ in Frobenius norm.

A.2 Reduction and gradient for the canonical OSSA clique s-block

Fix W and set $Q_i = W^\top A_i W$. Since $S = \mathrm{diag}(s)$,

$$(SB_iS)_{ab} = s_a (B_i)_{ab} s_b, \qquad SB_iS = (ss^\top) \circ B_i.$$

Using symmetry of Q_i and B_i,

$$\mathrm{Tr}(Q_i SB_i S) = \sum_{a,b} (Q_i)_{ab} (B_i)_{ab} s_a s_b = s^\top (Q_i \circ B_i) s.$$

Thus, with

$$D(W) = \sum_{i=1}^{n} (Q_i \circ B_i),$$

the fixed-W objective is $s^\top D(W) s$, and

$$\nabla_s J(W, s) = 2 D(W) s. \tag{8}$$

A.3 Projection onto the capped simplex

The support update requires Euclidean projection [48, 12] onto

$$\mathcal{S}_\kappa = \{s \in \mathbb{R}^p : 0 \le s \le 1,\ \mathbf{1}^\top s \le \kappa\}.$$

For a vector $y \in \mathbb{R}^p$, the Euclidean projection onto $\mathcal{S}_\kappa^=$ has the threshold form

$$[\Pi_{\mathcal{S}_\kappa^=}(y)]_j = \min\{1, \max\{0, y_j - \tau\}\},$$

where τ is chosen so that $\sum_j [\Pi_{\mathcal{S}_\kappa^=}(y)]_j = \kappa$. The scalar equation is monotone in τ and can be solved by bisection or by the standard sorting-based capped-simplex projection.

A.4 Multi-clique gradients

For the multi-clique objective (5), fix $\{s^{(\ell)}\}_{\ell=1}^{K}$ and define

$$M_i = \sum_{\ell=1}^{K} S^{(\ell)} B_i S^{(\ell)}.$$

The overlap penalty is constant in W, so the same calculation as Appendix A.1 gives

$$\nabla_W J_K(W, \{s^{(\ell)}\}) = 2 \sum_{i=1}^{n} A_i W M_i. \tag{9}$$

For fixed W, the objective is

$$J_K(\{s^{(\ell)}\} \mid W) = \sum_{\ell=1}^{K} {s^{(\ell)}}^{\top} D(W) s^{(\ell)} - \lambda \sum_{\ell<\ell'} \left\langle s^{(\ell)},\, s^{(\ell')} \right\rangle.$$

Holding all other supports fixed, the gradient with respect to $s^{(\ell)}$ is

$$\nabla_{s^{(\ell)}} J_K = 2 D(W) s^{(\ell)} - \lambda \sum_{r \neq \ell} s^{(r)}. \tag{10}$$

Each support is projected onto its own capped simplex $\mathcal{S}_\kappa$.

A.5 Monotonicity and nonconvexity

Algorithms 1 and 2 use backtracking steps that accept only updates that do not decrease the objective. Since $O(p) \times \mathcal{S}_\kappa$ and $O(p) \times \mathcal{S}_\kappa^K$ are compact, the corresponding objective sequences are bounded and therefore converge in value. The optimization problems remain nonconvex because of the orthogonality constraint and the quadratic support variables. We therefore treat the algorithms as first-order local solvers. In all experiments, we use multiple random restarts and retain the run with the largest training objective. This monotonicity and boundedness imply convergence of the objective values, but do not by themselves establish convergence of the iterates, vanishing stationarity residuals, or convergence to a global maximizer. The objective and iterate change tolerances are numerical stopping criteria and do not by themselves imply the residual bounds required in theory below.

B Technical details

For $W \in O(p)$, define

$$D_n(W) = \frac{1}{n} \sum_{i=1}^{n} \left\{ (W^{\top} A_i W) \circ B_i \right\}, \qquad D_0(W) = \mathbb{E} D_n(W). \tag{11}$$

Then

$$\bar{J}_n(W, s) = s^{\top} D_n(W) s, \qquad \mathcal{J}(W, s) = s^{\top} D_0(W) s. \tag{12}$$

For a vector v, write $\|v\|_1 = \sum_j |v_j|$, $\|v\|_2 = (\sum_j v_j^2)^{1/2}$, and $\|v\|_\infty = \max_j |v_j|$. For a matrix M, let $\|M\|_F = \{\mathrm{Tr}(M^{\top} M)\}^{1/2}$ and $\|M\|_{\mathrm{op}} = \sup_{\|v\|_2 = 1} \|Mv\|_2$. For a real random variable Z and $r \in \{1, 2\}$, define

$$\|Z\|_{\psi_r} = \inf \left\{ t > 0 : \mathbb{E} \exp\left(\frac{|Z|^r}{t^r} \right) \leq 2 \right\}. \tag{13}$$

The cases $r = 1$ and $r = 2$ are the sub-exponential and sub-Gaussian norms, respectively [47].

For matrices X, Y of the same size, write $\langle X,\, Y \rangle = \mathrm{Tr}(X^{\top} Y)$. We equip $O(p)$ with the Riemannian metric induced by this Frobenius inner product. For W, U in the same connected component, define

$$d_O(W, U) = \inf_{\substack{\gamma:[0,1] \to O(p) \text{ piecewise } C^1 \\ \gamma(0)=U,\ \gamma(1)=W}} \int_0^1 \|\dot{\gamma}(t)\|_F \, dt. \tag{14}$$

Set $d_O(W,U) = +\infty$ for points in different components. For every nonempty $\mathcal{W} \subset O(p)$, define

$$d_O(W,\mathcal{W}) = \inf_{U\in\mathcal{W}} d_O(W,U), \qquad d_V(W,\mathcal{W}) = \inf_{U\in\mathcal{W}} \|(W-U)P_V\|_F. \tag{15}$$

Differentiating $W^\top W = I_p$ along a differentiable curve gives

$$T_W O(p) = \left\{\xi \in \mathbb{R}^{p\times p} : W^\top\xi + \xi^\top W = 0\right\} = \left\{W\Omega : \Omega^\top = -\Omega\right\}. \tag{16}$$

For $U \in O(p)$ and $\xi \in T_U O(p)$, the Riemannian exponential for this metric is

$$\mathrm{Exp}_U(\xi) = U\exp(U^\top\xi), \qquad \exp(M) = \sum_{k=0}^{\infty}\frac{M^k}{k!}. \tag{17}$$

The curve $t \mapsto \mathrm{Exp}_U(t\xi)$ is the geodesic with initial position U and initial velocity ξ; see Edelman et al. [13, Section 2.2.2, Eq. (2.14)]. For $Z \in \mathbb{R}^{p\times p}$, note that

$$Z = \left\{Z - W\,\mathrm{sym}(W^\top Z)\right\} + W\,\mathrm{sym}(W^\top Z), \tag{18}$$

where $\mathrm{sym}(M) = (M+M^\top)/2$. The first term in (18) belongs to $T_W O(p)$, the second belongs to its orthogonal complement, and the two terms are orthogonal. Hence

$$\mathcal{P}_W(Z) = Z - W\,\mathrm{sym}(W^\top Z) \tag{19}$$

is the orthogonal projection onto $T_W O(p)$ (under the stated inner product) and satisfies

$$\|\mathcal{P}_W(Z)\|_F \le \|Z\|_F. \tag{20}$$

We remark that the disaplys (16) and (19) are the orthogonal-group specializations of Absil et al. [2, Examples 3.5.3 and 3.6.2]; see also Edelman et al. [13, Section 2.2.1].

For a differentiable function F on $O(p)$, $\mathrm{grad}_W F$ is the unique tangent vector satisfying

$$\langle \mathrm{grad}_W F,\, \xi\rangle = DF(W)[\xi], \qquad \xi \in T_W O(p). \tag{21}$$

If $\nabla_W F$ is the Euclidean gradient of any differentiable ambient extension of F, then

$$\mathrm{grad}_W F = \mathcal{P}_W(\nabla_W F) = \nabla_W F - W\,\mathrm{sym}(W^\top\nabla_W F). \tag{22}$$

The projected expression is independent of the chosen extension. For twice continuously differentiable F, the Riemannian Hessian is given by

$$\mathrm{Hess}_{WW} F(W)[\xi] = \mathcal{P}_W\left\{\frac{d}{dt}\,\mathrm{grad}_W F(\gamma(t))\Big|_{t=0}\right\}, \tag{23}$$

where γ is any smooth curve in $O(p)$ with $\gamma(0) = W$ and $\dot\gamma(0) = \xi$. This is the Riemannian covariant derivative of the gradient, which is further illustrated in Section 5 of Absil et al. [2].

We use the following conditions at the target $u^\star$.

Assumption A1. The pairs (A_i, B_i), $i = 1,\dots,n$, are independent and identically distributed, $A_i = A_i^\top$, $B_i = B_i^\top$, and $\mathrm{diag}(B_i) = 0$ almost surely. Moreover, $2 \le \kappa < p$.

Assumption A2. There exists $\nu_n > 0$ such that, for every $a, b \in \{1,\dots,p\}$ and every pair of unit vectors $x, y \in \mathbb{R}^p$,

$$\|B_{i,ab}x^\top A_i y\|_{\psi_1} \le \nu_n. \tag{24}$$

Assumption A3. There are a nonempty compact smooth embedded submanifold $\mathcal{W}^\star \subset O(p)$ and $J^\star \in \mathbb{R}$ such that

$$\mathcal{J}(U,u^\star) = J^\star, \qquad \mathrm{grad}_W\,\mathcal{J}(U,u^\star) = 0, \qquad U \in \mathcal{W}^\star. \tag{25}$$

There is also a rank-κ projector $\Pi^\star$ satisfying

$$UP_{V^\star}U^\top = \Pi^\star, \qquad U \in \mathcal{W}^\star. \tag{26}$$

The set contains every unselected-column completion: if $U \in \mathcal{W}^\star$, $R \in O(p)$, and $RP_{V^\star} = P_{V^\star}$, then $UR \in \mathcal{W}^\star$.

Assumption A4. There exists $\alpha_n > 0$ such that, for every $U \in \mathcal{W}^\star$ and every $\xi \in T_U O(p)$ orthogonal to $T_U \mathcal{W}^\star$,

$$-\langle \xi, \operatorname{Hess}_{WW} \mathcal{J}(U, u^\star)[\xi] \rangle \geq \alpha_n \|\xi\|_F^2. \tag{27}$$

Assumption A5. There exists $m_n > 0$ such that

$$\inf_{U \in \mathcal{W}^\star} \left[\min_{a \in V^\star} \{2D_0(U)u^\star\}_a - \max\left\{0, \max_{b \notin V^\star} \{2D_0(U)u^\star\}_b\right\}\right] \geq m_n. \tag{28}$$

The diagonal restriction in A1 applies only to B_i. In particular,

$$\{D_n(W)\}_{aa} = \{D_0(W)\}_{aa} = 0, \qquad a = 1, \ldots, p, \tag{29}$$

whereas a zero diagonal of A_i would not generally be preserved by the transformation $W^\top A_i W$.

The completion requirement in A3 ensures that the distance to $\mathcal{W}^\star$ is finite in both connected components of $O(p)$. Write e_j for the jth standard basis vector of $\mathbb{R}^p$. Choose $j \notin V^\star$ and set

$$R_j = I_p - 2e_j e_j^\top. \tag{30}$$

Then $R_j P_{V^\star} = P_{V^\star} R_j = P_{V^\star}$ and $\det(R_j) = -1$. Hence, for every $U \in \mathcal{W}^\star$,

$$(UR_j)P_{V^\star} = UP_{V^\star}, \qquad (UR_j)P_{V^\star}(UR_j)^\top = \Pi^\star, \tag{31}$$

but $\det(UR_j) = -\det(U)$. Thus $\mathcal{W}^\star$ intersects both components. For every $W \in O(p)$,

$$d_O(W, \mathcal{W}^\star) = \inf_{\substack{U \in \mathcal{W}^\star \\ \det(U) = \det(W)}} d_O(W, U), \tag{32}$$

and the infimum is finite and attained. If W_1 and W_2 lie in the same component, the triangle inequality gives

$$|d_O(W_1, \mathcal{W}^\star) - d_O(W_2, \mathcal{W}^\star)| \leq d_O(W_1, W_2). \tag{33}$$

Consequently, $W \mapsto d_O(W, \mathcal{W}^\star)$ is continuous on each component and hence on $O(p)$, because the two components are both open and closed.

Condition A4 permits directions tangent to $\mathcal{W}^\star$ to remain unidentified and requires curvature only in directions normal to the stationary manifold $\mathcal{W}^\star$. Condition A5 is the corresponding strict first-order requirement for the support block. To make the support requirement explicit, fix $U \in \mathcal{W}^\star$ and write

$$g(U) = \nabla_s \mathcal{J}(U, u^\star) = 2D_0(U)u^\star. \tag{34}$$

Equations (29) and (34) imply

$$g_a(U) = 2 \sum_{\ell \in V^\star \setminus \{a\}} \{D_0(U)\}_{a\ell}, \qquad a \in V^\star, \tag{35}$$

$$g_b(U) = 2 \sum_{\ell \in V^\star} \{D_0(U)\}_{b\ell}, \qquad b \notin V^\star, \tag{36}$$

where

$$\{D_0(U)\}_{a\ell} = \mathbb{E}\left[B_{i,a\ell}(Ue_a)^\top A_i (Ue_\ell)\right]. \tag{37}$$

Thus (28) compares every selected (active) coordinate's aggregate within-support SC–FC alignment with both zero and every outside (inactive) coordinate's aggregate alignment.

A vector d is a feasible first-order direction from $u^\star$ if and only if

$$d_a \leq 0 \quad (a \in V^\star), \qquad d_b \geq 0 \quad (b \notin V^\star), \qquad \mathbf{1}^\top d \leq 0. \tag{38}$$

The directions $-e_a$ and $e_b - e_a$ remove active mass and replace active mass by inactive mass, respectively; their directional derivatives are $-g_a(U)$ and $g_b(U) - g_a(U)$. The comparison with zero in (28) is therefore essential since (2) permits active mass to be removed without being reallocated.

More generally, for d satisfying (38), set

$$A_d = -\sum_{a \in V^\star} d_a, \qquad B_d = \sum_{b \notin V^\star} d_b, \qquad c(U) = \max\left\{0, \max_{b \notin V^\star} g_b(U)\right\}. \tag{39}$$

Then $0 \le B_d \le A_d$, and (28) implies

$$\begin{aligned}\langle g(U),\, d\rangle &= -\sum_{a \in V^\star} g_a(U)(-d_a) + \sum_{b \notin V^\star} g_b(U) d_b \\ &\le -\{c(U) + m_n\} A_d + c(U) B_d \le -m_n A_d \le -\frac{m_n}{2}\|d\|_1.\end{aligned} \tag{40}$$

Hence the directional derivative is uniformly negative along every nonzero feasible direction.

For $x > 0$, define

$$R_n(x) = 2\log p + p\log 9 + x, \qquad \eta_n(x) = C\nu_n\left\{\sqrt{\frac{R_n(x)}{n}} + \frac{R_n(x)}{n}\right\}, \tag{41}$$

where $C > 0$ is a universal constant. For $a, b \in \{1, \dots, p\}$, let

$$H_{ab} = \frac{1}{n}\sum_{i=1}^{n} B_{i,ab} A_i - \mathbb{E}(B_{i,ab} A_i), \tag{42}$$

and define

$$\mathcal{G}_x = \left\{\max_{a,b} \|H_{ab}\|_{\mathrm{op}} \le \eta_n(x)\right\}. \tag{43}$$

Lemma 1 bounds $\max_{a,b} \|H_{ab}\|_{\mathrm{op}}$ and gives the probability of $\mathcal{G}_x$.

Lemma 1. *Under A1–A2, the universal constant in* (41) *can be chosen so that*

$$\mathbb{P}(\mathcal{G}_x) \ge 1 - 2e^{-x}. \tag{44}$$

Proof. Fix a, b and a unit vector $z \in \mathbb{R}^p$. The variables

$$Y_i = B_{i,ab} z^\top A_i z - \mathbb{E}(B_{i,ab} z^\top A_i z) \tag{45}$$

are independent and centered. By (24), Jensen's inequality, and the triangle inequality for the ψ_1 norm, $\|Y_i\|_{\psi_1} \le C_0\nu_n$ for a universal constant C_0. The sub-exponential Bernstein inequality [43] therefore yields

$$\mathbb{P}\big(|z^\top H_{ab} z| > t\big) \le 2\exp\left[-cn\min\left\{\frac{t^2}{\nu_n^2}, \frac{t}{\nu_n}\right\}\right], \qquad t > 0, \tag{46}$$

for a universal constant $c > 0$.

Let $\mathcal{N}$ be a $1/4$-net of the unit sphere in $\mathbb{R}^p$ with $|\mathcal{N}| \le 9^p$. Because H_{ab} is symmetric,

$$\|H_{ab}\|_{\mathrm{op}} \le 2\max_{z \in \mathcal{N}} |z^\top H_{ab} z|. \tag{47}$$

Indeed, for every unit vector v, choose $z \in \mathcal{N}$ such that $\|v - z\|_2 \le 1/4$. Then

$$|v^\top H_{ab} v| \le |z^\top H_{ab} z| + \frac{1}{2}\|H_{ab}\|_{\mathrm{op}}. \tag{48}$$

A union bound over at most $p^2 9^p$ pairs and net points yields

$$\mathbb{P}\left(\max_{a,b} \|H_{ab}\|_{\mathrm{op}} > 2t\right) \le 2\exp\left[2\log p + p\log 9 - cn\min\left\{\frac{t^2}{\nu_n^2}, \frac{t}{\nu_n}\right\}\right]. \tag{49}$$

Taking $t = C_1\nu_n\{\sqrt{R_n(x)/n} + R_n(x)/n\}$ with C_1 sufficiently large makes the exponent in (49) at most $-x$. Absorbing the fixed constants into C in (41) proves (44). □

Define

$$M_0 = \max_{a,b} \|\mathbb{E}(B_{i,ab}A_i)\|_{\mathrm{op}}. \tag{50}$$

This quantity is finite under A2. Indeed, symmetry and (24) imply

$$\|\mathbb{E}(B_{i,ab}A_i)\|_{\mathrm{op}} = \sup_{\|z\|_2=1} \left|\mathbb{E}(B_{i,ab}z^\top A_i z)\right| \le C_0\nu_n. \tag{51}$$

The possible dependence of M_0 on n and p is suppressed.

Lemma 2 bounds $(\bar{J}_n - \mathcal{J})(W,s)$ and $\nabla_s(\bar{J}_n - \mathcal{J})(W,s)$ uniformly on $O(p) \times \mathcal{S}_\kappa$ using $\mathcal{G}_x$.

Lemma 2. *Under A1–A2, on $\mathcal{G}_x$, uniformly over $W \in O(p)$ and $s \in \mathcal{S}_\kappa$,*

$$\text{(i)} \quad \left|(\bar{J}_n - \mathcal{J})(W,s)\right| \le \kappa^2\eta_n(x), \tag{52}$$

$$\text{(ii)} \quad \left\|\nabla_s(\bar{J}_n - \mathcal{J})(W,s)\right\|_\infty \le 2\kappa\eta_n(x). \tag{53}$$

Proof. Write w_a for column a of W. The symmetry in A1 gives $H_{ab} = H_{ba} = H_{ab}^\top$. Expanding $\bar{J}_n$ and (11) yields

$$(\bar{J}_n - \mathcal{J})(W,s) = \sum_{a,b=1}^{p} s_a s_b w_a^\top H_{ab} w_b. \tag{54}$$

Since $s \ge 0$, $\mathbf{1}^\top s \le \kappa$, and $\|w_a\|_2 = 1$, the event (43) implies

$$\left|(\bar{J}_n - \mathcal{J})(W,s)\right| \le \eta_n(x) \sum_{a,b=1}^{p} s_a s_b = \eta_n(x)(\mathbf{1}^\top s)^2 \le \kappa^2\eta_n(x), \tag{55}$$

which is (52). The ath coordinate of the support-gradient difference is

$$2\sum_{b=1}^{p} s_b w_a^\top H_{ab} w_b. \tag{56}$$

Equation (43) bounds the absolute value of (56) by $2\eta_n(x)\mathbf{1}^\top s \le 2\kappa\eta_n(x)$, proving (53). □

Lemma 3 bounds $\mathrm{grad}_W(\bar{J}_n - \mathcal{J})(W,u^\star)$ and the change in $\mathrm{grad}_W \bar{J}_n(W,s)$ when s is replaced by $u^\star$.

Lemma 3. *Under A1–A2, on $\mathcal{G}_x$, uniformly over $W \in O(p)$ and $s \in \mathcal{S}_\kappa$,*

$$\text{(i)} \quad \left\|\mathrm{grad}_W(\bar{J}_n - \mathcal{J})(W,u^\star)\right\|_F \le 2(\kappa-1)\sqrt{\kappa}\,\eta_n(x), \tag{57}$$

$$\text{(ii)} \quad \left\|\mathrm{grad}_W \bar{J}_n(W,s) - \mathrm{grad}_W \bar{J}_n(W,u^\star)\right\|_F \le 2(\kappa+\sqrt{\kappa})\{M_0 + \eta_n(x)\}\|s-u^\star\|_1. \tag{58}$$

Proof. Write w_a for column a of W. The Euclidean W-gradient of the stochastic difference at $u^\star$ has column a

$$2u_a^\star \sum_{b \in V^\star} H_{ab} w_b. \tag{59}$$

Equations A1 and (42) give $H_{aa} = 0$. Hence each of the at most κ possibly nonzero columns in (59) has norm at most $2(\kappa-1)\eta_n(x)$, and the Euclidean gradient has Frobenius norm at most $2(\kappa-1)\sqrt{\kappa}\,\eta_n(x)$. Applying (20) proves (57).

For (58), let

$$\widehat{M}_{ab} = \frac{1}{n}\sum_{i=1}^{n} B_{i,ab}A_i, \qquad h = s - u^\star. \tag{60}$$

Equations (43) and (50) imply

$$\max_{a,b} \|\widehat{M}_{ab}\|_{\mathrm{op}} \le M_0 + \eta_n(x). \tag{61}$$

The difference between column a of the two Euclidean gradients is

$$2h_a \sum_{b=1}^{p} s_b \widehat{M}_{ab} w_b + 2u_a^\star \sum_{b=1}^{p} h_b \widehat{M}_{ab} w_b. \tag{62}$$

By (61), the matrices formed by the first and second terms in (62) have Frobenius norms at most

$$2\kappa\{M_0 + \eta_n(x)\}\|h\|_2 \qquad \text{and} \qquad 2\sqrt{\kappa}\{M_0 + \eta_n(x)\}\|h\|_1, \tag{63}$$

respectively. Since $\|h\|_2 \le \|h\|_1$, the Euclidean-gradient difference is bounded by the right-hand side of (58). Both gradients are evaluated at the same W, so (22) applies the same linear map $\mathcal{P}_W$ to both. Equation (20) therefore proves (58) for the Riemannian gradients. □

B.1 Bounds near $\mathcal{W}^\star \times \{u^\star\}$

Lemma 4 relates the distance to $\mathcal{W}^\star$ to the population gradient norm and objective loss through (64) and (65) below.

Lemma 4. *Under Assumptions A1–A4, there exists $\rho_{2,n} > 0$ such that, for every $W \in O(p)$, $d_O(W, \mathcal{W}^\star) \le \rho_{2,n}$ implies*

$$\text{(i)} \quad \|\mathrm{grad}_W \mathcal{J}(W, u^\star)\|_F \ge \frac{\alpha_n}{2} d_O(W, \mathcal{W}^\star), \tag{64}$$

$$\text{(ii)} \quad J^\star - \mathcal{J}(W, u^\star) \ge \frac{\alpha_n}{4} d_O(W, \mathcal{W}^\star)^2. \tag{65}$$

Proof. Equation (24) implies that every coefficient defining $\mathcal{J}$ is finite. Since (12) is a finite polynomial in the entries of W and s, the map $(W, s) \mapsto \mathcal{J}(W, s)$ is smooth. By Lee [26, Theorem 5.25], the tubular-neighborhood theorem applied within each connected component of $O(p)$ gives a radius $\rho_{1,n} > 0$ on which the normal exponential map around $\mathcal{W}^\star$ is one-to-one. By Lee [26, Theorem 6.40], the radius can be chosen so that every W satisfying $d_O(W, \mathcal{W}^\star) < \rho_{1,n}$ has the unique representation

$$\begin{aligned} W &= \mathrm{Exp}_{U(W)}(\xi_W), \\ U(W) &\in \mathcal{W}^\star, \\ \xi_W &\in T_{U(W)}O(p), \qquad \xi_W \perp T_{U(W)}\mathcal{W}^\star, \\ \|\xi_W\|_F &= d_O(W, \mathcal{W}^\star). \end{aligned} \tag{66}$$

The same metric tubular-neighborhood result also gives the converse distance identity

$$d_O(\mathrm{Exp}_U(\zeta), \mathcal{W}^\star) = \|\zeta\|_F \tag{67}$$

for every $U \in \mathcal{W}^\star$ and every $\zeta \in T_U O(p)$ satisfying $\zeta \perp T_U\mathcal{W}^\star$ and $\|\zeta\|_F < \rho_{1,n}$. The two componentwise radii can be replaced by their minimum because $\mathcal{W}^\star$ intersects both components by (31).

We next extend the curvature in (27) uniformly to nearby points. Consider

$$\mathcal{K} = \left\{(U, \nu) \in \mathcal{W}^\star \times \mathbb{R}^{p\times p} : \nu \in T_U O(p),\ \nu \perp T_U\mathcal{W}^\star,\ \|\nu\|_F = 1\right\}. \tag{68}$$

The set $\mathcal{K}$ is compact. Indeed, take any sequence (U_j, ν_j) in $\mathcal{K}$. Compactness of $\mathcal{W}^\star$ and of the Euclidean unit sphere yields a subsequence converging to some (U, ν). The identities $U_j^\top \nu_j + \nu_j^\top U_j = 0$ pass to the limit, so $\nu \in T_U O(p)$ by (16). A smooth local parametrization of $\mathcal{W}^\star$ around U allows every vector in $T_U\mathcal{W}^\star$ to be approximated by vectors in $T_{U_j}\mathcal{W}^\star$. Passing to the limit in the corresponding inner products gives $\nu \perp T_U\mathcal{W}^\star$. Thus $(U, \nu) \in \mathcal{K}$, so $\mathcal{K}$ is a closed subset of a compact set.

For $(U, \nu) \in \mathcal{K}$, define

$$\gamma_{U,\nu}(r) = \mathrm{Exp}_U(r\nu), \qquad \zeta_{U,\nu}(r) = \frac{d}{dr}\gamma_{U,\nu}(r). \tag{69}$$

Lee [26, Proposition 5.19(b)–(c)] implies that (69) is a geodesic and depends smoothly on (U, ν, r). Therefore

$$\Phi(U, \nu, r) = \langle \zeta_{U,\nu}(r),\, \mathrm{Hess}_{WW}\, \mathcal{J}(\gamma_{U,\nu}(r), u^\star)[\zeta_{U,\nu}(r)] \rangle \tag{70}$$

is continuous. At $r = 0$, $\zeta_{U,\nu}(0) = \nu$, and (27) gives $\Phi(U, \nu, 0) \le -\alpha_n$ on $\mathcal{K}$. Compactness of (68) therefore yields a radius satisfying

$$0 < \rho_{2,n} < \rho_{1,n} \tag{71}$$

such that

$$\Phi(U, \nu, r) \le -\frac{\alpha_n}{2}, \qquad (U, \nu) \in \mathcal{K}, \quad 0 \le r \le \rho_{2,n}. \tag{72}$$

Fix W with $d_O(W, \mathcal{W}^\star) \le \rho_{2,n}$ and use (66). If $\xi_W \neq 0$, let

$$\nu_W = \frac{\xi_W}{\|\xi_W\|_F}, \qquad \gamma_W(t) = \mathrm{Exp}_{U(W)}(t\xi_W), \qquad q_W(t) = \mathcal{J}(\gamma_W(t), u^\star). \tag{73}$$

Then

$$\dot\gamma_W(t) = \|\xi_W\|_F \zeta_{U(W),\nu_W}(t\|\xi_W\|_F), \qquad \|\dot\gamma_W(t)\|_F = \|\xi_W\|_F. \tag{74}$$

Since γ_W is a geodesic, Absil et al. [2, Definition 5.5.1 and Proposition 5.5.2] yields

$$\begin{aligned} q_W''(t) &= \langle \dot\gamma_W(t), \mathrm{Hess}_{WW}\, \mathcal{J}(\gamma_W(t), u^\star)[\dot\gamma_W(t)] \rangle \\ &\le -\frac{\alpha_n}{2}\|\xi_W\|_F^2, \qquad 0 \le t \le 1, \end{aligned} \tag{75}$$

where the inequality follows from (72) and (74). Equation (25) gives

$$q_W'(0) = \langle \mathrm{grad}_W\, \mathcal{J}(U(W), u^\star),\, \xi_W \rangle = 0. \tag{76}$$

Integrating (75) once and using (74) yields

$$\frac{\alpha_n}{2}\|\xi_W\|_F^2 \le -q_W'(1) \le \|\mathrm{grad}_W\, \mathcal{J}(W, u^\star)\|_F\, \|\xi_W\|_F. \tag{77}$$

Equations (66) and (77) prove (64). Taylor's expansion, together with (76) and (75), yields

$$\mathcal{J}(W, u^\star) = J^\star + \int_0^1 (1-t) q_W''(t)\, dt \le J^\star - \frac{\alpha_n}{4}\|\xi_W\|_F^2. \tag{78}$$

Equations (66) and (78) prove (65). □

Lemma 5 shows that the selected-coordinate derivatives of $\mathcal{J}(W, s)$ remain positive and uniformly exceed the unselected-coordinate derivatives throughout (79) below.

Lemma 5. *Under Assumptions A1–A5, there is $r_0 \in (0, 1)$, possibly depending on n and p, such that every $(W, s) \in O(p) \times \mathcal{S}_\kappa$ satisfying*

$$d_O(W, \mathcal{W}^\star) \le r_0, \qquad \|s - u^\star\|_2 \le r_0 \tag{79}$$

also satisfies

$$\min_{a \in V^\star} \{2D_0(W)s\}_a - \max\left\{0, \max_{b \notin V^\star} \{2D_0(W)s\}_b\right\} \ge \frac{3m_n}{4}. \tag{80}$$

Proof. Define

$$\Delta(W, s) = \min_{a \in V^\star} \{2D_0(W)s\}_a - \max\left\{0, \max_{b \notin V^\star} \{2D_0(W)s\}_b\right\}. \tag{81}$$

The map in (81) is continuous. By (28) and compactness of $\mathcal{W}^\star$, there exists $\rho_{3,n} > 0$ such that

$$\Delta(W, s) \ge \frac{3m_n}{4} \tag{82}$$

whenever $d_O(W, \mathcal{W}^\star) \le \rho_{3,n}$ and $\|s - u^\star\|_2 \le \rho_{3,n}$. Let

$$L_{0,n} = \sup_{W \in O(p)} \|D_0(W)\|_{\mathrm{op}}. \tag{83}$$

Equations (11) and (50) imply $L_{0,n} \le pM_0 < \infty$. Choose r_0 so that

$$0 < r_0 \le \min\left\{\rho_{2,n}, \rho_{3,n}, \frac{1}{2}\right\}, \qquad r_0 L_{0,n} \le \frac{m_n}{8}. \tag{84}$$

The choice $r_0 \le \rho_{3,n}$ and (82) prove (80). □

Corollary 1 bounds $J^\star - \mathcal{J}(W,s)$ using $d_O(W,\mathcal{W}^\star)$ and $\|s-u^\star\|_1$.

Corollary 1. *Under Assumptions A1–A5, the radius r_0 satisfying* (84) *can be chosen so that every $(W,s)\in O(p)\times\mathcal{S}_\kappa$ satisfying* (79) *obeys*

$$J^\star - \mathcal{J}(W,s) \geq \frac{\alpha_n}{4} d_O(W,\mathcal{W}^\star)^2 + \frac{m_n}{4}\|s-u^\star\|_1. \tag{85}$$

Proof. Let

$$h = s-u^\star, \qquad A_s = \sum_{a\in V^\star}(1-s_a), \qquad B_s = \sum_{b\notin V^\star} s_b. \tag{86}$$

Feasibility in (2) gives $0\leq B_s\leq A_s$ and

$$\|h\|_1 = A_s + B_s \leq 2A_s. \tag{87}$$

Equations (79) and (84) give $d_O(W,\mathcal{W}^\star)\leq r_0\leq \rho_{3,n}$. Thus (82) applies with $s=u^\star$. Using $0\leq B_s\leq A_s$ and (82), we obtain

$$\begin{aligned}
\langle 2D_0(W)u^\star,\, u^\star - s\rangle &= \sum_{a\in V^\star}\{2D_0(W)u^\star\}_a(1-s_a) - \sum_{b\notin V^\star}\{2D_0(W)u^\star\}_b s_b \\
&\geq A_s \min_{a\in V^\star}\{2D_0(W)u^\star\}_a - B_s\max\left\{0, \max_{b\notin V^\star}\{2D_0(W)u^\star\}_b\right\} \\
&\geq A_s\left[\min_{a\in V^\star}\{2D_0(W)u^\star\}_a - \max\left\{0, \max_{b\notin V^\star}\{2D_0(W)u^\star\}_b\right\}\right] \\
&\geq \frac{3m_n}{4}A_s \geq \frac{3m_n}{8}\|s-u^\star\|_1.
\end{aligned} \tag{88}$$

The following identity

$$\mathcal{J}(W,u^\star) - \mathcal{J}(W,s) = \langle 2D_0(W)u^\star,\, u^\star - s\rangle - h^\top D_0(W)h \tag{89}$$

and (88) yield

$$\mathcal{J}(W,u^\star) - \mathcal{J}(W,s) \geq \frac{3m_n}{8}\|h\|_1 - \|D_0(W)\|_{\mathrm{op}}\|h\|_2^2. \tag{90}$$

If $\|h\|_2\leq r_0$, then $\|h\|_2^2\leq r_0\|h\|_1$. Equations (90) and (84) therefore give

$$\mathcal{J}(W,u^\star) - \mathcal{J}(W,s) \geq \frac{m_n}{4}\|s-u^\star\|_1. \tag{91}$$

Equations (79) and (84) give $d_O(W,\mathcal{W}^\star)\leq r_0\leq\rho_{2,n}$, so (65) applies. Combining (65) and (91) proves (85). □

For an empirical pair (W,s), define

$$r_W(W,s) = \|\operatorname{grad}_W \bar{J}_n(W,s)\|_F, \qquad r_s(W,s) = \sup_{v\in\mathcal{S}_\kappa}\left\langle \nabla_s \bar{J}_n(W,s),\, v-s\right\rangle. \tag{92}$$

The second residual is nonnegative because $v=s$ is feasible. It is the maximum of a linear function over the compact convex set in (2). Both residuals are zero at a local maximizer.

For rank-κ orthogonal projectors P and Q, $\Theta(P,Q)$ denotes the diagonal matrix of principal angles between their ranges.

Let $(\widehat{W},\widehat{s})\in O(p)\times\mathcal{S}_\kappa$ and $\varepsilon_W,\varepsilon_s\geq 0$ satisfy

$$d_O(\widehat{W},\mathcal{W}^\star) < r_0, \qquad \|\widehat{s}-u^\star\|_2 < r_0, \tag{93}$$

$$r_W(\widehat{W},\widehat{s})\leq\varepsilon_W, \qquad r_s(\widehat{W},\widehat{s})\leq\varepsilon_s. \tag{94}$$

For the following residual bounds, suppose that

$$\eta_n(x)\leq\frac{m_n}{16\kappa}. \tag{95}$$

Theorem 2 bounds $\|\widehat{s}-u^\star\|_1$ using $r_s(\widehat{W},\widehat{s})$.

Theorem 2. *Under Assumptions A1–A5, on $\mathcal{G}_x$, conditions* (93)–(95) *imply*

$$\text{(i)} \quad \|\widehat{s} - u^\star\|_1 \le \frac{4\varepsilon_s}{m_n}, \tag{96}$$

$$\text{(ii)} \quad \min_{a \in V^\star} \widehat{s}_a > \max_{b \notin V^\star} \widehat{s}_b, \quad \mathrm{Top}_\kappa(\widehat{s}) = V^\star \tag{97}$$

In particular, $\varepsilon_s = 0$ implies $\widehat{s} = u^\star$.

Proof. Equations (80) and (53) imply

$$\min_{a \in V^\star} \{\nabla_s \bar{J}_n(\widehat{W}, \widehat{s})\}_a - \max\left\{0, \max_{b \notin V^\star} \{\nabla_s \bar{J}_n(\widehat{W}, \widehat{s})\}_b\right\} \ge \frac{3m_n}{4} - 4\kappa\eta_n(x) \ge \frac{m_n}{2}, \tag{98}$$

where the last inequality uses (95).

Define

$$\widehat{A} = \sum_{a \in V^\star} (1 - \widehat{s}_a), \qquad \widehat{B} = \sum_{b \notin V^\star} \widehat{s}_b. \tag{99}$$

Feasibility gives $0 \le \widehat{B} \le \widehat{A}$. As in the mass comparison leading to (88), the empirical score gap in (98) yields

$$\begin{aligned} &\left\langle \nabla_s \bar{J}_n(\widehat{W}, \widehat{s}),\, u^\star - \widehat{s} \right\rangle \\ &\quad \ge \widehat{A}\left[\min_{a \in V^\star} \{\nabla_s \bar{J}_n(\widehat{W}, \widehat{s})\}_a - \max\left\{0, \max_{b \notin V^\star} \{\nabla_s \bar{J}_n(\widehat{W}, \widehat{s})\}_b\right\}\right] \\ &\quad \ge \frac{m_n}{2}\widehat{A}. \end{aligned} \tag{100}$$

Since $u^\star \in \mathcal{S}_\kappa$, (92) and (94) bound the left-hand side of (100) by ε_s. Thus

$$\widehat{A} \le \frac{2\varepsilon_s}{m_n}, \qquad \|\widehat{s} - u^\star\|_1 = \widehat{A} + \widehat{B} \le \frac{4\varepsilon_s}{m_n}, \tag{101}$$

which proves (96).

For every coordinate j, (93) and (84) give

$$|\widehat{s}_j - u^\star_j| \le \|\widehat{s} - u^\star\|_2 < r_0 \le \frac{1}{2}. \tag{102}$$

Consequently, $\widehat{s}_a > 1/2$ for $a \in V^\star$ and $\widehat{s}_b < 1/2$ for $b \notin V^\star$, which proves (97). The last assertion follows from (96) with $\varepsilon_s = 0$. □

We denote

$$b_n(x, \varepsilon_W, \varepsilon_s) = \frac{2}{\alpha_n}\left\{\varepsilon_W + 2(\kappa - 1)\sqrt{\kappa}\,\eta_n(x) + \frac{8(\kappa + \sqrt{\kappa})\{M_0 + \eta_n(x)\}\varepsilon_s}{m_n}\right\}. \tag{103}$$

Corollary 2 bounds $d_O(\widehat{W}, \mathcal{W}^\star)$ in terms of ε_W, ε_s, and $\eta_n(x)$.

Corollary 2. *Under Assumptions A1–A5, on $\mathcal{G}_x$, conditions* (93)–(95) *imply*

$$d_{V^\star}(\widehat{W}, \mathcal{W}^\star) \le d_O(\widehat{W}, \mathcal{W}^\star) \le b_n(x, \varepsilon_W, \varepsilon_s). \tag{104}$$

Proof. For the alignment bound, the triangle inequality, followed by (58), (57), (94), and (96), yields

$$\begin{aligned} &\left\|\mathrm{grad}_W\, \mathcal{J}(\widehat{W}, u^\star)\right\|_F \\ &\quad \le r_W(\widehat{W}, \widehat{s}) + \left\|\mathrm{grad}_W\, \bar{J}_n(\widehat{W}, \widehat{s}) - \mathrm{grad}_W\, \bar{J}_n(\widehat{W}, u^\star)\right\|_F + \left\|\mathrm{grad}_W(\bar{J}_n - \mathcal{J})(\widehat{W}, u^\star)\right\|_F \\ &\quad \le \varepsilon_W + \frac{8(\kappa + \sqrt{\kappa})\{M_0 + \eta_n(x)\}\varepsilon_s}{m_n} + 2(\kappa - 1)\sqrt{\kappa}\,\eta_n(x). \end{aligned}$$

Combining (64) and (105) gives the upper bound on d_O in (104).

For every $U \in \mathcal{W}^\star$, the chordal distance is no larger than the geodesic distance, so

$$\|(\widehat{W} - U)P_{V^\star}\|_F \le \|\widehat{W} - U\|_F \le d_O(\widehat{W}, U). \tag{105}$$

Taking the infimum in (105) proves $d_{V^\star} \le d_O$ in (104). □

Corollary 3 transfers (104) to $\widehat{W}P_{V^\star}\widehat{W}^\top - \Pi^\star$.

Corollary 3. *Under Assumptions A1–A5, on $\mathcal{G}_x$, conditions* (93)–(95) *imply*

$$\text{(i)} \quad \left\|\widehat{W}P_{V^\star}\widehat{W}^\top - \Pi^\star\right\|_F \le 2b_n(x, \varepsilon_W, \varepsilon_s), \tag{106}$$

$$\text{(ii)} \quad \left\|\sin\Theta\Big(\widehat{W}P_{V^\star}\widehat{W}^\top, \Pi^\star\Big)\right\|_F^2 \le 2b_n(x, \varepsilon_W, \varepsilon_s)^2. \tag{107}$$

Proof. For every $U \in \mathcal{W}^\star$, (26) gives

$$\widehat{W}P_{V^\star}\widehat{W}^\top - \Pi^\star = (\widehat{W} - U)P_{V^\star}\widehat{W}^\top + UP_{V^\star}(\widehat{W} - U)^\top. \tag{108}$$

The Frobenius norm of each term on the right-hand side of (108) is at most $\|(\widehat{W} - U)P_{V^\star}\|_F$. Taking the infimum over $U \in \mathcal{W}^\star$ yields

$$\left\|\widehat{W}P_{V^\star}\widehat{W}^\top - \Pi^\star\right\|_F \le 2d_{V^\star}(\widehat{W}, \mathcal{W}^\star), \tag{109}$$

which proves (106). Finally, rank-κ orthogonal projectors P and Q satisfy

$$\|P - Q\|_F^2 = 2\kappa - 2\,\mathrm{Tr}(PQ) = 2\|\sin\Theta(P, Q)\|_F^2. \tag{110}$$

Equations (106) and (110) prove (107). □

B.2 Existence of an empirical local maximizer

All interiors and boundaries in this section are taken relative to

$$\mathcal{X}_\kappa = O(p) \times \mathcal{S}_\kappa. \tag{111}$$

For $0 < r < r_0$, define

$$\overline{\mathcal{N}}_r = \{(W, s) \in \mathcal{X}_\kappa : d_O(W, \mathcal{W}^\star) \le r,\ \|s - u^\star\|_2 \le r\}, \tag{112}$$

and

$$\mathcal{N}_r = \{(W, s) \in \mathcal{X}_\kappa : d_O(W, \mathcal{W}^\star) < r,\ \|s - u^\star\|_2 < r\}. \tag{113}$$

The following lemma characterizes the relative boundary of the set above.

Lemma 6. *The relative boundary of* (112) *is*

$$\begin{aligned}\partial_{\mathcal{X}_\kappa}\overline{\mathcal{N}}_r = \mathcal{B}_r = {} & \{(W, s) \in \mathcal{X}_\kappa : d_O(W, \mathcal{W}^\star) = r,\ \|s - u^\star\|_2 \le r\} \\ & \cup \{(W, s) \in \mathcal{X}_\kappa : d_O(W, \mathcal{W}^\star) \le r,\ \|s - u^\star\|_2 = r\}.\end{aligned} \tag{114}$$

Proof. The continuity of the two distance maps shows that every point in (113) is in the relative interior of $\overline{\mathcal{N}}_r$. Conversely, suppose $(W, s) \in \overline{\mathcal{N}}_r$ and $d_O(W, \mathcal{W}^\star) = r$. Equation (66) gives $U \in \mathcal{W}^\star$ and $\xi \perp T_U\mathcal{W}^\star$ such that

$$W = \mathrm{Exp}_U(\xi), \qquad \|\xi\|_F = r. \tag{115}$$

For every sufficiently small $\delta > 0$,

$$W_\delta = \mathrm{Exp}_U\{(1 + \delta)\xi\}, \qquad d_O(W_\delta, \mathcal{W}^\star) = (1 + \delta)r > r, \qquad W_\delta \longrightarrow W. \tag{116}$$

Equations (84) and (71) give $r < r_0 \le \rho_{2,n} < \rho_{1,n}$. Hence δ can be chosen so that (116) remains in the tubular neighborhood; its asserted distance equality follows from (115) and (67). Thus (W, s) is not in the relative interior of $\overline{\mathcal{N}}_r$.

Now suppose $(W, s) \in \overline{\mathcal{N}}_r$ and $\|s - u^\star\|_2 = r$. Set $h = s - u^\star$. Feasibility gives

$$h_a \le 0 \quad (a \in V^\star), \qquad h_b \ge 0 \quad (b \notin V^\star), \qquad \mathbf{1}^\top h \le 0. \tag{117}$$

Since $r < r_0 < 1$, $\max_j |h_j| < 1$. Hence, for every sufficiently small $\delta > 0$,

$$s_\delta = u^\star + (1 + \delta)h \in \mathcal{S}_\kappa, \qquad \|s_\delta - u^\star\|_2 = (1 + \delta)r > r, \qquad s_\delta \longrightarrow s. \tag{118}$$

To verify the feasibility asserted in (118), choose δ small enough that every coordinate remains in $[0, 1]$; then

$$\mathbf{1}^\top s_\delta = \kappa + (1 + \delta)\mathbf{1}^\top h \le \kappa \tag{119}$$

by (117). Thus (W, s) is not in the relative interior of $\overline{\mathcal{N}}_r$. Since $\overline{\mathcal{N}}_r$ is relatively closed, these arguments prove (114). □

Assumption A6. For $0 < r < r_0$,

$$\eta_n(x) \le \frac{m_n}{16\kappa}, \tag{120a}$$

$$2\kappa^2\eta_n(x) < \frac{1}{4}\min\{\alpha_n r^2, m_n r\}. \tag{120b}$$

On $\mathcal{G}_x$, (120a) ensures that each $\{\nabla_s \bar{J}_n(W,s)\}_a$, $a \in V^\star$, exceeds both zero and every $\{\nabla_s \bar{J}_n(W,s)\}_b$, $b \notin V^\star$, by at least $m_n/2$ for (W,s) satisfying (79), as in (98). Condition (120b) makes the uniform bound in (52) sufficiently smaller than the lower bound for $J^\star - \mathcal{J}(W,s)$ on $\mathcal{B}_r$ as in (122) below.

Theorem 3 establishes that $\bar{J}_n$ has local maximizers in $\mathcal{N}_r$.

Theorem 3. *Under Assumptions A1–A6, on $\mathcal{G}_x$. Every maximizer $(\widehat{W}, \widehat{s})$ of $\bar{J}_n$ over $\overline{\mathcal{N}}_r$ is a local maximizer on $\mathcal{X}_\kappa$ and satisfies*

$$d_O(\widehat{W}, \mathcal{W}^\star) < r, \qquad \|\widehat{s} - u^\star\|_2 < r. \tag{121}$$

At least one such maximizer exists.

Proof. The set $\mathcal{X}_\kappa$ is compact. Equation (33) gives continuity of the set-distance, so $\overline{\mathcal{N}}_r$ in (112) is a nonempty compact subset of $\mathcal{X}_\kappa$. Since $\bar{J}_n$ is continuous, it attains its maximum over $\overline{\mathcal{N}}_r$.

We adapt the argument used in the proof of Theorem 3 in Ning and Liu [31]. The comparison is over the relative boundary in (114). For every $(W,s) \in \mathcal{B}_r$, (85) and $\|s - u^\star\|_1 \ge \|s - u^\star\|_2$ give

$$J^\star - \mathcal{J}(W,s) \ge \frac{1}{4}\min\{\alpha_n r^2, m_n r\}. \tag{122}$$

The curvature term in (85) gives the $\alpha_n r^2/4$ bound on the first face of (114), whereas its support term gives the $m_n r/4$ bound on the second face. For any $U \in \mathcal{W}^\star$, equations (52) and (122) yield

$$\begin{aligned} \bar{J}_n(U,u^\star) - \bar{J}_n(W,s) &= J^\star - \mathcal{J}(W,s) + (\bar{J}_n - \mathcal{J})(U,u^\star) - (\bar{J}_n - \mathcal{J})(W,s) \\ &\ge \frac{1}{4}\min\{\alpha_n r^2, m_n r\} - 2\kappa^2\eta_n(x) > 0, \end{aligned} \tag{123}$$

where the strict inequality is due to (120b). The point $(U,u^\star)$ belongs to $\overline{\mathcal{N}}_r$. Hence (123) excludes $\mathcal{B}_r$ from the set of restricted maximizers. By (114), every restricted maximizer belongs to $\mathcal{N}_r$, proving (121).

Since $\mathcal{N}_r$ is open relative to $\mathcal{X}_\kappa$, every restricted maximizer has a relative neighborhood in $\mathcal{X}_\kappa$ contained in $\overline{\mathcal{N}}_r$. It is therefore a local maximizer of $\bar{J}_n$ on the original feasible set $\mathcal{X}_\kappa$, including when its support vector lies on a coordinate or respective face of $\mathcal{S}_\kappa$. □

Corollary 4 evaluates r_W and r_s at the maximizers in Theorem 3.

Corollary 4. *Under Assumptions A1–A6, on $\mathcal{G}_x$. Every maximizer in Theorem 3 satisfies*

$$\text{(i)} \quad r_W(\widehat{W}, \widehat{s}) = 0, \qquad \text{(ii)} \quad r_s(\widehat{W}, \widehat{s}) = 0. \tag{124}$$

Proof. Fix any such maximizer. Holding its support block fixed and varying the orthogonal block along smooth curves gives

$$\operatorname{grad}_W \bar{J}_n(\widehat{W}, \widehat{s}) = 0. \tag{125}$$

For every $v \in \mathcal{S}_\kappa$, the segment

$$t \longmapsto \widehat{s} + t(v - \widehat{s}), \qquad 0 \le t \le 1, \tag{126}$$

is feasible. The one-sided derivative at $t = 0$ is nonpositive by local maximality, and hence

$$\left\langle \nabla_s \bar{J}_n(\widehat{W}, \widehat{s}),\, v - \widehat{s} \right\rangle \le 0, \qquad v \in \mathcal{S}_\kappa. \tag{127}$$

Taking the supremum in (127), with $v = \widehat{s}$ attaining zero, gives $r_s(\widehat{W}, \widehat{s}) = 0$ by (92). Equation (125) gives $r_W(\widehat{W}, \widehat{s}) = 0$, proving (124). □

Corollary 5 combines (124) and (96) to obtain $\widehat{s} = u^\star$.

Corollary 5. *Under Assumptions A1–A6, on $\mathcal{G}_x$. Every maximizer in Theorem 3 satisfies*

$$\begin{aligned} &\text{(i)} \quad \widehat{s} = u^\star, \\ &\text{(ii)} \quad \operatorname{supp}(\widehat{s}) = \operatorname{Top}_\kappa(\widehat{s}) = V^\star. \end{aligned} \tag{128}$$

Proof. Equation (121), (124), and (120a) verify (93), (94), and (95), respectively, with $\varepsilon_W = \varepsilon_s = 0$. Therefore Theorem 2 applies, and (96) gives $\widehat{s} = u^\star$. Since $u^\star = \mathbf{1}_{V^\star}$ by definition, its support and its κ largest coordinates both equal $V^\star$, proving (128). □

Corollary 6 specializes (104) to $\varepsilon_W = \varepsilon_s = 0$.

Corollary 6. *Under Assumptions A1–A6, on $\mathcal{G}_x$. Every maximizer in Theorem 3 satisfies*

$$d_{V^\star}(\widehat{W}, \mathcal{W}^\star) \le d_O(\widehat{W}, \mathcal{W}^\star) \le \frac{4(\kappa-1)\sqrt{\kappa}\,\eta_n(x)}{\alpha_n}. \tag{129}$$

Proof. Equations (121) and (124), together with (120a), verify (93)–(95) with $\varepsilon_W = \varepsilon_s = 0$. Hence Corollary 2 applies. Equation (103) gives

$$b_n(x, 0, 0) = \frac{4(\kappa-1)\sqrt{\kappa}\,\eta_n(x)}{\alpha_n}. \tag{130}$$

Substitution of (130) into (104) yields (129). □

Corollary 7 specializes (106) and (107) to $\varepsilon_W = \varepsilon_s = 0$.

Corollary 7. *Under Assumptions A1–A6, on $\mathcal{G}_x$. Every maximizer in Theorem 3 satisfies*

$$\text{(i)} \quad \left\| \widehat{W} P_{V^\star} \widehat{W}^\top - \Pi^\star \right\|_F \le \frac{8(\kappa-1)\sqrt{\kappa}\,\eta_n(x)}{\alpha_n}, \tag{131}$$

$$\text{(ii)} \quad \left\| \sin\Theta\Big(\widehat{W} P_{V^\star} \widehat{W}^\top, \Pi^\star\Big) \right\|_F^2 \le \frac{32\kappa(\kappa-1)^2\eta_n(x)^2}{\alpha_n^2}. \tag{132}$$

Proof. The same substitutions into (93)–(95) allow Corollary 3 to be applied with $\varepsilon_W = \varepsilon_s = 0$. Substitution of (130) into (106) and (107) yields (131) and (132), respectively. □

If the deterministic inequalities in (120) hold, then (44) and Theorem 3 together with Corollaries 4–7 imply that (121), (128), (129), (131), and (132) hold simultaneously with probability at least $1-2e^{-x}$.

Theorem 1 follows from Theorem 3, Theorem 2, and Corollary 3.

Proof of Theorem 1. By Lemma 1, $\mathbb{P}(\mathcal{G}_x) \ge 1 - 2e^{-x}$. On $\mathcal{G}_x$, Theorem 3 and (120) give at least one local maximizer of $\bar{J}_n$ satisfying

$$d_O(\widehat{W}, \mathcal{W}^\star) < r, \quad \|\widehat{s} - u^\star\|_2 < r. \tag{133}$$

Since $J = n\bar{J}_n$ with $n > 0$, the local maximizers of J and $\bar{J}_n$ coincide.

Now let $(\widehat{W}, \widehat{s})$ be any local maximizer satisfying (133). Varying W along smooth curves in $O(p)$ and s along the feasible segments $\widehat{s} + t(v - \widehat{s})$ yields

$$\operatorname{grad}_W \bar{J}_n(\widehat{W}, \widehat{s}) = 0, \qquad \Big\langle \nabla_s \bar{J}_n(\widehat{W}, \widehat{s}),\, v - \widehat{s} \Big\rangle \le 0, \quad v \in \mathcal{S}_\kappa. \tag{134}$$

Consequently, both residuals in (92) are zero. Since $r < r_0$, (133) implies (93). Thus Theorem 2, with $\varepsilon_W = \varepsilon_s = 0$, yields $\widehat{s} = u^\star$ and $\widehat{\mathcal{V}} = \operatorname{Top}_\kappa(\widehat{s}) = V^\star$. Corollary 3 and (130) give the projector bound in (4). All these conclusions hold on the event $\mathcal{G}_x$, which proves the theorem. □

B.3 Sufficient conditions

We verify the assumptions for a specified shared-signal family of the observed matrices. For the fixed $V^\star$ introduced above, suppose

$$2 \le |V^\star| = \kappa < p. \tag{135}$$

For $\ell = 1, \dots, L$, let the fixed matrices $C^{(\ell)} \in \mathbb{R}^{p\times p}$ satisfy

$$C^{(\ell)} = C^{(\ell)\top}, \qquad \operatorname{diag}\Big(C^{(\ell)}\Big) = 0, \qquad C^{(\ell)} = P_{V^\star} C^{(\ell)} P_{V^\star}. \tag{136}$$

Suppose that

$$C_i = \sum_{\ell=1}^{L} z_{i\ell} C^{(\ell)}, \qquad z_i = (z_{i1}, \dots, z_{iL})^\top, \qquad \Sigma_z = \mathbb{E}(z_i z_i^\top) \succ 0, \tag{137}$$

where Σ_z is a finite second-moment matrix, and that, for a fixed $W^\star \in O(p)$ and $\sigma_A, \sigma_B \ge 0$,

$$A_i = W^\star C_i W^{\star\top} + \sigma_A E_i^A, \qquad B_i = C_i + \sigma_B E_i^B. \tag{138}$$

The noise matrices satisfy

$$\mathbb{E}(E_i^A) = \mathbb{E}(E_i^B) = 0_{p\times p}, \qquad E_i^A = E_i^{A\top}, \quad E_i^B = E_i^{B\top}, \quad \operatorname{diag}(E_i^B) = 0 \quad \text{almost surely.} \tag{139}$$

Assume that, for every a, b,

$$\begin{aligned} \mathbb{E}(C_{i,ab} E_i^A) &= 0_{p\times p}, \\ \mathbb{E}\big(E_{i,ab}^B W^\star C_i W^{\star\top}\big) &= 0_{p\times p}, \\ \mathbb{E}(E_{i,ab}^B E_i^A) &= 0_{p\times p}, \end{aligned} \tag{140}$$

and that

$$\bigcap_{\ell=1}^{L} \ker\Big(C^{(\ell)}\Big) = \operatorname{span}\{e_j : j \notin V^\star\} = \ker(P_{V^\star}). \tag{141}$$

Suppose further that

$$\{(z_i, E_i^A, E_i^B)\}_{i=1}^n \quad \text{are independent and identically distributed,} \tag{142}$$

and that finite, possibly dimension-dependent constants K_A, K_B satisfy

$$\max_{a,b} \|B_{i,ab}\|_{\psi_2} \le K_B, \qquad \sup_{\|x\|_2=\|y\|_2=1} \|x^\top A_i y\|_{\psi_2} \le K_A. \tag{143}$$

For this model, the quantities in Assumption A3 are given by

$$\begin{aligned} &\mathcal{W}^\star = \{W^\star Q : Q \in O(p),\ Q^\top C^{(\ell)} Q = C^{(\ell)} \text{ for } \ell = 1, \dots, L\}, \\ &\Pi^\star = W^\star P_{V^\star} W^{\star\top}, \qquad J^\star = \mathbb{E}\|C_i\|_F^2. \end{aligned} \tag{144}$$

Set

$$m_n = 2 \min_{a\in V^\star} \sum_{b\in V^\star\setminus\{a\}} \mathbb{E}(C_{i,ab}^2). \tag{145}$$

Proposition 1 verifies Assumptions A1–A5 for the model in (138).

Proposition 1. *Under* (135)*–*(143)*, Assumptions A1–A5 hold with* (144)*, the strictly positive margin in* (145)*, $\nu_n = CK_AK_B$ for a universal constant $C > 0$, and the strictly positive curvature constant $\alpha_n = \alpha_C$ defined in* (161) *below.*

Proof. Every $W \in O(p)$ has the unique representation $W = W^\star Q$, where $Q = W^{\star\top} W \in O(p)$. Expanding the population criterion yields

$$\mathcal{J}(W^\star Q, u^\star) = \mathbb{E}\operatorname{Tr}(Q^\top C_i Q C_i). \tag{146}$$

Indeed, for every $a, b \in V^\star$, the three remaining terms in the entrywise expansion have coefficients σ_A, σ_B, and $\sigma_A\sigma_B$, respectively, and their unscaled expectations vanish by (140):

$$\mathbb{E}\big[C_{i,ab} e_a^\top Q^\top W^{\star\top} E_i^A W^\star Q e_b\big] = e_a^\top Q^\top W^{\star\top} \mathbb{E}(C_{i,ab} E_i^A) W^\star Q e_b = 0, \tag{147}$$

$$\mathbb{E}\big[E_{i,ab}^B e_a^\top Q^\top C_i Q e_b\big] = e_a^\top Q^\top W^{\star\top} \mathbb{E}\big(E_{i,ab}^B W^\star C_i W^{\star\top}\big)\, W^\star Q e_b = 0,$$

$$\mathbb{E}\big[E_{i,ab}^B e_a^\top Q^\top W^{\star\top} E_i^A W^\star Q e_b\big] = e_a^\top Q^\top W^{\star\top} \mathbb{E}(E_{i,ab}^B E_i^A) W^\star Q e_b = 0. \tag{148}$$

Symmetry identifies the remaining sum with the trace in (146). Orthogonal invariance of the Frobenius norm therefore yields

$$J^\star - \mathcal{J}(W^\star Q, u^\star) = \frac{1}{2}\mathbb{E}\|Q^\top C_i Q - C_i\|_F^2. \tag{149}$$

Moreover, expansion in the fixed matrices $C^{(\ell)}$ gives the second-moment identity

$$\begin{aligned}\mathbb{E}\|Q^\top C_i Q - C_i\|_F^2 &= \sum_{\ell,m=1}^{L} (\Sigma_z)_{\ell m} \left\langle Q^\top C^{(\ell)} Q - C^{(\ell)},\, Q^\top C^{(m)} Q - C^{(m)} \right\rangle \\ &\geq \lambda_{\min}(\Sigma_z) \sum_{\ell=1}^{L} \|Q^\top C^{(\ell)} Q - C^{(\ell)}\|_F^2. \end{aligned} \tag{150}$$

The inequality follows by applying the smallest-eigenvalue bound for Σ_z to the L coefficients at each matrix entry and summing over the entries. Since $\Sigma_z \succ 0$, the expression in (150) is zero exactly when every $C^{(\ell)}$ is preserved by Q. Thus (144) gives the full population maximizer set at $u^\star$.

Define

$$\mathcal{H}_C = \{Q \in O(p) : Q^\top C^{(\ell)} Q = C^{(\ell)} \text{ for } \ell = 1, \dots, L\}. \tag{151}$$

This is a closed subgroup of the compact group $O(p)$. Cartan's closed subgroup theorem implies that it is an embedded Lie subgroup [25, Theorem 20.12]. Consequently, $\mathcal{W}^\star = W^\star \mathcal{H}_C$ is a nonempty compact smooth embedded submanifold of $O(p)$. Each of its points maximizes $\mathcal{J}(\cdot, u^\star)$ by (149); hence its Riemannian gradient is zero.

For $Q \in \mathcal{H}_C$, the defining equalities are equivalent to

$$C^{(\ell)} Q = Q C^{(\ell)}, \qquad \ell = 1, \dots, L. \tag{152}$$

It follows that Q preserves the common kernel in (141). Orthogonality then implies preservation of its orthogonal complement, $\mathrm{range}(P_{V^\star})$, and therefore

$$Q P_{V^\star} Q^\top = P_{V^\star}, \qquad (W^\star Q) P_{V^\star} (W^\star Q)^\top = \Pi^\star. \tag{153}$$

If $R \in O(p)$ satisfies $R P_{V^\star} = P_{V^\star}$, orthogonality gives

$$R^\top P_{V^\star} = P_{V^\star} R = P_{V^\star}. \tag{154}$$

The support condition in (136) now implies

$$R^\top C^{(\ell)} R = R^\top P_{V^\star} C^{(\ell)} P_{V^\star} R = C^{(\ell)}. \tag{155}$$

Thus $R \in \mathcal{H}_C$, and the group property gives $UR \in \mathcal{W}^\star$ whenever $U \in \mathcal{W}^\star$. These identities verify every part of Assumption A3.

The Lie algebra of $\mathcal{H}_C$ is

$$\mathfrak{h}_C = \{\Omega \in \mathbb{R}^{p\times p} : \Omega^\top = -\Omega,\ C^{(\ell)}\Omega = \Omega C^{(\ell)} \text{ for } \ell = 1, \dots, L\}. \tag{156}$$

Indeed, differentiating the defining equalities of $\mathcal{H}_C$ at the identity gives the displayed commutation relations; conversely, these relations imply that $\exp(t\Omega)$ commutes with every $C^{(\ell)}$ for all t. For a skew-symmetric Ω, differentiating (149) along $Q(t) = \exp(t\Omega)$ yields

$$-\frac{d^2}{dt^2} \mathcal{J}(W^\star \exp(t\Omega), u^\star)\Big|_{t=0} = \mathbb{E}\|C_i\Omega - \Omega C_i\|_F^2. \tag{157}$$

To see the coefficient explicitly, the matrix $F_i(t) = \exp(-t\Omega) C_i \exp(t\Omega) - C_i$ satisfies $F_i(0) = 0$ and $F_i'(0) = C_i\Omega - \Omega C_i$; hence the second derivative of $\|F_i(t)\|_F^2/2$ at zero is $\|F_i'(0)\|_F^2$. The finite

second moment in (137) justifies differentiation after expectation, or equivalently differentiation of the finite second-moment expansion.

The same expansion as in (150) gives

$$\mathbb{E}\|C_i\Omega - \Omega C_i\|_F^2 = \sum_{\ell,m=1}^{L} (\Sigma_z)_{\ell m} \left\langle C^{(\ell)}\Omega - \Omega C^{(\ell)},\, C^{(m)}\Omega - \Omega C^{(m)} \right\rangle$$

$$\geq \lambda_{\min}(\Sigma_z) \sum_{\ell=1}^{L} \|C^{(\ell)}\Omega - \Omega C^{(\ell)}\|_F^2. \tag{158}$$

Consequently,

$$\mathbb{E}\|C_i\Omega - \Omega C_i\|_F^2 = 0 \quad \Longleftrightarrow \quad \Omega \in \mathfrak{h}_C. \tag{159}$$

The orthogonal complement of $\mathfrak{h}_C$ within the skew-symmetric matrices has positive dimension. To verify this, choose unit vectors $v \in \mathrm{range}(P_{V^\star})$ and $w \in \ker(P_{V^\star})$. By (141), there is an ℓ such that $C^{(\ell)}v \neq 0$. The skew-symmetric matrix $\Omega_{v,w} = vw^\top - wv^\top$ satisfies

$$\{C^{(\ell)}\Omega_{v,w} - \Omega_{v,w}C^{(\ell)}\}w = C^{(\ell)}v \neq 0, \tag{160}$$

so $\mathfrak{h}_C$ is a proper subspace. Define

$$\alpha_C = \min_{\substack{\Omega^\top = -\Omega,\ \Omega \perp \mathfrak{h}_C \\ \|\Omega\|_F = 1}} \mathbb{E}\|C_i\Omega - \Omega C_i\|_F^2. \tag{161}$$

The constraint set is nonempty and compact, and the objective is continuous and strictly positive there by (159). Hence $\alpha_C > 0$.

For $Q_0 \in \mathcal{H}_C$ and $U = W^\star Q_0$, left multiplication by U is an isometry for the Frobenius metric, and

$$\begin{aligned} T_U\mathcal{W}^\star &= \{U\Omega : \Omega \in \mathfrak{h}_C\}, \\ \{\xi \in T_U O(p) : \xi \perp T_U\mathcal{W}^\star\} &= \{U\Omega : \Omega^\top = -\Omega,\ \Omega \perp \mathfrak{h}_C\}. \end{aligned} \tag{162}$$

For such $\xi = U\Omega$, the curve $t \mapsto U\exp(t\Omega)$ is the geodesic with initial velocity ξ [13, Section 2.2.2, Eq. (2.14)]. Since Q_0 preserves every C_i, one has

$$\mathcal{J}(W^\star Q_0 \exp(t\Omega), u^\star) = \mathcal{J}(W^\star \exp(t\Omega), u^\star). \tag{163}$$

The geodesic second-derivative identity [2, Proposition 5.5.4], (157), and (161) therefore imply

$$-\langle \xi,\, \mathrm{Hess}_{WW}\, \mathcal{J}(U, u^\star)[\xi] \rangle \geq \alpha_C \|\Omega\|_F^2 = \alpha_C \|\xi\|_F^2. \tag{164}$$

This proves Assumption A4 with $\alpha_n = \alpha_C$.

For $U = W^\star Q \in \mathcal{W}^\star$, expansion of (11), followed by (152) and (140), yields

$$\{D_0(U)\}_{ab} = \begin{cases} \mathbb{E}(C_{i,ab}^2), & a, b \in V^\star, \\ 0, & \{a, b\} \not\subseteq V^\star. \end{cases} \tag{165}$$

As $\mathrm{diag}(C_i) = 0$, this implies

$$\{2D_0(U)u^\star\}_a = \begin{cases} 2 \displaystyle\sum_{b \in V^\star \setminus \{a\}} \mathbb{E}(C_{i,ab}^2), & a \in V^\star, \\ 0, & a \notin V^\star. \end{cases} \tag{166}$$

The selected sums are positive. Indeed, if such a sum were zero for some $a \in V^\star$, then $\mathbb{E}(C_{i,ab}^2) = 0$ for every b. By (137) and $\Sigma_z \succ 0$, this would give $C_{ab}^{(\ell)} = 0$ for all b, ℓ. Its symmetry would then imply $C^{(\ell)}e_a = 0$ for every ℓ, contradicting (141). Thus (145) is strictly positive, and (166) proves Assumption A5.

The structural conditions and (142) give Assumption A1. For unit vectors x, y, Lemma 2.7.7 of [43] implies

$$\|B_{i,ab}x^\top A_i y\|_{\psi_1} \leq C\|B_{i,ab}\|_{\psi_2}\|x^\top A_i y\|_{\psi_2} \leq CK_AK_B. \tag{167}$$

This proves Assumption A2 and completes the verification. □

For a concrete instance, take $L = 1$, $C_i = z_i C$, and

$$C_{ab} = \begin{cases} c, & a, b \in V^\star,\ a \neq b, \\ 0, & \text{otherwise}, \end{cases} \qquad c \neq 0. \tag{168}$$

Let $z_i \sim N(0,1)$, and let $G_i^A, G_i^B \in \mathbb{R}^{p\times p}$ have independent $N(0,1)$ entries, with all these variables and matrices mutually independent across modalities and subjects. Set

$$E_i^A = \frac{1}{2}(G_i^A + G_i^{A\top}), \qquad E_{i,ab}^B = \begin{cases} \frac{1}{2}(G_{i,ab}^B + G_{i,ba}^B), & a \neq b, \\ 0, & a = b. \end{cases} \tag{169}$$

Define A_i, B_i by (138) for any fixed $W^\star \in O(p)$. Symmetry, centering, and independence verify (139), (140), and (142).

Note that three distinct eigenvalues of C are $c(\kappa-1)$, $-c$, 0, with multiplicities 1, $\kappa-1$, and $p-\kappa$, respectively. Thus its kernel is exactly $\ker(P_{V^\star})$. Let $\lambda_1, \ldots, \lambda_p$ denote these eigenvalues in an orthonormal eigenbasis. In this basis, the identity $(C\Omega - \Omega C)_{ab} = (\lambda_a - \lambda_b)\Omega_{ab}$ shows that $\mathfrak{h}_C$ consists of the skew-symmetric blocks within the eigenspaces and that its orthogonal complement consists of the blocks between distinct eigenspaces. The smallest absolute difference between distinct eigenvalues within the spectrum is $|c|$. Hence

$$\|C\Omega - \Omega C\|_F^2 \geq c^2 \|\Omega\|_F^2, \qquad \Omega^\top = -\Omega, \quad \Omega \perp \mathfrak{h}_C. \tag{170}$$

Since $\mathbb{E}(z_i^2) = 1$, the constants are therefore $\alpha_n = c^2$, $m_n = 2(\kappa-1)c^2$, and $J^\star = \kappa(\kappa-1)c^2$. For every pair of unit vectors x, y, independence and the symmetrized Gaussian construction yield

$$\mathrm{Var}(x^\top A_i y) = (x^\top W^\star C W^{\star\top} y)^2 + \frac{\sigma_A^2}{2}\{1 + (x^\top y)^2\} \leq c^2(\kappa-1)^2 + \sigma_A^2, \tag{171}$$

$$\max_{a,b} \mathrm{Var}(B_{i,ab}) = c^2 + \frac{\sigma_B^2}{2}. \tag{172}$$

Therefore, for a universal constant $C_\psi > 0$ determined by the convention for the ψ_2 norm, (143) holds with $K_A = C_\psi \sqrt{c^2(\kappa-1)^2 + \sigma_A^2}$, and $K_B = C_\psi \sqrt{c^2 + \sigma_B^2/2}$. The population coefficient bound (50) holds with $\mathbb{E}(B_{i,ab} A_i) = C_{ab} W^\star C W^{\star\top}$, and $M_0 = c^2(\kappa-1)$.

C Full OSSA algorithm pseudocode

Algorithm 1 gives the canonical OSSA single-support OSSA solver. Algorithm 2 gives the overlap-regularized K-support extension.

When $K = 1$ and $\lambda = 0$, the inner ℓ-loop collapses to a single s-update, the aggregate M_i reduces to SB_iS, and the algorithm reduces exactly to Algorithm 1. For $K \geq 2$, the Gauss-Seidel sweep updates each support against the current W and the current values of the other supports, with the overlap penalty entering linearly through $-\lambda \sum_{r\neq\ell} s^{(r)}$. As in the canonical OSSA case, backtracking guarantees a monotone objective sequence, and multiple random restarts are used to mitigate local optima.

Algorithm 2 Overlap-regularized K-clique OSSA

Require: Symmetric matrices $\{A_i, B_i\}_{i=1}^n$ in one community; number of supports K; per-support budget κ; overlap penalty $\lambda \geq 0$; initial steps η_W, η_s; backtracking factor $\beta \in (0,1)$

Ensure: Alignment W, relaxed supports $\{s^{(\ell)}\}_{\ell=1}^K$, and hard supports $\{\widehat{\mathcal{V}}^{(\ell)}\}_{\ell=1}^K$

Initialize $W \leftarrow I_p$ or a random orthogonal matrix; initialize $s^{(\ell)} \in \mathcal{S}_\kappa$ for $\ell = 1, \ldots, K$
repeat
  $S^{(\ell)} \leftarrow \mathrm{diag}(s^{(\ell)})$ for $\ell = 1, \ldots, K$
  $M_i \leftarrow \sum_{\ell=1}^K S^{(\ell)} B_i S^{(\ell)}$ for $i = 1, \ldots, n$
  $G_W \leftarrow 2\sum_{i=1}^n A_i W M_i$ and $W^+(\eta) \leftarrow \Pi_{O(p)}(W + \eta G_W)$
  Choose the largest $\eta \in \{\eta_W \beta^m : m \geq 0\}$ satisfying
  $J_K(W^+(\eta), \{s^{(\ell)}\}) \geq J_K(W, \{s^{(\ell)}\})$, and set $W \leftarrow W^+(\eta)$
  $D \leftarrow \mathrm{sym}\{\sum_{i=1}^n (W^\top A_i W) \circ B_i\}$
  for $\ell = 1, \ldots, K$ **do** ▷ Gauss-Seidel sweep over supports
    $g^{(\ell)} \leftarrow 2Ds^{(\ell)} - \lambda \sum_{r \neq \ell} s^{(r)}$
    $s^{(\ell,+)}(\eta) \leftarrow \Pi_{\mathcal{S}_\kappa}(s^{(\ell)} + \eta g^{(\ell)})$
    Choose the largest $\eta \in \{\eta_s \beta^m : m \geq 0\}$ satisfying
  $J_K(W, \ldots, s^{(\ell,+)}(\eta), \ldots) \geq J_K(W, \ldots, s^{(\ell)}, \ldots)$,
  and set $s^{(\ell)} \leftarrow s^{(\ell,+)}(\eta)$
  end for
until the objective and iterate changes satisfy the stopping tolerances
$\widehat{\mathcal{V}}^{(\ell)} \leftarrow \mathrm{Top}_\kappa(s^{(\ell)})$ for $\ell = 1, \ldots, K$
return W, $\{s^{(\ell)}\}_{\ell=1}^K$, $\{\widehat{\mathcal{V}}^{(\ell)}\}_{\ell=1}^K$

Implementation constants. Unless otherwise stated, we use initial step sizes $\eta_W = 5 \times 10^{-2}$, $\eta_s = 5 \times 10^{-2}$, backtracking factor $\beta = 0.5$ with a floor $\eta_{\min} = 10^{-12}$, relative objective tolerance $\mathrm{tol}_{\text{outer}} = 10^{-5}$, per-block objective-improvement threshold $\mathrm{tol}_{\text{inner}} = 10^{-7}$, and a maximum of 20 outer block-coordinate iterations (with up to 30 projected-gradient iterations per W-block and per support-vector update, each using backtracking line search). Random orthogonal initializations are generated by QR decomposition of an i.i.d. $N(0,1)$ Gaussian matrix (equivalently, samples from the Haar measure on $O(p)$). Support initializations are generated by drawing each entry i.i.d. uniform on $[0,1]$ and then projecting onto $\mathcal{S}_\kappa$ via the clip-and-soft-threshold simplex projection, we verified that uniform $s = (\kappa/p)\mathbf{1}_p$ and binary κ-hot random initializations produce indistinguishable selections within restart-to-restart variability. For the ABCD experiments, we use 16 random restarts for canonical OSSA evaluation (Algorithm 1), 4 for per-κ selection, and 4 for multi-clique fits (Algorithm 2). The canonical single-clique simulation sweeps use 8 random restarts; settings for the remaining simulation diagnostics are reported in Appendices D.4–D.7. The restart with the largest training objective is retained. For the ABCD bootstrap stability analysis we use 2 restarts with a lighter $(\mathrm{max_outer}, \mathrm{max_iter}_{W,s}) = (10, 15)$ schedule and $\mathrm{tol}_{\text{outer}} = 10^{-4}$, $\mathrm{tol}_{\text{inner}} = 10^{-6}$ to keep the 100-bootstrap budget tractable, retaining the same step sizes and backtracking factor. Detailed optimization settings are listed in Table 5.

Computational cost. In a community of size p, one dense outer iteration costs $O(np^3)$ for the W-block matrix products and $O(np^2)$ for forming $D(W)$, plus $O(Kp^2)$ for support gradients and $O(Kp \log p)$ for capped-simplex projections. The polar projection costs $O(p^3)$. Memory is $O(np^2)$ for dense subject matrices. In the ABCD experiments, the largest community has about 30 nodes, so runtime is dominated by the number of subject-level matrix products and restarts rather than by the cubic polar step.

D Additional experimental results and diagnostics

This appendix collects simulation sweep curves, robustness analyses, and ABCD training-side diagnostics referenced from Sections 3 and 4. Unless otherwise stated, the simulation experiments use the canonical OSSA generative model of Section 3.1 or the multi-clique extension of Section 3.2, with the same nuisance structure.

Table 5: OSSA optimization budgets used at each stage of the pipeline. Joint support-and-alignment fits use initial step sizes $\eta_W = \eta_s = 5\times10^{-2}$, backtracking factor $\beta = 0.5$ with floor $\eta_{\min} = 10^{-12}$, and projected-uniform s initialization. Fixed-support baseline W-refits hold s fixed and use three W initializations, including the identity (Appendix D.9). "Outer" is the maximum number of block-coordinate sweeps for joint fits; for fixed-support W-refits, it gives the maximum number of W-update iterations per initialization. "Inner" is the maximum number of projected-gradient iterations per W-block and per support-vector update; each iteration uses backtracking line search. "Restarts" is the number of optimization runs from the specified initializations; the run with the largest training objective is retained.

Stage	Restarts	Outer	Inner	$\eta_W = \eta_s$	$(\mathrm{tol}_{\mathrm{outer}}, \mathrm{tol}_{\mathrm{inner}})$
Canonical OSSA evaluation (Table 3)	16	20	30	5×10^{-2}	$(10^{-5}, 10^{-7})$
Per-κ selection	4	15	20	5×10^{-2}	$(10^{-5}, 10^{-7})$
Multi-clique fits ($K \geq 2$)	4	15	20	5×10^{-2}	$(10^{-5}, 10^{-7})$
ABCD bootstrap stability	2	10	15	5×10^{-2}	$(10^{-4}, 10^{-6})$
Canonical single-clique sweeps	8	20	40	5×10^{-2}	$(10^{-5}, 10^{-7})$
Per-baseline W-refit (fixed s)	3	200	-	5×10^{-2}	$\mathrm{tol}_{\mathrm{improve}} = 10^{-9}$

D.1 Complete simulation generator

In Section 3.1, the main text writes $\zeta_i^A = \xi_i^A + \tau_i^A$ and $\zeta_i^B = \xi_i^B + \tau_i^B$ to keep the simulation model concise. For each replicate we draw the data $\{(A_i, B_i)\}_{i=1}^n$ from the following procedure (single-clique unless noted; multi-clique substitutions in brackets). Default ambient size is $p = 40$ for the canonical OSSA experiments and $p = 50$ for the multi-clique experiments, with $n = 30$ subjects per replicate. "$\mathrm{LR}_r(V, c)$" denotes a symmetric rank-r random block of magnitude c, supported on $V \subseteq [p]$, generated as $UU^\top \circ \Sigma$ with $U \in \mathbb{R}^{|V|\times r}$ having i.i.d. $N(0, c^2/r)$ entries and Σ a symmetrized random ±1 sign matrix (entries $\Sigma_{ab} = \frac{1}{2}(\sigma_{ab} + \sigma_{ba})$ with $\sigma_{ab} \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}\{-1, +1\}$), embedded as zero outside $V \times V$.

- **Draw planted support(s).** Sample $\mathcal{V}^\star \subseteq [p]$ uniformly without replacement of size $\kappa = 10$. For the multi-clique generator, sample $K \in \{2, 3\}$ supports $\mathcal{V}^{\star,(\ell)} \subseteq [p]$ via a random permutation of $[p]$ and consecutive blocks of size κ, with adjacent supports sharing exactly $r \in \{0, 4\}$ nodes (so they are pairwise disjoint when $r = 0$). Defaults: K=2, κ=10, r=0 (*disjoint*); K=2, κ=10, r=4 (*overlap-4*); K=3, κ=8, r=0 (K=3 *disjoint*).
- **Draw low-rank block U_i and sign matrix Σ_i.** Sample $U_i \in \mathbb{R}^{\kappa\times r_{\mathrm{sig}}}$ with i.i.d. $N(0, c_{\mathrm{sig}}^2/r_{\mathrm{sig}})$ entries and the symmetric sign mask Σ_i as above. Defaults $r_{\mathrm{sig}} = 3$, $c_{\mathrm{sig}} = 1.5$.
- **Form C_i (Canonical OSSA) or C_i^{tot} (multi) and embed.** Canonical OSSA: $C_i = \mathrm{LR}_{r_{\mathrm{sig}}}(\mathcal{V}^\star, c_{\mathrm{sig}})$. Multi: $C_i^{\mathrm{tot}} = \sum_{\ell=1}^K \mathrm{LR}_{r_{\mathrm{sig}}}(\mathcal{V}^{\star,(\ell)}, c_{\mathrm{sig}})$ with an independent $(U_i^{(\ell)}, \Sigma_i^{(\ell)})$ for each support. Both are zero outside their planted supports.
- **Draw background matrices ξ_i^A, ξ_i^B.** Each is a fresh i.i.d. rank-r_{bg} symmetric block with the same LR recipe but supported on all p nodes, with amplitudes $c_{\mathrm{bg}}^A = c_{\mathrm{bg}}^B = 0.6$ (Canonical OSSA) and $c_{\mathrm{bg}}^A = c_{\mathrm{bg}}^B = 0.4$ (multi-clique), $r_{\mathrm{bg}} = 4$. Backgrounds are independent across modalities.
- **Draw modality-specific nuisance blocks τ_i^A, τ_i^B (Canonical OSSA only).** Two disjoint modality-specific nuisance variation node sets $\mathcal{D}^A, \mathcal{D}^B \subseteq [p] \setminus \mathcal{V}^\star$, each of size $|\mathcal{D}| = 12$, are sampled once per replicate (not per subject) by random permutation of $[p] \setminus \mathcal{V}^\star$ and consecutive slicing. For each subject i we then draw $\tau_i^B = \mathrm{LR}_{r_{\mathrm{dec}}}(\mathcal{D}^B, c_{\mathrm{dec}}^B)$ and $\tau_i^A = \mathrm{LR}_{r_{\mathrm{dec}}}(\mathcal{D}^A, c_{\mathrm{dec}}^A)$ independently, with $r_{\mathrm{dec}} = 3$ and $c_{\mathrm{dec}}^A = c_{\mathrm{dec}}^B = 1.5$. The multi-clique generator omits this step.
- **Draw noise matrices E_i^A, E_i^B.** Symmetric Gaussian: $E_i^A = \frac{1}{2}(G_i^A + G_i^{A\top})$ with $G_i^A \sim N(0, 1)^{p\times p}$ (and analogously for E_i^B), then scaled by σ_A, σ_B. Defaults $\sigma_A = \sigma_B = 0.3$ (Canonical OSSA) and $\sigma_A = \sigma_B = 0.4$ (multi-clique). Sweep ranges in the appendix vary $\sigma \in [0.1, 0.5]$.

- **Symmetrize and zero the diagonal.** Set

$$B_i \;=\; \tfrac{1}{2}(\widetilde{B}_i + \widetilde{B}_i^{\top}), \qquad A_i \;=\; \tfrac{1}{2}(\widetilde{A}_i + \widetilde{A}_i^{\top}),$$

where $\widetilde{B}_i = C_i + \xi_i^B + \tau_i^B + E_i^B$ and $\widetilde{A}_i = W^\star C_i (W^\star)^\top + \xi_i^A + \tau_i^A + E_i^A$ (replace C_i by C_i^{tot} and drop the τ terms in the multi-clique case), then enforce $\mathrm{diag}(A_i) = \mathrm{diag}(B_i) = 0$.

- **Draw $W^\star$.** Sample $G \sim N(0,1)^{p\times p}$, take its (sign-corrected) QR factor Q so that $Q \sim \mathrm{Haar}(O(p))$, and set

$$W^\star \;=\; \mathrm{polar}\big((1-\alpha)\, I_p + \alpha\, Q\big) \;=\; \tilde{U}\tilde{V}^{\top}, \qquad \tilde{U}\,\Sigma\,\tilde{V}^{\top} \;=\; \mathrm{SVD}\big((1-\alpha)\, I_p + \alpha\, Q\big),$$

which equals I_p at $\alpha{=}0$ and a Haar-random orthogonal matrix at $\alpha{=}1$, and is itself orthogonal for every $\alpha \in [0,1]$. The rotation strength is swept over $\alpha \in \{0, 0.25, 0.5, 0.75, 1\}$ for OSSA, OSSA-NOW, and the support-selection baselines; the OSSA-ORACLE reference uses $\alpha \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$. The headline canonical OSSA configuration uses $\alpha{=}1$ and the multi-clique configuration uses $\alpha{=}0.6$.

D.2 Full canonical OSSA sweep curves

Figure 3 reports the full rotation and noise trajectories summarized by Table 1. The rotation sweep isolates the coordinate-mismatch setting, while the noise sweep shows that OSSA remains above the fixed-W baselines even when the planted support is observed through stronger subject-level noise.

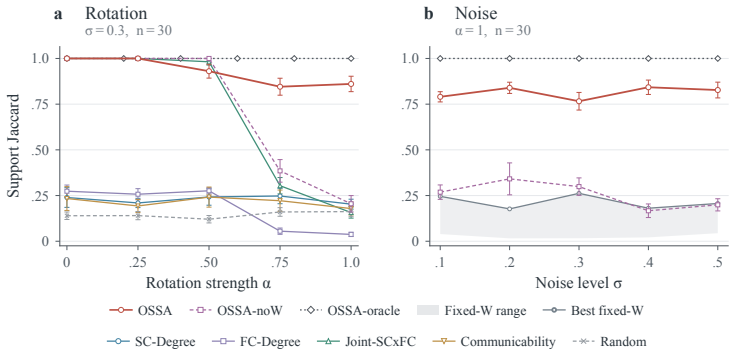


Figure 3: Canonical OSSA support recovery (Jaccard with $\mathcal{V}^\star$, mean $\pm$ s.e., 10 replicates). (a) Rotation sweep $\alpha \in [0,1]$ at $\sigma = 0.3$, $n = 30$. (b) Noise sweep $\sigma \in [0.1, 0.5]$ at $\alpha = 1$, $n = 30$. In panel (b), the grey band spans the minimum and maximum mean Jaccard values of SC-Degree, FC-Degree, Joint-SCxFC, Spectral-SC, and Spectral-FC at each noise level; the "Best fixed-W" line follows the upper boundary.

D.3 ABCD community-detection diagnostics

Figure 4 reports the training-side diagnostics used to fix the number of FC communities in the baseline ABCD cohort. The diagnostics are computed only from the training absolute-FC mean within each split and therefore do not use validation or test data. Modularity peaks at $Q = 3$, while eigengap and silhouette show a strong coarser $Q = 2$ split followed by a stable elbow at $Q = 3$. We use $Q = 3$ as the smallest stable nontrivial partition beyond the coarse two-community split, and the choice is stable across split seeds. In the experiments, communities are estimated by spectral clustering on the training mean absolute-FC matrix. The multilayer DCSBM result motivates the shared-community assumption but is not used as a literal generative claim for ABCD.

D.4 Sparsity-budget misspecification

In practice the budget κ is rarely known exactly. We sweep the supplied $\kappa \in \{4, 6, 8, 10, 12, 14, 16\}$ at planted $\kappa^\star = 10$, holding all other settings at the moderate-rotation, moderate-noise slice $p = 40$, $n = 30$, $\alpha = 0.6$, $\sigma = 0.4$ (8 replicates per supplied κ, 6 random restarts), and report Jaccard, precision, and recall against $\mathcal{V}^\star$.

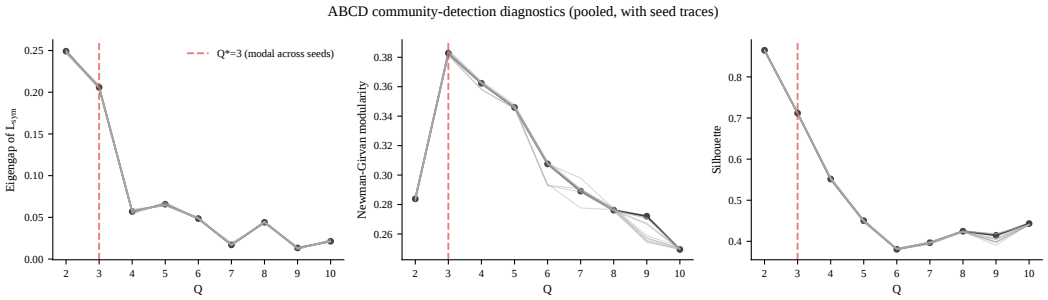


Figure 4: ABCD community-detection diagnostics on the training absolute-FC mean. Light gray traces are per-seed; dark line is the pooled training average. The diagnostics support a stable $Q = 3$ working partition: modularity peaks at $Q = 3$, and eigengap/silhouette show a coarse $Q = 2$ split with a clear elbow at $Q = 3$. No split seed selects a different working partition under this rule.

OSSA degrades gracefully on both sides of the truth (Figure 5). Under-specification ($\kappa < \kappa^\star$) yields high precision but limited recall (e.g., $\kappa = 4$: precision 1.00, recall 0.40, Jaccard 0.40); over-specification ($\kappa > \kappa^\star$) yields perfect recall and reduced precision (e.g., $\kappa = 16$: precision 0.62, recall 1.00, Jaccard 0.62). At the truth $\kappa = 10$, OSSA attains Jaccard 0.81 ± 0.06 and precision 0.89. The strongest non-OSSA baseline at $\kappa = 10$ is JOINT-SCXFC at 0.49, which tracks below OSSA across the entire grid. OSSA-NOW matches or slightly exceeds OSSA at $\alpha = 0.6$ for $\kappa \leq 10$ (where $W = I$ is a useful approximation at moderate rotation) but degrades faster on the over-specified side. For this budget-misspecification sweep, we also include SPECTRAL-SC and SPECTRAL-FC, which rank nodes by the absolute loading of the leading eigenvector of the training mean SC and FC matrices, respectively. Missing or extra budget translates into the corresponding precision/recall trade-off at the planted nodes rather than catastrophic Jaccard loss, indicating that OSSA is a usable estimator when κ is supplied heuristically.

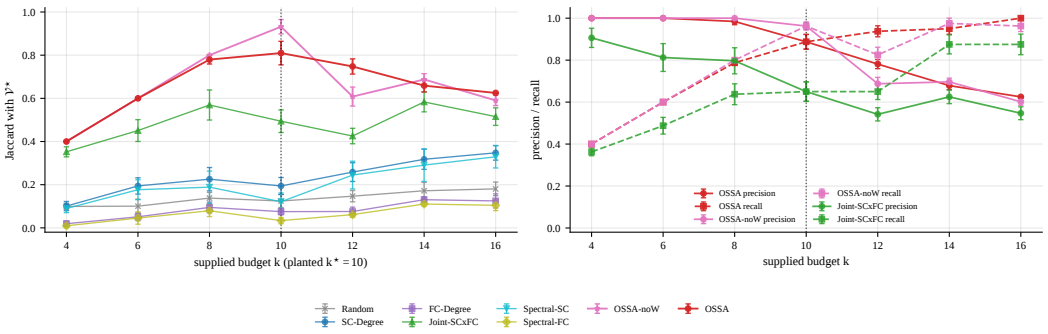


Figure 5: Robustness to sparsity-budget misspecification at $\kappa^\star = 10$, $\alpha = 0.6$, $\sigma = 0.4$ (mean $\pm$ s.e., 8 replicates). (a) Jaccard with $\mathcal{V}^\star$ vs. supplied κ. (b) Precision and recall vs. supplied κ for OSSA, OSSA-NOW, and the strongest non-OSSA baseline (JOINT-SCXFC); solid = precision, dashed = recall.

D.5 Selection stability

We assess stability of the selected support along two complementary axes on a single base dataset at $\alpha = 0.6$, $\sigma = 0.4$, $n = 30$, $\kappa^\star = 10$: per-node selection frequency across 40 bootstrap resamples (with replacement) of the subjects. Each bootstrap fit uses one random initialization, so the bootstrap curve captures both subject-resampling variability and single-start optimization variability. We also examine the distribution of pairwise Jaccard between 15 random initializations on the same (non-resampled) data for OSSA and OSSA-NOW (Figure 6b).

Bootstrap selection frequency on the 10 planted nodes is substantially higher than on non-planted nodes for OSSA (0.46 vs. 0.18; Figure 6a), consistent with concentration on $\mathcal{V}^\star$ rather than a restart-dependent local optimum. The moderate pairwise Jaccard similarity across restarts indicates sensitivity of the fitted support to initialization, consistent with the nonconvexity of the joint optimization problem. For the main experiments, we use multiple restarts and keep the run with the largest training objective. However, we note that this selection rule does not guarantee a global maximizer or eliminate variation across independently generated sets of restarts. OSSA-NOW achieves higher restart concentration at this moderate rotation ($\alpha = 0.6$, where $W = I$ is a useful approximation to $W^\star$ and the smaller-dimensional optimization is easier to stabilize) but pays the cost in recovery quality at $\alpha = 1$ (Section 3.1, Table 1).

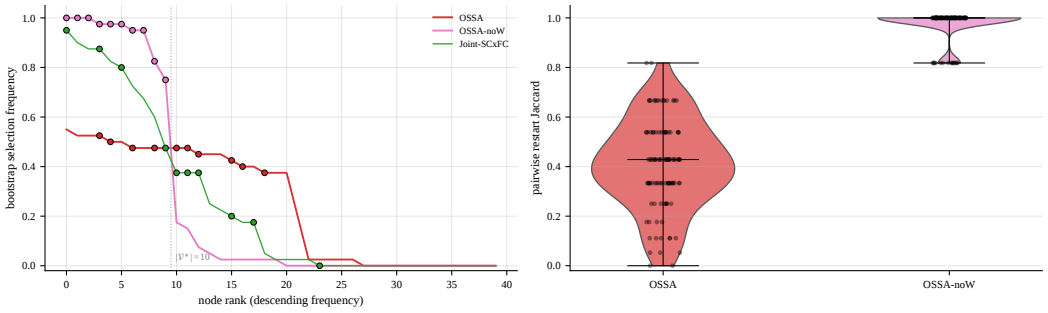


Figure 6: Selection stability of OSSA at $\alpha = 0.6$, $\sigma = 0.4$, $n = 30$. (a) Per-node bootstrap selection frequency, sorted descending; circled markers indicate planted nodes ($|\mathcal{V}^\star| = 10$). (b) Violin distribution of pairwise Jaccard between 15 random restarts on the same data, for OSSA and OSSA-NOW.

D.6 Multi-clique simulation details

This subsection gives the full design behind Section 3.2. We use one community with $p = 50$ nodes and $n = 30$ subjects. Each replicate contains K planted supports $\{\mathcal{V}_\ell^\star\}_{\ell=1}^K$, each carrying an independently sampled mixed-sign low-rank signal block. The subject-level signal is

$$C_i^{\text{tot}} = \sum_{\ell=1}^{K} C_i^{(\ell)},$$

and the observed matrices are

$$B_i = C_i^{\text{tot}} + \xi_i^B + \sigma_B E_i^B, \qquad A_i = W^\star C_i^{\text{tot}} W^{\star\top} + \xi_i^A + \sigma_A E_i^A.$$

We fix rotation strength $\alpha = 0.6$ and noise level $\sigma = 0.4$. We evaluate three planted regimes: two disjoint settings, $(K, \kappa) = (2, 10)$ and $(3, 8)$, and one overlapping setting with $K = 2$, $\kappa = 10$, and exactly four shared nodes. The overlapping nodes receive additive signal from both planted supports. No extra modality-specific nuisance variation blocks are added, because the other planted supports already act as structured distractors for each target support.

We compare OSSA-K against four size-matched competitors. SPECTRAL-SC-K: compute the top K eigenvectors of the training mean SC matrix $\bar{B}$. For component ℓ, rank nodes by $|v_\ell(j)|$ and select the top κ nodes. If a disjoint version is used, previously selected nodes are removed before the next component. COMMUNICABILITY-K clusters nodes using communicability distance. SEQUENTIAL-OSSA repeatedly fits canonical OSSA, masks the selected nodes, and refits until K supports are obtained. RANDOM-K samples K supports uniformly. Unlike OSSA-K, SEQUENTIAL-OSSA estimates a fresh rotation at each step and enforces disjoint supports by construction.

For this diagnostic simulation, λ for OSSA-K is selected on the grid reported in Appendix D.7, choosing the smallest value within one standard error of the peak best-match Jaccard. Because planted labels are available here, this λ_{best} rule is used only to summarize the recoverability envelope of the method; it is not used as a real-data tuning rule. We report best-match Jaccard after Hungarian assignment to the planted supports, together with the mean pairwise overlap of the fitted supports. All methods use six random restarts, with eight replicates for $K = 2$ and six for $K = 3$.

Table 6: Full multi-clique support recovery (mean $\pm$ s.e.). JACC: best-match Jaccard. OVL: mean pairwise overlap of fitted supports. In the two disjoint regimes all methods produce fitted overlap 0.00 ± 0.00; overlap is therefore reported only for the $r = 4$ regime.

Method	**Disjoint** K=2	**Overlap** r=4, K=2		**Disjoint** K=3
	JACC	JACC	OVL	JACC
RANDOM-K	0.150 ± 0.009	0.123 ± 0.016	0.00 ± 0.00	0.133 ± 0.012
SPECTRAL-SC-K	0.522 ± 0.048	0.368 ± 0.026	0.00 ± 0.00	0.401 ± 0.039
COMMUNICABILITY-K	0.652 ± 0.056	0.284 ± 0.023	0.00 ± 0.00	0.552 ± 0.061
SEQUENTIAL-OSSA	0.710 ± 0.040	0.448 ± 0.026	0.00 ± 0.00	0.744 ± 0.063
OSSA-K (λ_{best})	$\mathbf{0.945 \pm 0.023}$	$\mathbf{0.910 \pm 0.066}$	$\mathbf{3.75 \pm 0.56}$	$\mathbf{0.975 \pm 0.016}$

D.7 Continuous λ-trajectory

The λ_{best} values summarized in Section 3.2 and Appendix D.6 are read off a dense log-spaced grid spanning four decades. We fit OSSA-K at $\lambda \in \{0, 1, 2, 3, 5, 10, 20, 30, 50, 70, 100, 150, 200, 300, 500, 700, 1000, 2000, 3000\}$ on each of the three regimes from Section 3.2 ($K = 2$ disjoint, $K = 2$ with $r = 4$ shared nodes, $K = 3$ disjoint), using the same seeds and replicate counts as the main experiment (8 replicates for $K = 2$, 6 for $K = 3$, 6 random restarts per fit). The vertical line in each panel of Figure 7 marks the selected λ_{best} for that regime. This oracle selection is used only to summarize whether the λ-path contains a recoverability region. It is not a deployable tuning rule and is not used in ABCD.

The trajectories share a common shape: at small λ the supports collapse (high pairwise overlap, low Jaccard against the planted partition); raising λ separates the supports until fitted overlap matches the planted value (with λ_{best} near this crossing); further increases drive the supports into full disjointness. In the disjoint regimes, fitted overlap reaches zero between $\lambda = 100$ and $\lambda = 200$. For $K = 2$, Jaccard remains near 0.95 across $\lambda \in [150, 300]$, whereas for $K = 3$ it peaks at $\lambda = 150$ and declines over the same interval. In the $r = 4$ regime, fitted overlap shows an overall decrease from $\kappa = 10$ at $\lambda = 0$ to approximately $r = 4$ near $\lambda_{\text{best}} = 200$ (3.75 ± 0.56), with Jaccard peaking near this crossing and declining as larger penalties force the supports toward full disjointness.

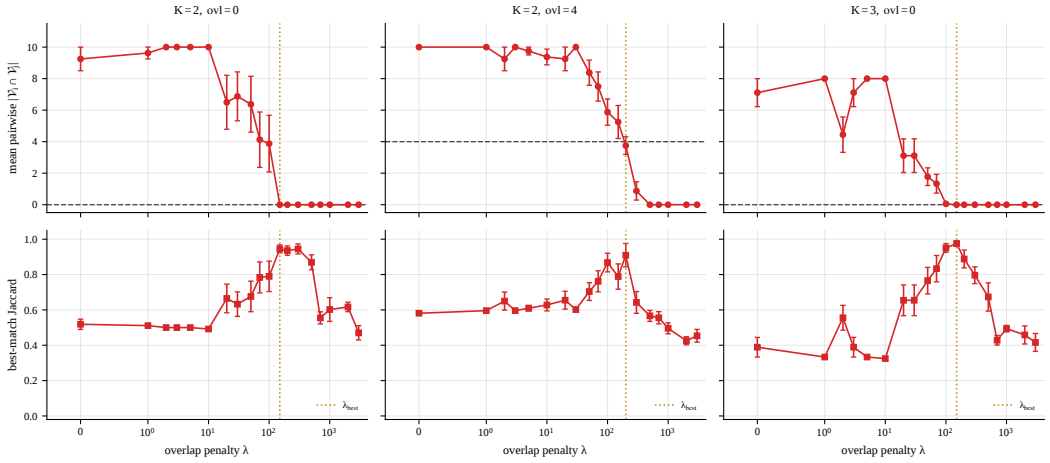


Figure 7: Continuous λ-trajectory of OSSA-K across three planted regimes (mean $\pm$ s.e.; 8 replicates for $K = 2$, 6 for $K = 3$). Top row: mean pairwise overlap of fitted supports; dashed line marks the planted overlap. Bottom row: best-match Jaccard with the planted supports. Columns: $K = 2$ disjoint, $K = 2$ with $r = 4$ shared nodes, $K = 3$ disjoint. Vertical dotted line marks λ_{best} summarized in Table 2. Full baseline values are reported in Appendix Table 6.

D.8 ABCD κ-robustness

We verify that OSSA's lead over the nine comparison methods in Table 7 is preserved at $\kappa \in \{3, 5, 8, 10\}$, not only at the design choice $\kappa = 3$. Each row summarizes the same 30 (seed, community) units. The SC-degree-quartile-matched null uses 1000 draws per unit at $\kappa = 3$ and 30 draws per unit at $\kappa \in \{5, 8, 10\}$ (Table 7).

The headline pattern is that OSSA achieves the largest mean null-corrected gain $M_{3,\text{aln}}$ at every tested budget. At $\kappa = 3$, SC-DEGREE, JOINT-SCXFC, and COMMUNICABILITY have positive mean null-corrected gains ($+0.59$ to $+0.69$), showing that these selectors achieve higher aligned correlations than the degree-matched null average. At the tested budgets $\kappa \geq 5$, SC-DEGREE, JOINT-SCXFC, COMMUNICABILITY, FC-DEGREE, and SDI have near-zero or negative mean null-corrected gains $M_{3,\text{aln}}$. OSSA, in contrast, retains $M_{3,\text{aln}}$ gain of $+0.37$ to $+0.41$ at $\kappa \in \{5, 8, 10\}$, exceeding all compared baselines at each of these budgets.

Table 7: ABCD κ-robustness: OSSA vs. nine comparison methods at $\kappa \in \{3, 5, 8, 10\}$, mean $\pm$ s.e. across 30 (seed, community) units. The table reports $M_{3,\text{aln}}$ gain, defined as $M_{2,\text{aln}}$ minus the mean SC-degree-quartile-matched null.

	$M_{3,\textbf{aln}}$ **gain**			
Method	$\kappa = 3$	$\kappa = 5$	$\kappa = 8$	$\kappa = 10$
OSSA	$\mathbf{+0.81 \pm 0.03}$	$\mathbf{+0.41 \pm 0.03}$	$\mathbf{+0.39 \pm 0.03}$	$\mathbf{+0.37 \pm 0.02}$
OSSA-NOW	$+0.60 \pm 0.05$	$+0.13 \pm 0.02$	$+0.09 \pm 0.02$	-0.03 ± 0.01
SC-DEGREE	$+0.69 \pm 0.06$	$+0.02 \pm 0.03$	-0.01 ± 0.01	-0.02 ± 0.02
JOINT-SCXFC	$+0.67 \pm 0.06$	-0.06 ± 0.03	-0.03 ± 0.02	-0.07 ± 0.01
COMMUNICABILITY	$+0.59 \pm 0.05$	-0.04 ± 0.01	$+0.00 \pm 0.01$	-0.04 ± 0.01
FC-DEGREE	$+0.06 \pm 0.09$	-0.23 ± 0.05	-0.21 ± 0.02	-0.14 ± 0.02
BRAINGNN	$+0.42 \pm 0.09$	$+0.03 \pm 0.03$	$+0.01 \pm 0.04$	$+0.03 \pm 0.03$
BRAINNETCNN	$+0.28 \pm 0.08$	-0.08 ± 0.05	$+0.04 \pm 0.04$	$+0.00 \pm 0.03$
SDI	$+0.38 \pm 0.08$	-0.01 ± 0.03	-0.11 ± 0.02	-0.06 ± 0.02
RANDOM	$+0.38 \pm 0.09$	$+0.01 \pm 0.06$	-0.06 ± 0.04	-0.02 ± 0.03

D.9 ABCD baseline W-fitting protocol

For each baseline support $\widehat{\mathcal{V}}_q^m$, the per-method alignment W_q^m is obtained by running the OSSA W-block of Algorithm 1 on training subjects with $s = \mathbf{1}_{\widehat{\mathcal{V}}_q^m}$ held fixed, using three initializations (including the identity), up to 200 W-update iterations per initialization, and monotone backtracking. The run with the largest training objective is retained. At $\kappa = 3$, the $M_{3,\text{aln}}$ random-null W_r is fit using a lighter schedule of up to 50 iterations from a single random orthogonal initialization to keep the 1000-draw null tractable per (seed, community) unit. The deep-learning baselines (BRAINGNN, BRAINNETCNN) are fit on a 3×3 (learning rate, weight decay) grid selected by validation loss.

Deep-learning baseline details. BRAINGNN and BRAINNETCNN are reimplemented in NumPy with hand-derived gradients and trained per (split-seed, community, κ) on the same train/val/test split as OSSA.

- **Input modality.** BRAINGNN: per-community training mean $\bar{D} = \frac{1}{n}\sum_i A_i \circ B_i$. BRAINNETCNN: per-subject SC submatrix B_i.
- **Prediction target.** BRAINGNN: maximize attention-pooled alignment energy $a^\top \bar{D}\, a$ ($a = \sigma(\theta)$). BRAINNETCNN: regress the z-scored scalar $y_i = \langle A_i, B_i \rangle_F$.
- **Loss.** BRAINGNN: $-a^\top \bar{D}_{\text{tr}} a$. BRAINNETCNN: mean squared error.
- **Architecture.** BRAINGNN: attention vector $\theta \in \mathbb{R}^p$, $a = \sigma(\theta)$, init $N(0, 0.1^2)$. BRAINNETCNN: 3-stage E2E$\to$E2N$\to$N2G with $W_{\text{row}}, W_{\text{col}}, W_2 \in \mathbb{R}^{p \times p}$, $W_3 \in \mathbb{R}^p$, init i.i.d. $N(0, 1/p)$.
- **Optimizer.** Adam (β_1=0.9, β_2=0.999, ε=10^{-8}), full-batch over the community.
- **Learning-rate grid.** $\eta \in \{10^{-2}, 10^{-3}, 10^{-4}\}$.

- **Weight-decay grid.** $\lambda_{\text{wd}} \in \{0, 10^{-4}, 10^{-3}\}$.
- **Batch size.** Full batch (n_{tr} subjects).
- **Epochs.** BRAINGNN: 200. BRAINNETCNN: 20.
- **Restarts and early stopping.** 16 restarts per (seed,q,κ); each draws $(\eta, \lambda_{\text{wd}})$ from the 3×3 grid by a seeded RNG. No within-restart early stopping. The best restart is selected on validation.
- **Validation criterion.** BRAINGNN: $-a^\top \bar{D}_{\text{val}} a$. BRAINNETCNN: validation MSE.
- **Saliency.** BRAINGNN: $a_v = \sigma(\theta_v)$. BRAINNETCNN: closed-form gating gradient $\text{sal}_v = \frac{1}{n}\sum_i |(W_3^\top h_i^{(2)} - y_i)(W_3)_v h_{i,v}^{(2)}|$ at $m=\mathbf{1}$.
- **Saliency $\to$ κ-node support.** $\widehat{\mathcal{V}} = \text{Top}_\kappa(\cdot)$, ties broken by ascending node index. The hard support is then passed through the same W-refit and held-out metric pipeline as every other baseline.

Table 8: Additional size-matched baseline comparisons on ABCD. Held-out $M_{1,\text{aln}}$ is reported as mean $\pm$ s.e. across 30 seed-community units. W/T/L denotes OSSA wins, ties, and losses against each baseline. These counts are descriptive because repeated splits reuse subjects.

Method	$M_{1,\text{aln}}$	OSSA W/T/L
OSSA	$\mathbf{10.55 \pm 0.85}$	—
NMF	9.96 ± 0.82	$28/1/1$
Sparse PLS	9.22 ± 1.10	$17/11/2$
Path efficiency	8.98 ± 0.71	$20/10/0$
Mean first passage time	8.32 ± 0.60	$29/0/1$
FAQ graph matching	7.57 ± 0.70	$26/4/0$
SC-to-FC ridge prediction	4.06 ± 0.49	$30/0/0$
Random	1.98 ± 0.42	$30/0/0$

Table 9: Subject-level paired bootstrap comparisons on held-out subjects from one split. $\Delta M_{1,\text{aln}}$ denotes OSSA minus the comparator. The 95% confidence intervals use 10,000 bootstrap replicates with the same resampled subject indices for all methods. Training-fitted supports and the community partition are held fixed, so the intervals exclude support-selection variability.

Comparator	Mean $\Delta M_{1,\text{aln}}$	95% CI
NMF	$+0.468$	$[0.400, 0.534]$
COMMUNICABILITY	$+0.802$	$[0.715, 0.893]$
SC-DEGREE	$+1.803$	$[1.716, 1.892]$
JOINT-SCXFC	$+1.874$	$[1.785, 1.967]$
FC-DEGREE	$+2.829$	$[2.748, 2.911]$
RANDOM	$+7.427$	$[7.292, 7.568]$

D.10 ABCD multi-clique λ-sensitivity

Diagnostic criterion. The K-support fit is not used as the default estimator in ABCD; it is a diagnostic for whether one FC community contains several dominant aligned structural cores. Since the OSSA objective is a trace alignment criterion, the primary held-out diagnostic is $M_{1,\text{aln}}$, evaluated on each fitted κ-node support after the same training-side W-refit used in Table 3. The correlation metrics $M_{2,\text{aln}}$ and $M_{3,\text{aln}}$ gain are retained as scale-free diagnostics of aligned FC–SC patterns after rotation, with degree-matched correction applied only to $M_{3,\text{aln}}$ gain; they complement but do not replace the objective optimized by OSSA.

Support geometry. Figure 8 shows a collapse-to-separation transition. For $\lambda \le 10^2$, the fitted supports nearly coincide: for $K = 2$, the mean pairwise Jaccard is 0.983 and $|U|/\kappa = 1.011$; for $K = 3$, the supports fully collapse with Jaccard 1.000 and $|U|/\kappa = 1.000$. Increasing λ first

ABCD multi-clique λ-diagnostic.
Top row: structural support behaviour; Bottom row: held-out Table~3-style metrics on each κ = 3 support

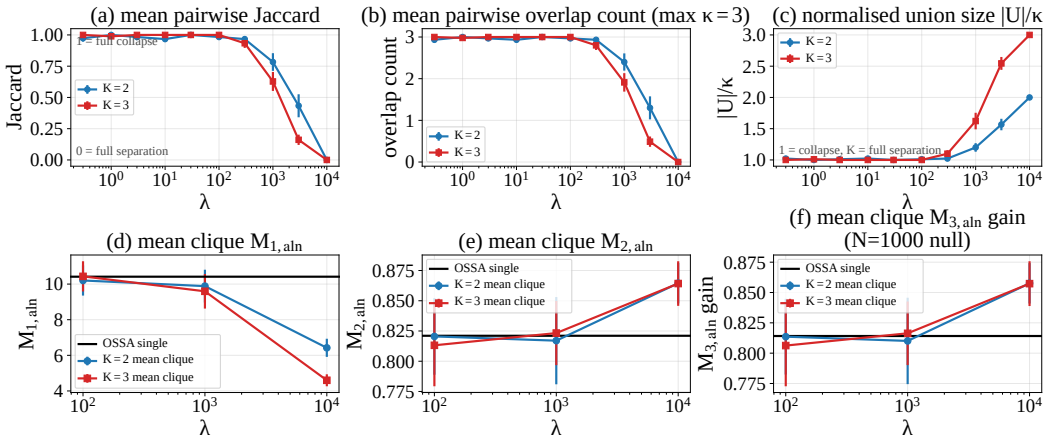


Figure 8: ABCD multi-clique λ-diagnostic. Top row: support geometry as λ varies. At small λ, the K supports collapse onto the same κ-node set. At large λ, the overlap penalty forces separation. Bottom row: held-out metrics evaluated on each $\kappa = 3$ support after the Table 3-style training-side W-refit. $M_{1,\mathrm{aln}}$ is the primary held-out OSSA objective and decreases after forced separation. $M_{2,\mathrm{aln}}$ and $M_{3,\mathrm{aln}}$ gain are scale-free correlation diagnostics and increase at large λ, indicating secondary correlation-structured motifs rather than multiple dominant high-M_1 alignment cores.

produces partial separation and then full disjointness. At $\lambda = 10^4$, the pairwise Jaccard is 0, with $|U|/\kappa = 2.000$ for $K = 2$ and 3.000 for $K = 3$. Thus the optimizer can separate the supports when forced, but the path does not reveal a stable intermediate multi-core regime.

Held-out metrics. The held-out metrics separate the primary objective from secondary correlation structure. At $\lambda = 10^2$, where the supports collapse, the mean clique-level $M_{1,\mathrm{aln}}$ is essentially the canonical OSSA value: 10.20 ± 0.85 for $K = 2$ and 10.43 ± 0.85 for $K = 3$, compared with 10.42 ± 0.80 for canonical OSSA. Once separation is forced, $M_{1,\mathrm{aln}}$ decreases: at $\lambda = 10^4$, it is 6.42 ± 0.51 for $K = 2$ and 4.60 ± 0.34 for $K = 3$. In contrast, $M_{2,\mathrm{aln}}$ and $M_{3,\mathrm{aln}}$ gain rise at $\lambda = 10^4$, reaching approximately 0.86. We interpret this as evidence for secondary scale-free FC-SC correlation motifs, not as evidence for several dominant OSSA cores, because these supports no longer remain competitive under the primary alignment-magnitude objective. Therefore, under the community construction and λ grid used here, ABCD supports a single dominant high-M_1 aligned core at $\kappa = 3$, with multi-clique OSSA serving as a diagnostic extension.

D.11 Multi-support extensions of ABCD baselines

Protocol. This auxiliary analysis checks whether the single-core conclusion is an artefact of comparing the K-support OSSA diagnostic only with single-support baselines. We extend the non-neural support selectors in Table 3 to $K \in \{2, 3\}$ supports at the same budget $\kappa = 3$. For ranking-based selectors, the K supports are consecutive disjoint κ-node blocks from the training-only node ranking. Thus $V^{(1)}$ contains the top κ nodes, $V^{(2)}$ contains the next κ nodes after removing $V^{(1)}$, and so on. RANDOM-K samples K disjoint κ-node blocks uniformly within each community and averages over $R = 200$ draws per seed-community unit. OSSA-K and OSSA-noW-K are evaluated on the same overlap-penalty grid $\lambda \in \{10^2, 10^3, 10^4\}$. Each fitted κ-node support is then evaluated by the same fixed-support training-side W-refit and held-out metrics used in Table 3. We report mean-clique metrics, averaging over the K supports within each seed-community unit, so that a larger K is judged by per-support yield rather than by a mechanically larger union.

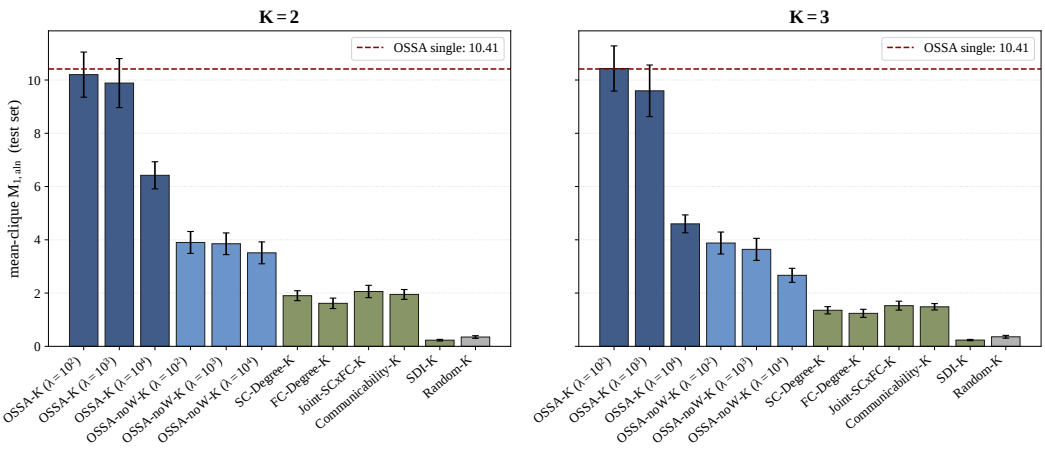


Figure 9: ABCD multi-support extensions of the Table 3 baselines, evaluated by the primary held-out objective $M_{1,\mathrm{aln}}$. Bars report mean-clique $M_{1,\mathrm{aln}}$, averaging over the K fitted $\kappa = 3$ supports within each seed-community unit; error bars are standard errors over the 30 seed-community units. The dashed line shows the single-support OSSA reference used in this multi-support diagnostic. OSSA-K remains close to this value only when supports collapse or partially overlap; forced separation lowers the primary alignment magnitude. Ranking-based multi-support extensions are substantially lower.

Primary held-out alignment. Figure 9 shows that extending the baselines to several supports does not produce multiple competitive high-M_1 cliques. OSSA-K remains close to the canonical single-support value only in collapsed or partially collapsed regimes. At $\lambda = 10^2$, the $K = 2$ fit has mean-clique $M_{1,\mathrm{aln}} = 10.20 \pm 0.85$, but its supports nearly coincide ($|U|/\kappa = 1.01$, mean pairwise Jaccard 0.98); the $K = 3$ fit fully collapses and gives $M_{1,\mathrm{aln}} = 10.43 \pm 0.85$. When the overlap penalty forces disjoint supports at $\lambda = 10^4$, mean-clique $M_{1,\mathrm{aln}}$ drops to 6.42 ± 0.51 for $K = 2$ and 4.60 ± 0.34 for $K = 3$. The ranking-based K-support baselines are lower throughout: their best mean-clique $M_{1,\mathrm{aln}}$ is 2.06 ± 0.23 for $K = 2$ and 1.53 ± 0.17 for $K = 3$. Thus forcing multiple disjoint supports does not appear beneficial as a dimension-reduction step in ABCD.

Support-index degradation. For ranking-based extensions, the support index makes the same pattern more explicit. Figure 10 shows that, for SC-DEGREE, FC-DEGREE, JOINT-SCXFC and COMMUNICABILITY, $V^{(1)}$, the first κ-node block, corresponds to the original single-support baseline, whereas the later supports have much smaller held-out $M_{1,\mathrm{aln}}$. This degradation is expected when the FC-derived communities are already small: after the first high-signal support has been removed, the remaining disjoint blocks do not provide comparable high-M_1 structural cores.

Scale-free diagnostics. The scale-free diagnostics tell a complementary story rather than changing the primary conclusion. As in Appendix D.10, high-λ separated OSSA-K fits can increase $M_{2,\mathrm{aln}}$ or $M_{3,\mathrm{aln}}$ gain, but these metrics measure within-support FC-SC correlation after scale normalization. They do not replace the trace-alignment magnitude optimized by OSSA. We therefore interpret the high-λ scale-free gains as secondary correlation-structured motifs, not as evidence for several dominant high-M_1 aligned cores.

D.12 Structural connectivity within consensus supports

Table 4 in the main text reports the subject-averaged SC edge weights and their degree-matched reference ratios. Each consensus support forms a within-lobe triangle. The temporal and occipital supports lie in the right hemisphere, whereas the frontal support includes the left medial orbitofrontal region and bilateral frontal poles.

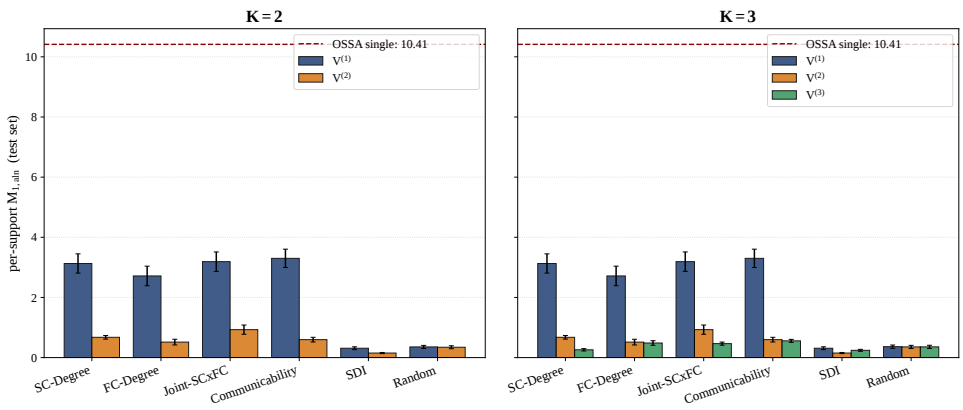


Figure 10: Per-support degradation for ranking-based baseline extensions with $K \in \{2, 3\}$. Each method forms K consecutive disjoint $\kappa = 3$ blocks from its training-only ranking. $V^{(1)}$ corresponds to the original single-support baseline; for SC-DEGREE, FC-DEGREE, JOINT-SCXFC and COMMUNICABILITY, the later supports have much smaller held-out $M_{1,\mathrm{aln}}$. The dashed line shows the single-support OSSA reference used in this multi-support diagnostic.

D.13 Empirical diagnostics of the fitted alignment

At $\kappa = 3$, the normalized squared projector distance $\|\widehat{W} P_{\widehat{V}} \widehat{W}^\top - P_{\widehat{V}}\|_F^2/(2\kappa)$ has community-wise means ranging from 0.54 to 0.67 over ten split seeds. Reference means based on $1{,}000$ Haar-random orthogonal matrices of each community's dimension range from 0.80 to 0.90. All 30 seed–community units lie below the 2.5th percentile of their corresponding Haar reference distribution.

For regional loadings $\ell_a = \sum_{j \in \widehat{V}} \widehat{W}_{aj}^2$, we quantify concentration using the participation ratio $\mathrm{PR} = (\sum_a \ell_a)^2 / \sum_a \ell_a^2$, which measures the effective number of contributing regions. Community-wise mean PR values over ten split seeds are approximately 5–10, compared with 10–19 under the Haar reference. The corresponding mean peak regional loadings, $\max_a \ell_a$, are approximately 0.59–0.69.

We measured the fraction of $\|\widehat{W} P_{\widehat{V}} \widehat{W}^\top\|_F^2$ contributed by matrix entries corresponding to contralateral homologous regions. Across the three communities, its means over ten split seeds are approximately 3.2–8.9 times the corresponding Haar-reference means.

For each community, we compared the fitted projectors $\widehat{W} P_{\widehat{V}} \widehat{W}^\top$ across all 45 pairs of the ten split seeds using squared Frobenius distance divided by 2κ. Mean distances were 0.0002, 0.113 and 0.0001 for the frontal, temporal and occipital communities, respectively. The observed stability is consistent with reproducibility of the spectral structure of mean training FC and does not by itself establish biological specificity.

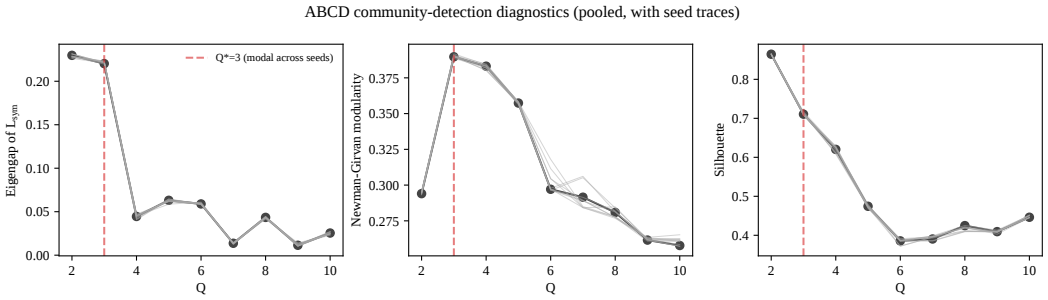


Figure 11: ABCD second-visit community-detection diagnostics. Light gray traces are per-seed; the dark line is the pooled training average. The eigengap of L_{sym}, Newman-Girvan modularity, and silhouette all peak at $Q^* = 3$, with no seed disagreeing, identical to the baseline cohort.

D.14 Sensitivity to community construction

Table 10: Sensitivity to community construction using the same training mean $|FC|$. Node recovery summarizes retention of the original nine selected regions under each construction. Held-out M_1 values are not directly comparable across Q, because community size changes the number of candidate edges.

Community construction	Sizes	Node recovery	Held-out M_1
$Q = 2$	52, 16	0.50	13.83
$Q = 3$, same partition refit (control)	30, 23, 15	0.89	10.30
$Q = 4$	17, 18, 14, 19	0.56	7.74
Spectral	22, 15, 31	0.61	10.71
Ward hierarchical	6, 20, 42	0.61	10.04
Louvain	24, 22, 22	0.67	8.85
Newman leading eigenvector	28, 17, 23	0.44	9.66

E ABCD second-visit replication details

This appendix documents the second-visit replication summarised in Section 4 on the ABCD Year-2 follow-up ("second-visit") release. The protocol is identical to the baseline (Year-1) evaluation: no hyperparameter is retuned, no seed is reshuffled, and the same per-method W-fit budget, multi-clique grid, and aggregation rules are reused.

Data and pipeline. The ABCD second-visit release contributes 1,909 subjects parcellated by the same 68-region Desikan-Killiany atlas as the baseline [11]. The canonical preprocessing pipeline (transposition to (N, p, p), exact symmetrization, diagonal enforcement, Fisher-z then non-negative clipping for FC, and $\log(1 + \cdot)$ for SC) and the cohort-median 5σ QC rule are applied without modification. QC drops 12 subjects (none with simultaneous FC and SC flags) and leaves $n_{\mathrm{ABCD2}} =$ 1,897 subjects. The $60/20/20$ deterministic splits over the same 10 seeds yield $n_{\mathrm{train}} = 1{,}138$ and $n_{\mathrm{val}} = n_{\mathrm{test}} \in \{379, 380\}$ per seed. The one-time test-evaluation rule of the baseline protocol is reused: each (method, metric, split) tuple is touched exactly once on the second-visit cohort.

Community partition. Spectral diagnostics on the training absolute-FC mean again select $Q^* = 3$ in every one of the 10 split seeds. The pooled training average likewise yields $Q_{\mathrm{pool}} = 3$. Cross-seed ARI on the modal-Q partitions is 0.99 ± 0.01 (min 0.96 across the 45 seed pairs), indicating near-perfect agreement up to label permutation (Figure 11). The pooled community sizes $(31, 23, 14)$ are within one node of the baseline pooled sizes $(30, 23, 15)$, supporting that the same three-community organisation is recovered at the second visit.

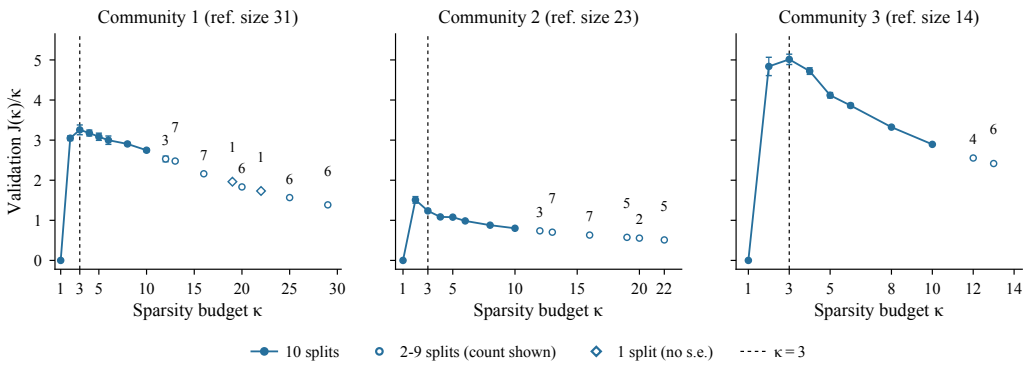


Figure 12: ABCD second-visit sparsity sweep after matching communities across splits to the pooled partition by maximum node overlap. Points show means and error bars show s.e.; counts below 10 are labeled, and single-split results have no error bar. Larger budgets may use different split subsets. The $\kappa = 1$ point is a trivial zero-objective floor. For tested budgets $\kappa \geq 3$, mean validation $J(\kappa)/\kappa$ decreases in all three communities.

Table 11: ABCD second-visit canonical OSSA evaluation, mean $\pm$ s.e. across 30 (seed, community) units, $\kappa = 3$, on $n_{\mathrm{ABCD2}} = 1{,}897$ subjects. Conventions match Table 3: $M_{3,\mathrm{aln}}$ gain is the difference between the method's $M_{2,\mathrm{aln}}$ and the mean of $N = 1000$ SC-degree-quartile-matched random nulls; $M_3\,p$ is mean (median) over the 30 units. The headline summary appears in Section 4 ("Second-visit replication"), with this appendix providing per-method detail.

Method	$M_{1,\mathrm{aln}}$	$M_{2,\mathrm{raw}}$	$M_{2,\mathrm{aln}}$	$M_{3,\mathrm{aln}}$ gain	$M_3\,p$
OSSA	$\mathbf{10.16 \pm 0.87}$	$+0.494$	$\mathbf{+0.811}$	$\mathbf{+0.81}$	$0.073\,(0.042)$
OSSA-NOW	3.23 ± 0.30	$+0.514$	$+0.539$	$+0.54$	$0.211\,(0.166)$
COMMUNICABILITY	3.19 ± 0.28	$\mathbf{+0.672}$	$+0.672$	$+0.67$	$0.145\,(0.083)$
JOINT-SCXFC	3.11 ± 0.31	$+0.623$	$+0.623$	$+0.62$	$0.173\,(0.084)$
SC-DEGREE	3.02 ± 0.31	$+0.652$	$+0.652$	$+0.65$	$0.158\,(0.044)$
FC-DEGREE	2.27 ± 0.31	-0.124	-0.124	-0.13	$0.549\,(0.633)$
BRAINGNN	0.64 ± 0.17	$+0.285$	$+0.324$	$+0.32$	$0.324\,(0.291)$
RANDOM	0.51 ± 0.14	$+0.260$	$+0.248$	$+0.25$	$0.370\,(0.406)$
BRAINNETCNN	0.37 ± 0.12	$+0.328$	$+0.432$	$+0.43$	$0.275\,(0.207)$
SDI	0.07 ± 0.02	$+0.257$	$+0.322$	$+0.32$	$0.322\,(0.315)$

Per-community budget. The per-community budget $\kappa = 3$ from the baseline cohort is reused without retuning. Figure 12 summarizes validation $J(\kappa)/\kappa$ after matching communities across split seeds to the pooled reference partition by maximum node overlap. Among the tested budgets $\kappa \geq 3$, the mean is highest at $\kappa = 3$ in each community, consistent with the baseline results.

Canonical OSSA evaluation. The summary appears in Section 4 ("Second-visit replication"); this appendix provides per-method numbers and per-community decomposition (Table 11). At $\kappa = 3$ over 30 (seed, community) units, OSSA attains $M_{1,\mathrm{aln}} = 10.16 \pm 0.87$, with the ablation OSSA-NOW (3.23 ± 0.30) and the classical baseline SC-DEGREE (3.02 ± 0.31) clustering at roughly one third of OSSA's value. The $M_{2,\mathrm{raw}} \to M_{2,\mathrm{aln}}$ shift again shows that high-degree or SC-summary baselines can look competitive in raw coordinates (e.g., COMMUNICABILITY: $+0.672$, SC-DEGREE: $+0.652$), whereas OSSA leads after rotation-aware evaluation ($+0.811$ vs. next-best $+0.672$). The same conclusion holds on the SC-degree-quartile-matched null gain $M_{3,\mathrm{aln}}$ ($+0.81$ vs. next-best $+0.67$). After matching each split's community labels to the pooled partition by maximum node overlap, OSSA leads on $M_{1,\mathrm{aln}}$ in all three communities. For pooled community sizes $(31, 23, 14)$, its scores are 10.17 ± 0.26, 4.32 ± 0.04, and 15.98 ± 0.07, respectively (mean $\pm$ s.e. across ten splits), versus highest baseline means of 3.66, 1.11, and 5.00.

Table 12: Cross-visit support replication between the ABCD baseline and Year-2 follow-up releases for canonical $\kappa = 3$ supports. Results are averaged over 30 (seed, community) units. CVSR denotes the support-set Jaccard overlap between visits. nCVSR is the null-corrected replication score relative to size-matched or SC-degree-quartile-matched random supports.

Method	Overlap	Fraction	CVSR	nCVSR		Exact	≥ 2	Null p
	count	$(/\kappa)$	(Jaccard)	size-matched	degree-matched	rate	rate	median
OSSA	2.86±0.08	0.95±0.03	0.928±0.033	0.921±0.034	0.935±0.031	0.85	1.00	0.022
JOINT-SCxFC	3.00±0.00	1.00±0.00	1.000±0.000	1.000±0.000	1.000±0.000	1.00	1.00	0.019
FC-DEGREE	2.87±0.06	0.96±0.02	0.933±0.032	0.929±0.034	0.910±0.043	0.87	1.00	0.019
COMMUNICABILITY	2.83±0.07	0.94±0.02	0.917±0.035	0.903±0.040	0.877±0.051	0.83	1.00	0.017
SC-DEGREE	2.83±0.07	0.94±0.02	0.917±0.035	0.909±0.038	0.893±0.044	0.83	1.00	0.035
RANDOM	1.73±0.20	0.58±0.07	0.503±0.072	0.455±0.079	0.430±0.080	0.33	0.57	0.205
SDI	1.73±0.11	0.58±0.04	0.433±0.038	0.370±0.045	0.370±0.040	0.07	0.67	0.058
OSSA-NOW	1.50±0.21	0.50±0.07	0.430±0.073	0.361±0.082	0.349±0.084	0.30	0.40	0.498
BRAINGNN	0.97±0.21	0.32±0.07	0.270±0.067	0.194±0.073	0.201±0.072	0.17	0.27	0.813
BRAINNETCNN	0.47±0.12	0.16±0.04	0.103±0.029	0.015±0.031	0.020±0.030	0.00	0.10	1.000

Cross-visit support replication To evaluate whether the selected structural supports remain stable across longitudinal neurodevelopment, we compare the canonical $\kappa = 3$ supports identified from the baseline and Year-2 ABCD releases. For each split seed and matched FC community, we compute a cross-visit support replication score

$$\mathrm{CVSR} = \frac{|\widehat{V}_{\mathrm{base}} \cap \widehat{V}_{\mathrm{Y2}}|}{|\widehat{V}_{\mathrm{base}} \cup \widehat{V}_{\mathrm{Y2}}|},$$

corresponding to the Jaccard overlap between the baseline and second-visit supports. We additionally report the overlap count, overlap fraction, exact replication rate, and a degree-matched null-corrected replication score (nCVSR). Community labels between visits are aligned by maximum-overlap matching prior to evaluation.

Table 12 shows that OSSA remains highly stable across visits, with CVSR $= 0.928 \pm 0.033$ and mean overlap $2.86/3$ nodes. Importantly, this stability is achieved simultaneously with the substantially larger held-out alignment scores reported in Tables 3 and 11. Several graph-summary baselines, including SC-DEGREE, FC-DEGREE, COMMUNICABILITY, and JOINT-SCXFC, also exhibit high cross-visit support consistency, suggesting that portions of the dominant structural organization are preserved longitudinally within FC communities. In contrast, the deep-learning baselines exhibit substantially weaker support replication and near-null corrected stability.

Multi-clique sensitivity (λ-diagnostic). We replay the same $K \in \{1, 2, 3\}$ and log-spaced λ-sweep at $\kappa = 3$ on the second-visit cohort over the same 30 (seed, community) units (Figure 13); the summary appears in Section 4 ("Second-visit replication"). The diagnostic returns the same single-core signature as the baseline cohort: at $\lambda \leq 10^2$ the $K = 2$ and $K = 3$ fits collapse onto a single dominant support (mean pairwise Jaccard ≥ 0.98, normalised union ≤ 1.03); at $\lambda \in [10^3, 10^4]$ a sharp transition forces them into geometric expansion of the union ($K = 2$: $1.0 \to 2.0$; $K = 3$: $1.0 \to 2.97$), with no intermediate λ yielding stable partial separation.

F Compute resources

All experiments were run on a high-performance computing cluster. CPU jobs used Intel Xeon 6342 processors, with 4 CPU cores and 16 GB of RAM allocated per job (4 GB per core). GPU resources were used only for the deep-learning baselines, which ran on an RTX 8000 GPU. Runtime was measured from loading the preprocessed SC/FC matrices to saving the final metrics and figures. Raw MRI preprocessing is excluded. Table 13 reports peak memory usage (GB) and wall-clock time (minutes) for the main experiment groups.

G Limitations and broader impact.

Our theory establishes local statistical guarantees for canonical OSSA, including the existence of an empirical local maximizer near the population target. Note that the implemented block ascent

Table 13: Compute resources for the main reported experiments.

Experiment group	Worker type	Peak memory (GB)	Wall-clock time (min)
Simulation experiments	CPU	0.87	13
ABCD main evaluation	CPU	1.23	57
Random-null/κ-robustness/multi-clique diagnostics	CPU	0.62	5
Deep-learning baselines	GPU	1.23	45
ABCD second-visit replication	CPU	1.06	48

method is a finite multi-restart local solver, and we do not claim that its output lies in the specified neighborhood. Extending guarantees to the multi-support objective is future work. Empirically, the results are limited to the Desikan–Killiany atlas, ABCD, predefined FC communities, and small support budgets, so the anatomical supports should be treated as cohort- and pipeline-dependent hypotheses until replicated across parcellations, acquisition protocols, and clinical populations. OSSA may support more interpretable studies of brain development and disease-related network organization, but connectome-derived supports should not be used for individual diagnosis, screening, or group profiling without external validation and uncertainty reporting.

The community-wise formulation can also be applied to predefined overlapping functional communities. Each community would retain its own alignment and support parameters, so the same objective could be fitted independently within each community, with a shared region eligible for selection in more than one. However, support selections across communities would remain uncoupled, and an SC edge included in multiple community subproblems could contribute fully to each corresponding objective. Such independent fits therefore do not describe how a shared region's structural role is distributed across its functional memberships. Whether the relevant biological structure is best described by independent community-specific cores or by a decomposition of shared structural roles across systems remains an open question that the current formulation cannot resolve. Jointly modeling these shared roles is a direction for future work.

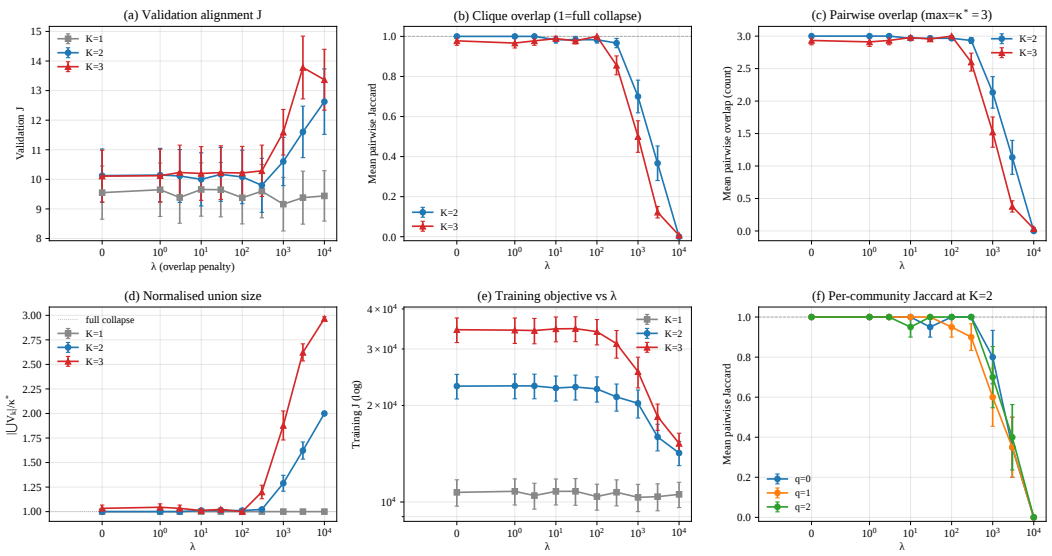


Figure 13: ABCD second-visit multi-clique λ-sensitivity on a log-spaced grid (mean $\pm$ s.e. over 30 (seed, community) units; x-axis on a symmetric-log scale to display $\lambda = 0$). Panels show: (a) Validation J evaluated on the union of supports; (b) Mean pairwise Jaccard (1 = full collapse); (c) Mean pairwise overlap count (maximum $\kappa = 3$); (d) Normalised union size $|\bigcup_\ell \widehat{\mathcal{V}}^{(\ell)}|/\kappa$ (1 = full collapse, K = full separation); (e) Training J on log scale; (f) Per-community pairwise Jaccard at $K = 2$. The single-core verdict (collapse at small λ, no per-node yield gain at large λ) holds in every community, replicating the baseline-cohort conclusion.

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 introduction state the OSSA formulation. The learned orthogonal alignment, the sparse node support, the multi-support extension, theoretical guarantees, simulations, and ABCD evaluation are supported in Sections 2–4 and Appendices B–E.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

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

   Answer: [Yes]

   Justification: Section 5 discusses the main limitations: the empirical conclusions are based mainly on one atlas and one developmental cohort, with fixed community partitions and

small support budgets, so the anatomical findings should be interpreted as cohort- and preprocessing-dependent until replicated across datasets, parcellations, and clinical populations.

Guidelines:

- The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

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 main theoretical result is Theorem 1, whose assumptions are stated as (A1)-(A6) in appendix B. Appendix B also provides the proof, auxiliary lemmas, and sufficient population conditions.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

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 experiments are described with the simulation generator, ABCD preprocessing, train/validation/test protocol, baselines, metrics, tuning rules, and optimization budgets in Sections 3–4 and Appendices C–E.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

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: Code is provided in the supplementary material. ABCD data access is subject to the corresponding data-use terms.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.

- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

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: Sections 3–4 and Appendices C–E specify the preprocessing, deterministic data splits, community selection, sparsity budgets, baseline definitions, validation choices, optimization constants, restarts, and deep-learning baseline training details.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

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: The paper reports means and standard errors over simulation replicates or ABCD seed-community units. The ABCD summaries are descriptive because the seed-community units are not independent samples.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
   - If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

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: Appendix F and Table 13 report the compute worker type, peak memory, and wall-clock time for simulation experiments, ABCD main evaluation, random-null/robustness/multi-clique diagnostics, deep-learning baselines, and second-visit replication.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

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 work is methodological, uses established connectome data, and does not describe any procedure that deviates from the NeurIPS Code of Ethics.

Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

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

Answer: [Yes]

Justification: Appendix G discusses both positive societal impact, namely supporting more interpretable studies of brain development and disease-related network organization through localized structure-function coupling, and negative impact, including the risks of misleading biological claims or unfair profiling if connectome-derived supports are over-interpreted or used outside validated research settings.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.

- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

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 paper does not release a high-risk pretrained model, generative model, scraped dataset, or other asset with an obvious misuse pathway requiring special release safeguards.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

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 ABCD data were obtained through the official ABCD data access mechanism and used in accordance with the applicable data-use agreement. The Desikan-Killiany atlas and all external datasets/software assets are cited in the manuscript. Code is provided in the supplementary material.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

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 manuscript does not announce the release of a new dataset, model checkpoint, or software package as a new asset; it introduces a method and reports experiments.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

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: The authors do not conduct new crowdsourcing or participant-facing experiments. The empirical analysis uses existing ABCD connectome data, so no new participant instructions, screenshots, or compensation details apply.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

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: [Yes]

Justification: The ABCD data were obtained through the official ABCD data access mechanism and used in accordance with the applicable data-use agreement. ABCD participants were enrolled under the study's institutional review and consent/assent procedures, and our analysis used de-identified connectome-derived data only. We do not attempt individual identification and do not release subject-level neuroimaging data.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.

- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

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: LLM use was limited to assistance with writing, code development, and script execution. LLMs were not components of the core research methodology.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.