Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Recursive Transport Topology: Predictive and Causal Organization in Protein Fitness Landscapes

Jeffery Scott Allbright

scott.allbright@yahoo.com

AIIA Technologies, Inc

Research Article

Keywords: recursive indexing, transport topology, protein fitness landscapes, graph geometry, admissibility constraints, phenotype locality, recursive compression, fragmentation, GFP, GB1, representation learning

Posted Date: June 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9889139/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare potential competing interests as follows: Jeffery Scott Allbright reports pending intellectual property related to the theoretical and computational frameworks described in this work and is a co-founder of Forma Substrate, Inc. and AIIA Technologies, Inc., which may commercialize these ideas. The author declares no other competing interests.

Abstract

Biological sequence-function relationships are commonly modeled through local sequence similarity, engineered features, or learned embeddings. Here we evaluate whether admissible transformations can organize protein fitness landscapes into recursive transport geometry before manifold formation. In this Phase I study, we analyze GFP ortholog systems and the GB1 fitness landscape using admissibility-constrained recursive indexing. Across GFP systems, recursive basins preserved phenotype-local structure without using phenotype labels for basin formation, including ~ 38% node compression, ~ 60% edge reduction, and near-perfect path preservation in ppluGFP. Transfer-compatible GFP systems preserved meaningful recursive transfer at low capacity (~ 2–4 dimensions), whereas incompatible systems required higher capacity and exhibited perturbation-sensitive fragmentation. In GB1, dense redundant topology supported robust rerouting rather than GFP-like fragmentation, indicating that transport behavior depends strongly on topology regime. Synthetic topology interventions further showed causal control within constructed transport graphs: clustering enhancement reduced maximum edge betweenness from ~ 0.0263 to ~ 0.00005, nearly eliminating bottleneck dominance. Across analyses, redundancy, rerouting capacity, bottleneck concentration, and spectral cohesion predicted robust, transitional, and fragmentation-sensitive phase regions. These results support a predictive and intervention-capable recursive transport-topology framework for biological organization and suggest that pre-manifold constraint-consistent indexing can induce measurable, compressible, and manipulable transport structure with relevance to biological modeling and artificial representation learning.

Introduction

Protein fitness landscapes are often represented through raw sequence similarity, engineered features, learned embeddings, or supervised models trained directly on measured phenotypes. These approaches are powerful, but they often ask downstream optimization to discover which transitions are lawful, which neighborhoods are biologically meaningful, and which pathways preserve function.

Recursive Indexing (RIX) starts from a different premise: useful geometry may be downstream of indexing. Instead of treating biological variants as unconstrained symbols, RIX reorganizes them into constraint-consistent transport neighborhoods before manifold learning or prediction. The central question of Phase I is whether admissibility-constrained organization produces measurable compression, phenotype locality, transfer sensitivity, perturbation robustness, and topology-defined phase behavior.

More generally, RIX treats representation quality as an operational alignment problem: a representation is useful to the extent that it preserves the relevant properties, links, constraints, and admissible transformations of the system being represented. From this perspective, indexing is not a neutral labeling step. It is the first representational act that determines which relationships remain visible, which transitions remain lawful, and which geometry downstream learning systems are asked to recover.

This framing creates a bridge to artificial intelligence. If valid transformations shape representations before embedding, learning systems may benefit from pre-manifold structure rather than relying entirely on scale and optimization to recover relational organization after the fact. Protein landscapes provide a tractable empirical setting because constraints, phenotypes, transfer regimes, and perturbations can be measured.

Primary contributions

- Define recursive admissibility organization as a pre-manifold indexing layer that transforms raw sequence variation into constrained transport topology.
- Show that admissibility-constrained organization preserves navigability and phenotype locality without using phenotype labels to form recursive basins.
- Separate local recursive organization from cross-system transfer alignment, showing that internally organized systems can still be externally misaligned.
- Identify global recursive transport geometry, rather than local similarity alone, as the layer most associated with transfer compatibility.
- Introduce topology variables - redundancy, rerouting capacity, bottleneck dominance, and spectral cohesion - that predict and causally influence perturbation behavior in constructed transport graphs.
- Map robust, transitional, and fragmentation-sensitive transport phase regions with implications for representation learning.

Methods Overview

Phase I used graph-based recursive transport analyses across three classes of systems: toy non-biological domains, GFP ortholog fitness landscapes, and the GB1 combinatorial protein fitness landscape. The goal was not to maximize predictive leaderboard performance, but to test whether admissibility-constrained indexing produces measurable organization before downstream manifold learning.

Raw variants were represented as graph nodes. Edges encoded local accessibility or admissible transitions. In the biological analyses, admissibility was operationalized using mutation position and amino-acid compatibility structure. The simplest recursive basin rule grouped variants by mutation position plus coarse biochemical class, creating an unsupervised pre-manifold organization layer. Phenotype labels were not used to construct these basins.

Operationally, recursive indexing is defined here as the repeated use of preserved admissibility structure to organize, compress, and evaluate higher-order transport geometry. In the protein analyses below, preserved structure corresponds to admissible sequence-function neighborhoods; recursive compression turns those neighborhoods into higher-order transport basins that can be tested for navigability, phenotype locality, transfer alignment, and perturbation robustness.

Operational definitions

Term	Operational meaning in Phase I
Admissible transition	An edge permitted by a predefined compatibility rule, such as mutation position plus amino-acid class or local sequence accessibility.
Recursive basin	A contracted group of variants sharing admissible local structure; in the GFP analyses, the primary rule was mutation position plus coarse amino-acid class.
Recursive transport geometry	The graph geometry induced after admissible transitions and recursive basins define pathways for traversal, compression, perturbation, or transfer.
Transfer retention	A source-frozen recursive traversal metric interpreted relative to shuffled, flattened, and incompatible-regime controls rather than as absolute prediction accuracy.
Recursive capacity threshold	The estimated minimum recursive representational capacity required to preserve a target transfer or phenotype-continuity signal.
Bottleneck dominance	Concentration of transport flow through limited pathways, measured here primarily by maximum edge betweenness.
Spectral cohesion	A global connectivity/stability proxy measured by graph spectral gap.
Fragmentation-sensitive regime	A topology regime in which perturbation reduces connectivity, spectral cohesion, or transport continuity.

Analysis workflow and controls

The analysis workflow had five steps: construct admissible transport graphs; recursively compress compatible neighborhoods into basins; measure preservation of navigability, phenotype locality, and transfer; perturb or ablate graph structure; and map topology-defined robustness, bottleneck, and phase behavior. This workflow allowed local organization, global transport alignment, and perturbation robustness to be evaluated separately.

Primary metrics included path fidelity, path ranking, compression retention, within-basin phenotype dispersion, recursive transfer capacity thresholds, geodesic distortion, simplified Forman-style graph curvature, edge betweenness, spectral gap, perturbation drift, and phase-region assignment. Randomized controls included matched basin shuffles, topology perturbations, phenotype shuffles, neighborhood disruption, hierarchy flattening, and synthetic topology interventions. Reproducibility materials report node/edge counts, random-control counts, fixed seeds, and available confidence intervals or empirical permutation summaries for the main numerical claims.

Methods summary table

Analysis	Graph nodes	Edge / topology rule	Controls	Primary metrics
Toy domains	3x3 navigation graphs and symbolic correction systems	Legal transitions or lawful correction transitions	Matched random edges, random coarsening	Path fidelity, ranking, compression retention
GFP compression	Observed GFP variants plus reference state	Mutation-position and amino-acid-class admissibility	Randomized recursive basins and phenotype shuffles	Compression, path distortion, phenotype dispersion
GFP transfer	Ortholog recursive graphs	Source-frozen recursive traversal geometry	Neighborhood shuffles, hierarchy flattening, phenotype randomization	Transfer retention, geodesic distortion, hierarchy benefit
GB1 regime analysis	GB1 combinatorial variants	Local sequence accessibility with recursive admissibility compression	Perturbation sweeps and randomized controls	Connectivity, phenotype drift, curvature variance
Synthetic topology	Constructed sparse and dense transport graphs	Controlled changes to redundancy, clustering, and bottleneck pressure	Baseline vs modified topology	Edge betweenness, spectral gap, phase region

Causal-language guardrail: Synthetic interventions are causal with respect to constructed transport topology. They should not yet be interpreted as direct experimental perturbations of protein function, evolution, or cellular biology.

Results

Phase I progression

The Phase I program progressed from validating constraint-consistent indexing to a predictive and causal transport-topology framework. The key transition was from asking whether lawful indexing is measurable to asking whether transport behavior can be predicted and manipulated from topology.

Phase	Milestones	Core discovery
Recursive indexing validity	2–3	Admissibility-constrained recursive geometry forms measurable biological organization
Transfer geometry	4–6	Global transport geometry governs transfer compatibility
Compression and efficiency	7	Compatible systems preserve transfer under extreme recursive compression
Fragmentation boundaries	8	Transport collapse occurs through measurable fragmentation transitions
Comparative regimes	9	Transport behavior depends on topology regime
Predictive and causal topology	10	Transport behavior is predictive, causal, and phase-boundary organized

This Phase I study moves from lawful indexing and recursive biological compression to transfer geometry, fragmentation boundaries, comparative regimes, causal topology intervention, and measurable phase regions.

Admissibility-constrained recursion preserves navigability and phenotype locality

Early benchmarks established that lawful indexing was computable and ablatable: legal-transition indexC preserved path structure better than matched random controls, and meaningful recursive coarsening retained more relational structure than random coarsening. The main text uses this result as a non-biological proof-of-concept; full toy-domain controls belong in Supplementary Note 1.

The first biological test extended the same principle to GFP ortholog landscapes. In ppluGFP, admissibility filtering reduced edges while preserving full connectivity, and recursive compression preserved near-perfect path structure. Across GFP systems, recursive basins preserved substantially lower phenotype dispersion than randomized basins even though phenotype labels were not used to construct the basins.

In the statistical-lock analysis, RIX basins were compared against 50 matched randomized recursive basin controls for each of ppluGFP and amacGFP. ppluGFP RIX basins had mean within-basin phenotype standard deviation 1144.1 versus randomized means of 2568–2571 (~ 55% reduction; z-like effect = 17.24), while amacGFP RIX basins had 883.2 versus 1787–1791 (~ 50% reduction; z-like effect = 16.42). Because 50 randomized controls were used, the minimum empirical one-sided p-value is approximately 1/51 = 0.0196 when no random control matches or exceeds the RIX result.

Key compression result: recursive admissibility achieved ~38% node compression and ~60% edge reduction in ppluGFP while preserving connectivity, navigability, and phenotype-local structure.

Admissibility-constrained recursive basins preserve substantially lower within-basin phenotype dispersion than matched randomized recursive basins. In the statistical-lock analysis, ppluGFP RIX basins had mean within-basin phenotype standard deviation 1144.1 versus randomized means of 2568–2571, while amacGFP RIX basins had 883.2 versus 1787–1791. Phenotype labels were not used to construct recursive basins, indicating that phenotype-local organization emerged from admissibility-constrained structure rather than phenotype supervision.

System / analysis	Main quantitative signal	Main interpretation
ppluGFP compression	~ 38% node compression; ~60% edge reduction; mean path distortion ~ 0.997	Redundancy removed without loss of navigability
ppluGFP phenotype locality	RIX std 1144.1 vs random 2568–2571; ~55% reduction; z-like effect 17.24; n = 50 controls	Phenotype locality emerged without phenotype-supervised basin formation
amacGFP phenotype locality	RIX std 883.2 vs random 1787–1791; ~50% reduction; z-like effect 16.42; n = 50 controls	Same admissibility rule preserved phenotype locality across ortholog systems

Global recursive transport geometry governs transfer compatibility

GFP ortholog analyses separated local organization from transferable alignment. Local basin structure could remain strong even when transfer failed; compatibility was associated instead with preserved global traversal geometry and phenotype-weighted geodesic organization.

Transfer retention is reported on the source-frozen recursive traversal metric; values are interpreted relative to matched shuffled, flattened, and incompatible-regime controls rather than as absolute prediction accuracy.

We define Recursive Transfer Capacity Threshold (RTCT) as the minimum recursive representational capacity required to preserve a specified transfer signal. Systems requiring low RTCT preserve transport organization efficiently, whereas systems requiring high RTCT require increasing representational complexity to maintain transfer performance. RTCT therefore acts as an operational measure of recursive transport coherence.

Compatible GFP systems maintain meaningful transfer performance at low recursive capacity, whereas incompatible systems require substantially greater capacity and exhibit transfer collapse. Intermediate systems occupy a transitional regime.

Recursive transport behavior separates into coherent, rerouting, transitional, and fragmented phase regions. Fragmentation boundaries emerge as a function of recursive capacity and perturbation pressure.

Together, these capacity and fragmentation analyses indicate that recursive transfer failure is not simply a loss of local organization, but a transition into a higher-capacity, perturbation-sensitive transport

regime.

Notably, compatible GFP systems occupy the coherent-to-transitional region at low recursive capacity, whereas incompatible systems move rapidly into fragmented regions as perturbation increases.

Comparative transport regimes: GFP vs GB1

Unlike the GFP ortholog systems, the GB1 fitness landscape did not exhibit fragmentation under recursive perturbation. Instead, transport remained robust across recursive compression levels, indicating that recursive organization can be preserved through alternative admissible pathways when sufficient topological redundancy is present.

Recursive compression reduced graph complexity substantially while preserving navigable structure. Node count decreased from 5,000 variants in the admissible graph to 889 compressed representatives, while edge count decreased from 45,512 to 16,714, corresponding to an 82.2% reduction in nodes and a 63.3% reduction in edges. Despite this compression, phenotype locality remained preserved. Mean within-basin phenotype standard deviation was 0.104 for recursive basins compared with 0.146 for random controls, corresponding to a 28.6% improvement in phenotype locality.

Recursive compression of the GB1 admissible graph reduced node count from 5,000 variants to 889 compressed representatives and edge count from 45,512 to 16,714, corresponding to 82.2% node reduction and 63.3% edge reduction. Despite this compression, recursive basins preserved phenotype locality relative to randomized controls.

Transport robustness was similarly preserved under perturbation. Mean phenotype drift remained nearly constant across recursive perturbation levels, varying only between 0.148 and 0.153. At the same time, curvature variance decreased from 716 in the unperturbed system to 33 under maximal perturbation, suggesting that recursive transport reorganized into smoother alternative pathways rather than collapsing into disconnected regions.

The resulting transport regime differs qualitatively from the fragmentation behavior observed in incompatible GFP transfers. Whereas incompatible GFP systems exhibited recursive transfer collapse under recursive capacity inflation and perturbation, GB1 maintained coherent transport despite substantial recursive perturbation. Comparative regime analysis therefore identifies two distinct classes of recursive topology: fragmentation-sensitive systems, in which admissible transport depends on a small number of critical pathways, and redundancy-supported systems, in which transport can be rerouted through multiple alternative admissible trajectories.

Additional evidence for redundancy-supported transport emerged from clustering interventions. In sparse structured topologies, maximum edge betweenness decreased from 0.0263 to 0.000052 following cluster enhancement, representing a greater than 500-fold reduction in bottleneck concentration. Spectral cohesion increased systematically across transport phases, with mean spectral

gap rising from 0.003 in fragmentation-sensitive systems to 0.136 in transitional systems and 0.274 in robust systems. Together, these results suggest that recursive transport robustness is governed not simply by local admissibility, but by the availability of redundant admissible routes capable of sustaining coherent traversal under perturbation.

The contrast between GB1 and GFP highlights a central conclusion of the framework: recursive transport behavior depends not only on admissible local structure but also on global topological organization. Dense redundant systems support transport rerouting and robustness under perturbation, whereas sparse bottleneck-dominated systems are more susceptible to fragmentation and transfer collapse. Recursive transport topology therefore provides a mechanistic framework for understanding how biological systems balance compression, robustness, and evolvability across distinct fitness landscapes.

System	Transport regime	Perturbation behavior	Mechanistic interpretation
GFP orthologs	Sparse structured	Fragmentation-sensitive	Bottleneck-dominated, low rerouting capacity
GB1	Dense redundant	Perturbation-robust	Distributed redundancy, high rerouting capacity

Unlike GFP-compatible transport systems, GB1 remains connected under perturbation and exhibits stable phenotype drift despite progressive recursive constraint disruption.

GFP-compatible and GB1 systems occupy distinct transport regimes. GFP exhibits fragmentation-sensitive behavior whereas GB1 exhibits perturbation robustness through redundant transport pathways.

Predictive and causal transport topology

Milestone 10 moved from description to prediction and intervention. Redundancy, rerouting capacity, bottleneck concentration, and spectral cohesion separated robust GB1-like transport from fragmentation-sensitive GFP-like transport before perturbation outcomes were considered.

Importantly, phase assignment could be predicted from redundancy and spectral cohesion prior to perturbation outcomes, indicating that transport behavior is forecastable from topology alone.

Synthetic clustering interventions dramatically reduce bottleneck concentration, demonstrating causal control over transport topology through rerouting enhancement.

Spectral cohesion cleanly separates fragmentation-sensitive, transitional, and robust transport regimes, indicating that global connectivity structure predicts transport stability.

Transport systems occupy distinct phase regions defined by redundancy and bottleneck pressure. Increasing redundancy systematically shifts systems toward robust transport behavior.

Phase mapping further separated systems into robust, transitional, and fragmentation-sensitive regions. Redundancy 2 behaved as fragmentation-sensitive, redundancy 4 as transitional, and redundancy 6–12 as increasingly robust. Spectral cohesion separated the same phase structure, suggesting that global graph cohesion tracks transport stability.

Discussion

The analysis supports the hypothesis that admissibility-constrained recursive transport structure can act as an organizing layer in protein sequence-function landscapes. Rather than depending primarily on local sequence similarity or post-hoc statistical optimization, admissible transformations can be recursively structured into coherent transport geometry prior to manifold formation. The resulting geometry is measurable, compressible, predictive within the tested graph-transport analyses, manipulable within constructed topology space, and organized into distinct topological phase regions.

A central contribution of this work is the separation of three frequently conflated layers: local similarity, local admissibility, and global recursive transport organization. Local admissible neighborhoods can generate internally coherent, phenotype-local basins even in transfer-incompatible systems. However, successful cross-system transfer requires coherent global traversal geometry across recursive coordinate systems. This distinction explains why systems that appear well organized locally can still fail to transfer: their recursive transport coordinates may be globally misaligned.

A second major finding is that robustness is strongly topological. Both GFP orthologs and GB1 exhibit recursive admissibility organization, yet they occupy sharply different transport regimes. Sparse, bottleneck-dominated systems are fragmentation-sensitive and perturbation-prone. Dense, redundant systems support effective rerouting and maintain global stability under stress. These differences are predictable from topology variables — redundancy, rerouting capacity, bottleneck concentration, and spectral cohesion — and can be altered through targeted synthetic interventions in constructed transport graphs. Most notably, increasing local rerouting capacity via clustering reduced maximum edge betweenness by nearly three orders of magnitude, demonstrating topology-level control over transport fragility.

Current self-supervised approaches such as LeJEPA seek stable and identifiable representations through alignment objectives, predictive structure, and distributional constraints. RIX addresses a complementary question: whether representational stability can also emerge from recursively conserved admissibility structure before manifold formation. In this sense, RIX is not proposed as a replacement for statistical representation learning, but as a candidate upstream organizational layer that may reduce the burden placed on downstream optimization.

Unlike protein language models, which learn representations from large-scale sequence statistics, or graph neural networks, which learn over a fixed graph structure, RIX focuses on constructing recursive transport geometry prior to representation learning. The framework is therefore complementary rather

than competitive: it proposes a pre-manifold structural prior that could potentially augment existing representation-learning architectures.

For biology, this framework provides a testable mechanistic lens on phenotype locality, ortholog compatibility, and evolutionary robustness. It suggests that biological systems may differ not only in local admissible structure, but also in the global topology of allowable transport. Dense, redundant landscapes may preserve function through rerouting, whereas sparse bottleneck-dominated landscapes may be more vulnerable to fragmentation and transfer collapse. This distinction may help explain why some protein systems remain robust under perturbation while others require higher representational capacity to preserve coherent traversal.

For artificial intelligence, the implications are architectural. Most current representation-learning pipelines begin with raw symbolic indices and delegate relational structure discovery to large-scale optimization. The RIX results motivate a different approach: construct or learn constraint-consistent recursive transport structure upstream of the primary learner. Such pre-manifold organization may improve transfer, sample efficiency, robustness, and interpretability in domains where not all transitions are valid and local similarity alone is insufficient.

Phase I does not claim that recursive transport topology explains all biological organization, nor that RIX supersedes existing methods. It provides evidence that pre-manifold constraint-consistent indexing is a computationally meaningful and architecturally consequential primitive. The framework now offers a foundation for broader empirical validation, controlled benchmark comparison, and practical implementation in biological modeling and artificial representation systems.

Scope, Limitations, and Falsifiability

Phase I does not claim universal biological generalization, closed-form transport laws, comprehensive biological prediction, AGI, or universal physics. It establishes a predictive and intervention-capable transport-topology framework that requires external replication and benchmark comparison.

- Early GFP analyses emphasize observed single-mutant graphs for interpretability and should be extended to broader combinatorial regimes.
- Initial admissibility rules use coarse amino-acid classes; future work should test BLOSUM-informed, structural, energetic, and evolutionary constraints.
- Some randomized-control analyses used limited random-control counts; external submission should use larger fixed-seed control suites.
- Curvature analyses use simplified Forman-style graph curvature; alternative transport curvature metrics should be compared.
- Synthetic perturbations and topology interventions probe transport structure but should not yet be interpreted as direct evolutionary dynamics.

- The framework requires controlled comparisons against GNNs, hierarchical graph methods, protein language models, and topology-preserving nulls.

The hypothesis is weakened if constraint-consistent indexing fails against matched topology controls, if recursive compression fails to preserve phenotype locality beyond randomized basins, if global transport geometry does not predict transfer compatibility across expanded systems, or if topology interventions fail to reproduce predicted changes under external replication.

Reproducibility and Benchmark Roadmap

The accompanying repository supports reproduction of the locked Phase I claims. Future work will extend this reproducibility package into a broader benchmark suite for evaluating recursive transport-topology methods against standard graph and protein-representation baselines.

Planned benchmark extensions include controlled comparisons against GCN, GAT, GIN, GraphSAGE, DiffPool, MinCutPool, HARP, ESM-2/ESM-3, and topology-preserving random controls. Evaluation will focus on ortholog transfer, sparse-label prediction, perturbation robustness, recursive compression efficiency, and topology-aware generalization.

This benchmark roadmap is intended to separate three questions: whether admissibility-constrained indexing preserves phenotype-local structure, whether recursive transport geometry improves transfer robustness, and whether topology-aware pre-manifold organization provides useful inductive bias for downstream biological and AI representation models.

Conclusion

This Phase I study supports the conclusion that admissibility-constrained recursive transport structure can induce measurable, compressible, and predictive organization in biological sequence-function landscapes, with topology-level manipulability demonstrated within constructed graph-topology analyses. Across GFP orthologs and GB1, compatible or redundant systems preserved compact, stable transport, while bottleneck-sensitive systems exhibited fragmentation, recursive capacity inflation, and perturbation-sensitive collapse.

For biology, this provides a testable graph-transport lens on phenotype locality, transfer compatibility, and perturbation robustness. For AI, it suggests a concrete architectural hypothesis: more broadly, the results suggest that representation geometry may emerge not only from optimization and scale, but also from the lawful organization of admissible transformations before learning begins.

Declarations

Availability of Data and Materials

All primary datasets used in this study are publicly available from the original sources cited in the manuscript. Processed datasets, canonical metric definitions, deterministic figure-generation scripts, and supporting reproducibility workflows for Proof RIX Phase I are available in the project repository: https://github.com/ScottA8275/AIIA-Proof-RIX-Phase1

All empirical figures in the main text are generated from the public reproducibility artifacts in the accompanying repository. Figure 1 is a conceptual schematic summarizing the Phase I workflow and is not used as quantitative evidence.

This repository is designed to reproduce the locked scientific claims of the analysis rather than the complete internal research history. The package emphasizes deterministic figure regeneration, canonical metrics, interpretable transport-topology analysis, and recursive phase-boundary reproducibility.

The repository intentionally exposes frozen processed datasets, canonical figure-generation scripts, public metric definitions, and deterministic transport-topology workflows. The repository focuses on the finalized analyses reported here and does not attempt to archive exploratory development history or unrelated internal infrastructure. This distinction is intended to preserve both scientific reproducibility and conceptual clarity.

Author Contributions

Jeffery Scott Allbright conceived the study, designed the analysis, performed experiments, and wrote the manuscript.

Competing Interests

Jeffery Scott Allbright reports pending intellectual property related to the theoretical and computational frameworks described in this work and is a co-founder of Forma Substrate, Inc. and AIIA Technologies, Inc., which may commercialize these ideas. The author declares no other competing interests.

Supplementary Information

Supplementary Information includes expanded graph-construction details, randomized-control summaries, perturbation sweeps, transfer-metric definitions, phase-map generation details, random seeds, and figure-source mappings.

References

1. Sarkisyan KS, Bolotin DA, Meer MV et al (2016) Local fitness landscape of the green fluorescent protein. Nature 533:397–401. https://doi.org/10.1038/nature17995
2. Gonzalez Somermeyer L, Fleiss A, Mishin AS et al (2022) Heterogeneity of the GFP fitness landscape and data-driven protein design. eLife, 11, e75842

3. Wu NC, Dai L, Olson CA, Lloyd-Smith JO, Sun R (2016) Adaptation in protein fitness landscapes is facilitated by indirect paths. eLife, 5, e16965
4. Forman R (2003) Bochner's method for cell complexes and combinatorial Ricci curvature. Discrete Comput Geom 29:323–374
5. Hamming RW (1950) Error detecting and error correcting codes. Bell Syst Tech J 29(2):147–160
6. Kipf TN, Welling M (2017) Semi-supervised classification with graph convolutional networks. International Conference on Learning Representations
7. Velickovic P, Cucurull G, Casanova A et al (2018) Graph attention networks. International Conference on Learning Representations
8. Rives A, Meier J, Sercu T et al (2021) Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. Proceedings of the National Academy of Sciences, 118(15), e2016239118
9. Belkin M, Niyogi P (2003) Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput 15(6):1373–1396
10. You J, Ying R, Leskovec J (2020) Design space for graph neural networks. NeurIPS
11. Ying R, You J, Morris C, Ren X, Hamilton WL, Leskovec J (2018) Hierarchical graph representation learning with differentiable pooling. NeurIPS
12. Ollivier Y (2009) Ricci curvature of Markov chains on metric spaces. J Funct Anal 256(3):810–864
13. Balestriero R, LeCun Y (2025) *LeJEPA: Provable and Scalable Self-Supervised Learning Without the Heuristics*. arXiv:2511.08544
14. White M (2007) The G-Ball, a New Icon for Codon Symmetry and the Genetic Code. arXiv q-bio/0702056. https://arxiv.org/abs/q-bio/0702056

Figures

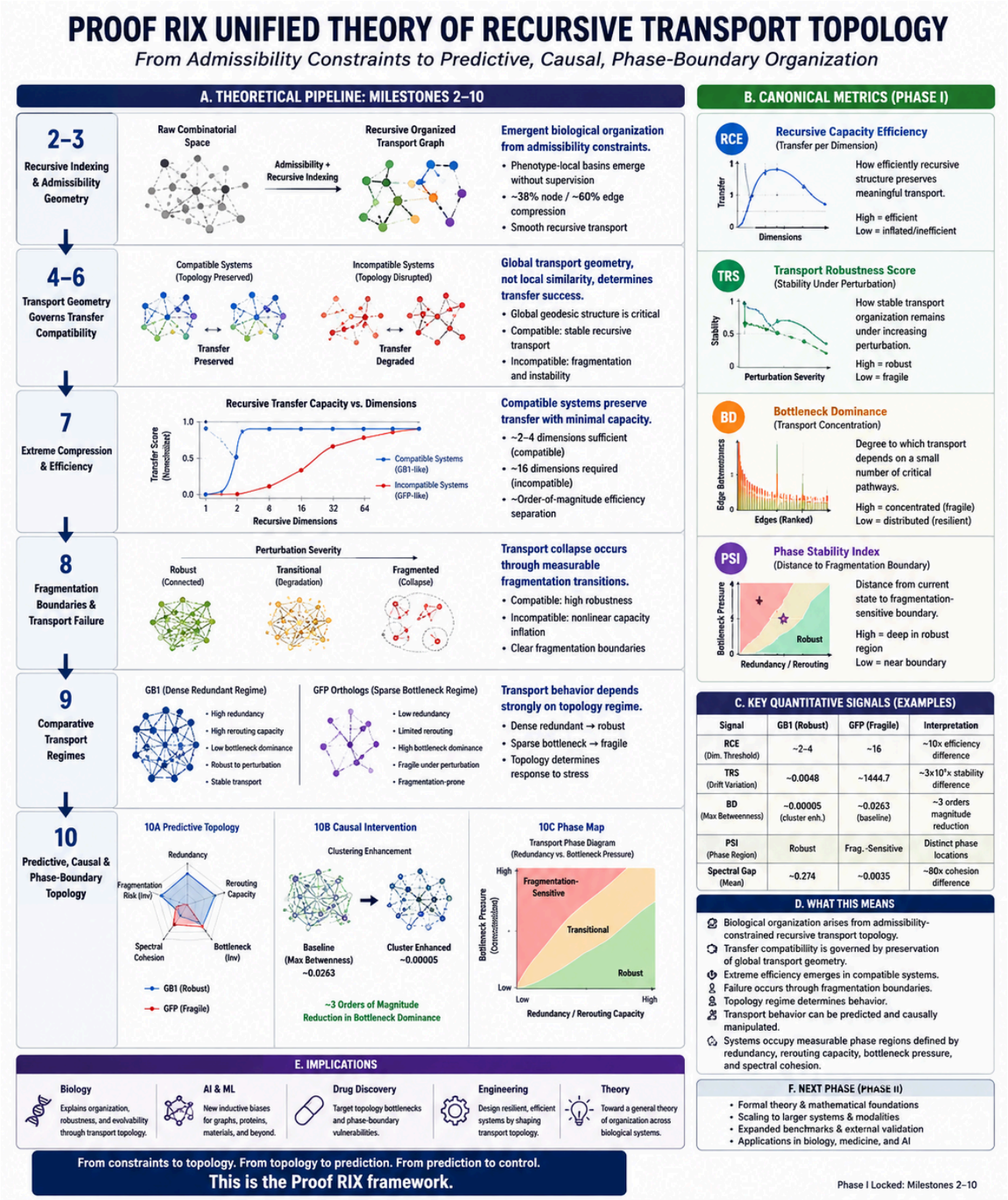

Figure 1

Unified theory progression

This Phase I study moves from lawful indexing and recursive biological compression to transfer geometry, fragmentation boundaries, comparative regimes, causal topology intervention, and measurable phase regions.

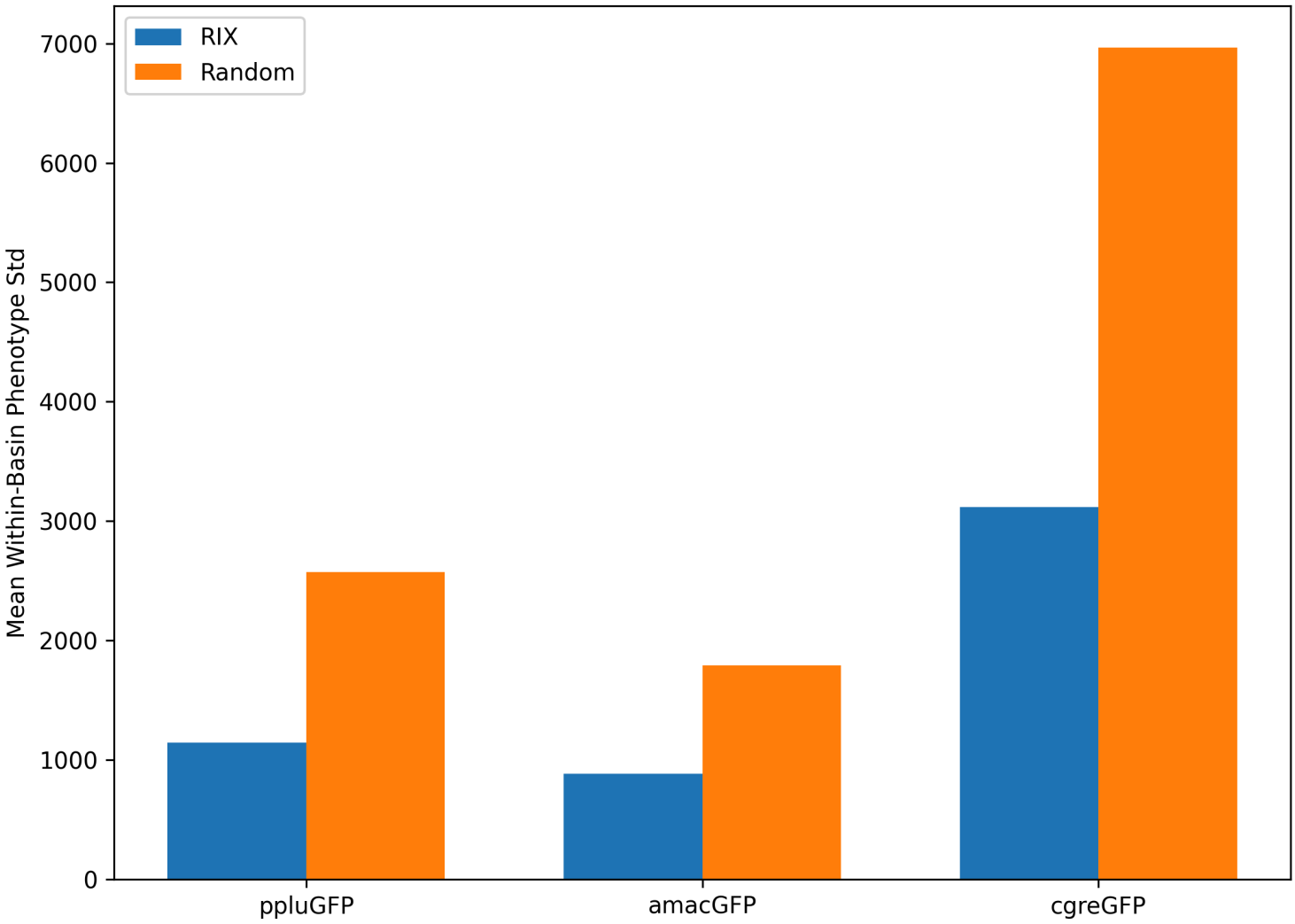

Figure 2

Recursive GFP compression preserves navigability and phenotype-local basin structure

Admissibility-constrained recursive basins preserve substantially lower within-basin phenotype dispersion than matched randomized recursive basins. In the statistical-lock analysis, ppluGFP RIX basins had mean within-basin phenotype standard deviation 1144.1 versus randomized means of 2568–2571, while amacGFP RIX basins had 883.2 versus 1787–1791. Phenotype labels were not used to construct recursive basins, indicating that phenotype-local organization emerged from admissibility-constrained structure rather than phenotype supervision.

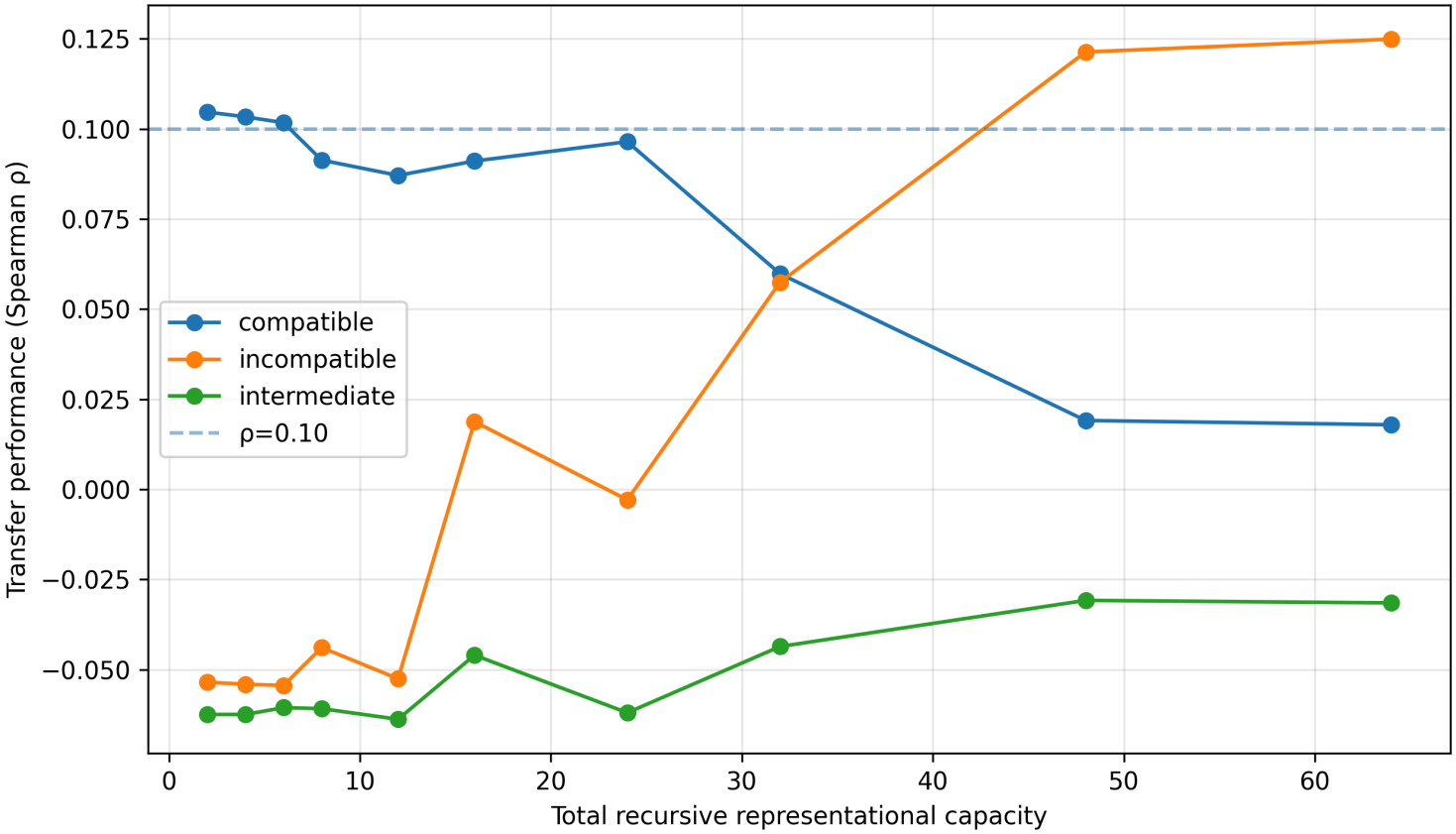

Figure 3

Recursive Transfer Capacity Threshold Curves

Compatible GFP systems maintain meaningful transfer performance at low recursive capacity, whereas incompatible systems require substantially greater capacity and exhibit transfer collapse. Intermediate systems occupy a transitional regime.

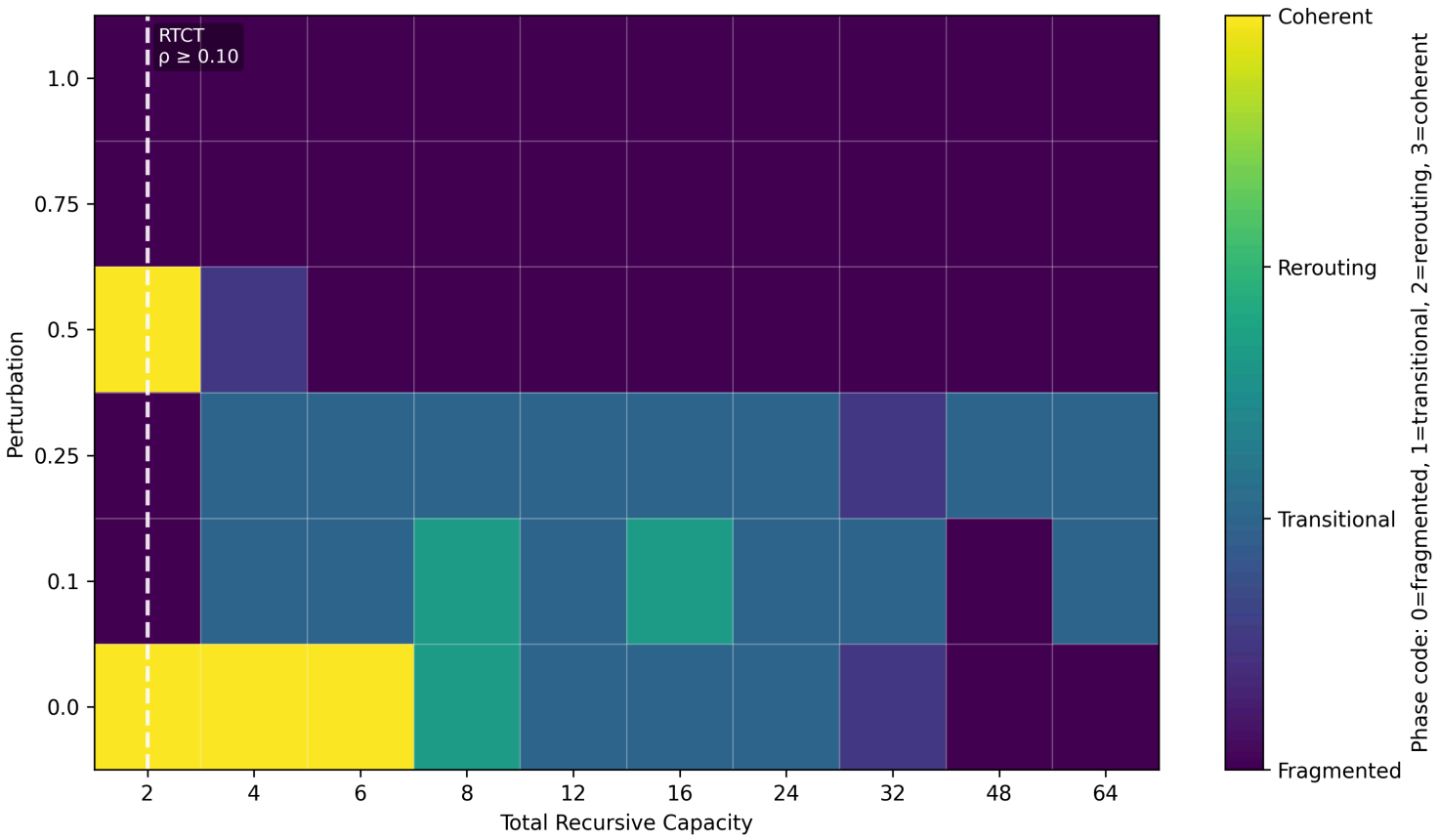

Figure 4

Recursive Transport Fragmentation Boundaries and Phase Regions

Recursive transport behavior separates into coherent, rerouting, transitional, and fragmented phase regions. Fragmentation boundaries emerge as a function of recursive capacity and perturbation pressure.

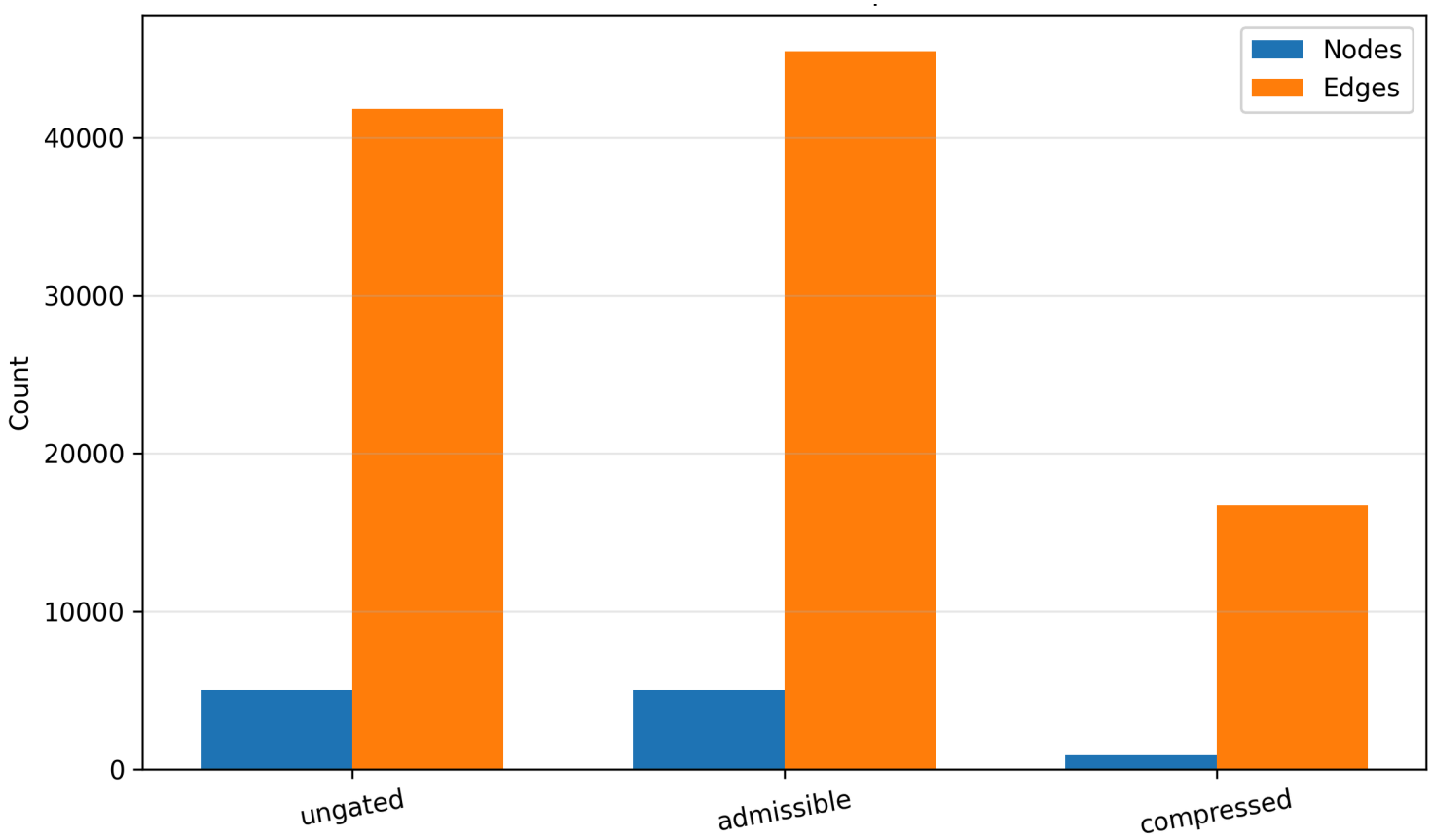

Figure 5

GB1 recursive compression preserves navigable topology

Recursive compression of the GB1 admissible graph reduced node count from 5,000 variants to 889 compressed representatives and edge count from 45,512 to 16,714, corresponding to 82.2% node reduction and 63.3% edge reduction. Despite this compression, recursive basins preserved phenotype locality relative to randomized controls.

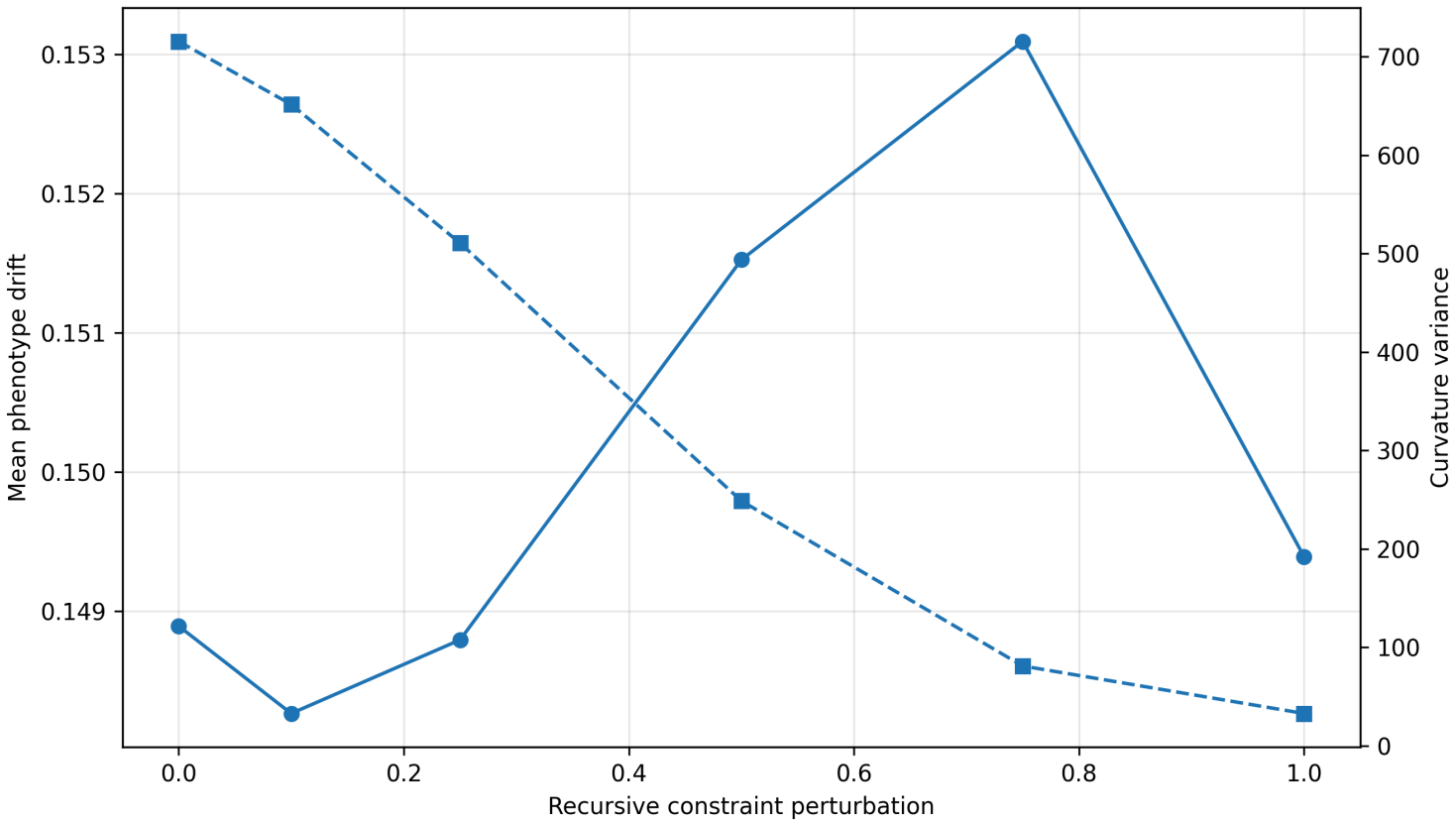

Figure 6

GB1 Robustness Profile

Unlike GFP-compatible transport systems, GB1 remains connected under perturbation and exhibits stable phenotype drift despite progressive recursive constraint disruption.

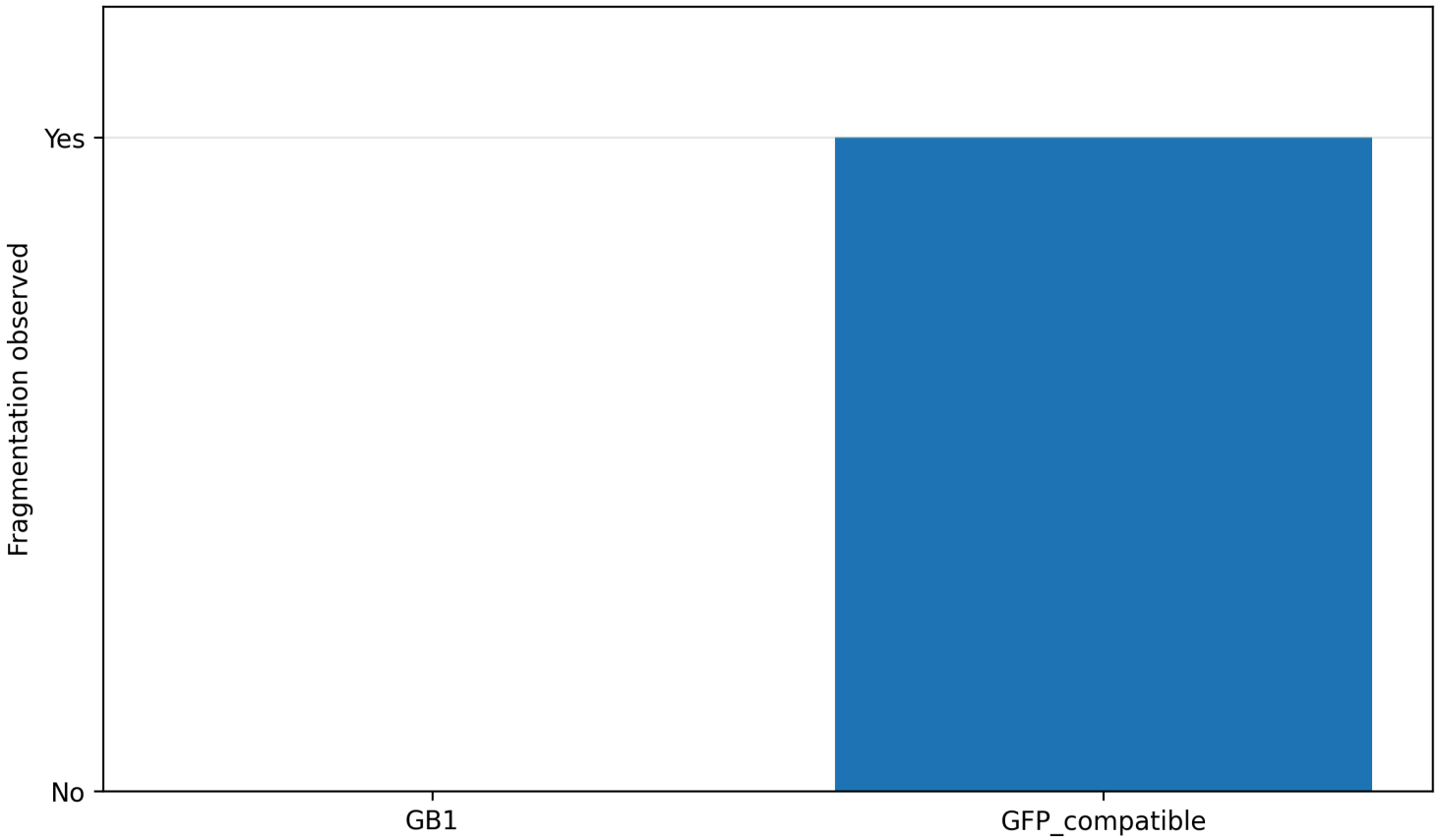

Figure 7

Comparative Transport Regimes

GFP-compatible and GB1 systems occupy distinct transport regimes. GFP exhibits fragmentation-sensitive behavior whereas GB1 exhibits perturbation robustness through redundant transport pathways.

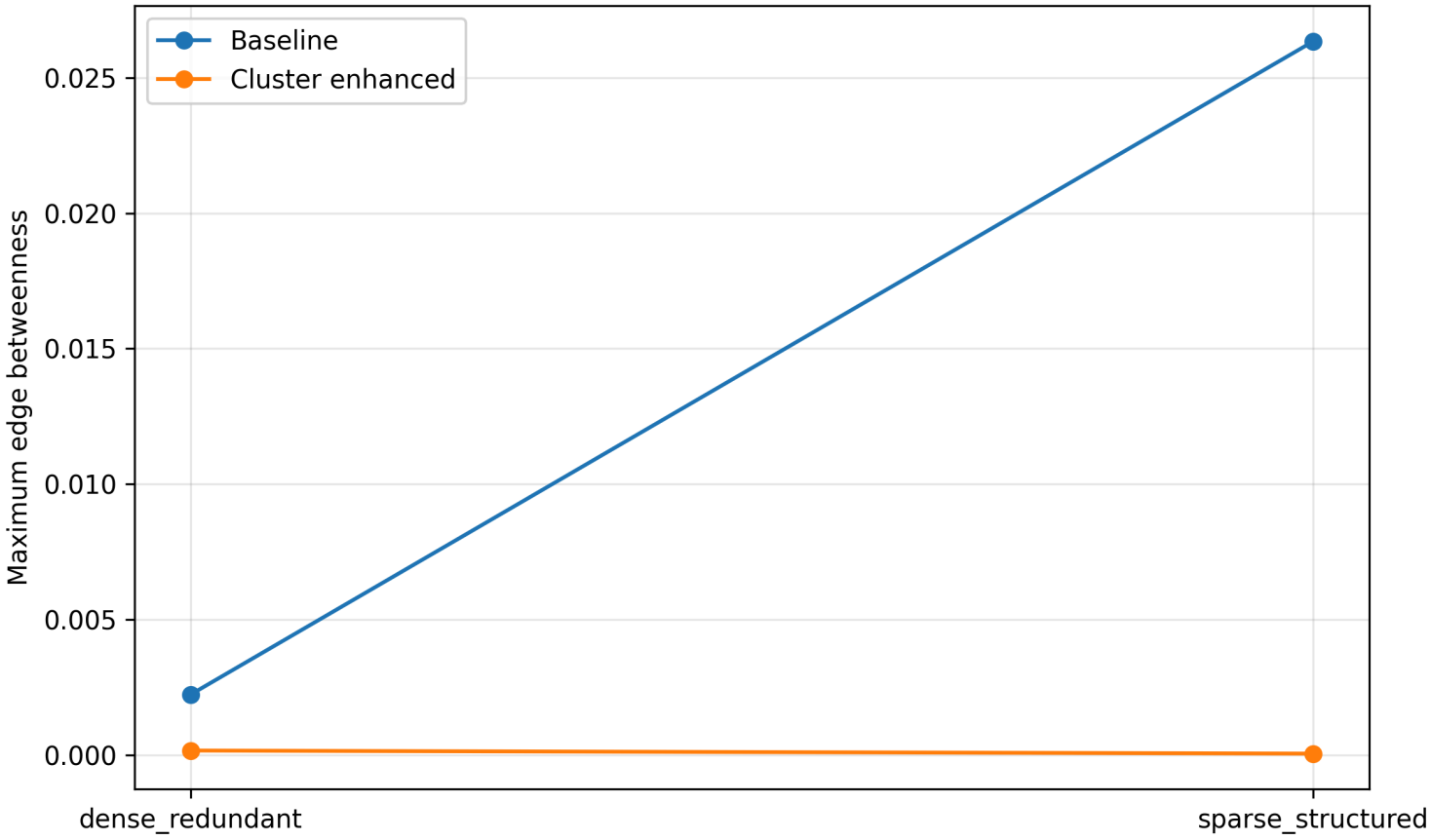

Figure 8

Clustering Enhancement Eliminates Bottleneck-Dominated Transport

Synthetic clustering interventions dramatically reduce bottleneck concentration, demonstrating causal control over transport topology through rerouting enhancement.

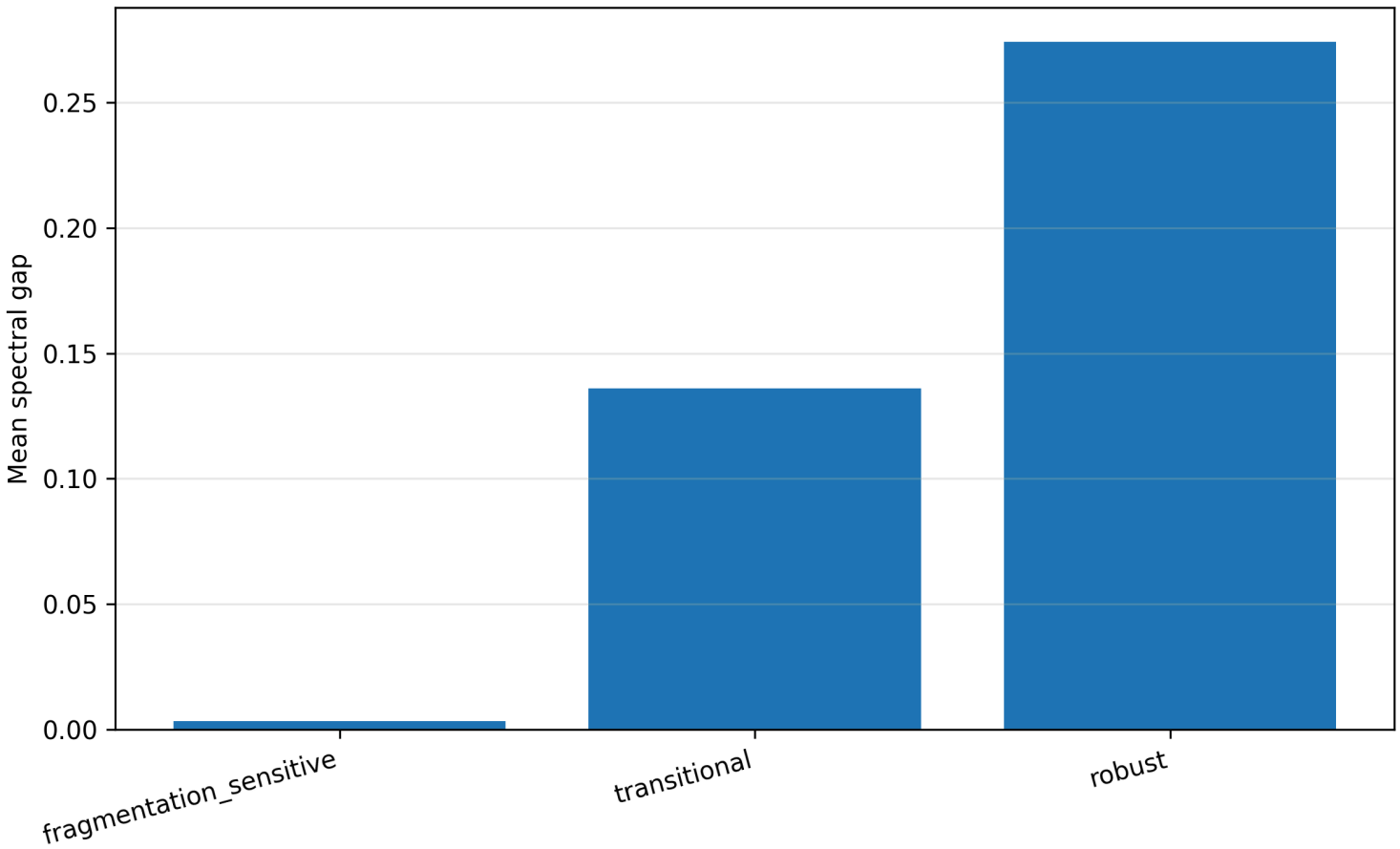

Figure 9

Spectral Cohesion Separates Transport Phases

Spectral cohesion cleanly separates fragmentation-sensitive, transitional, and robust transport regimes, indicating that global connectivity structure predicts transport stability.

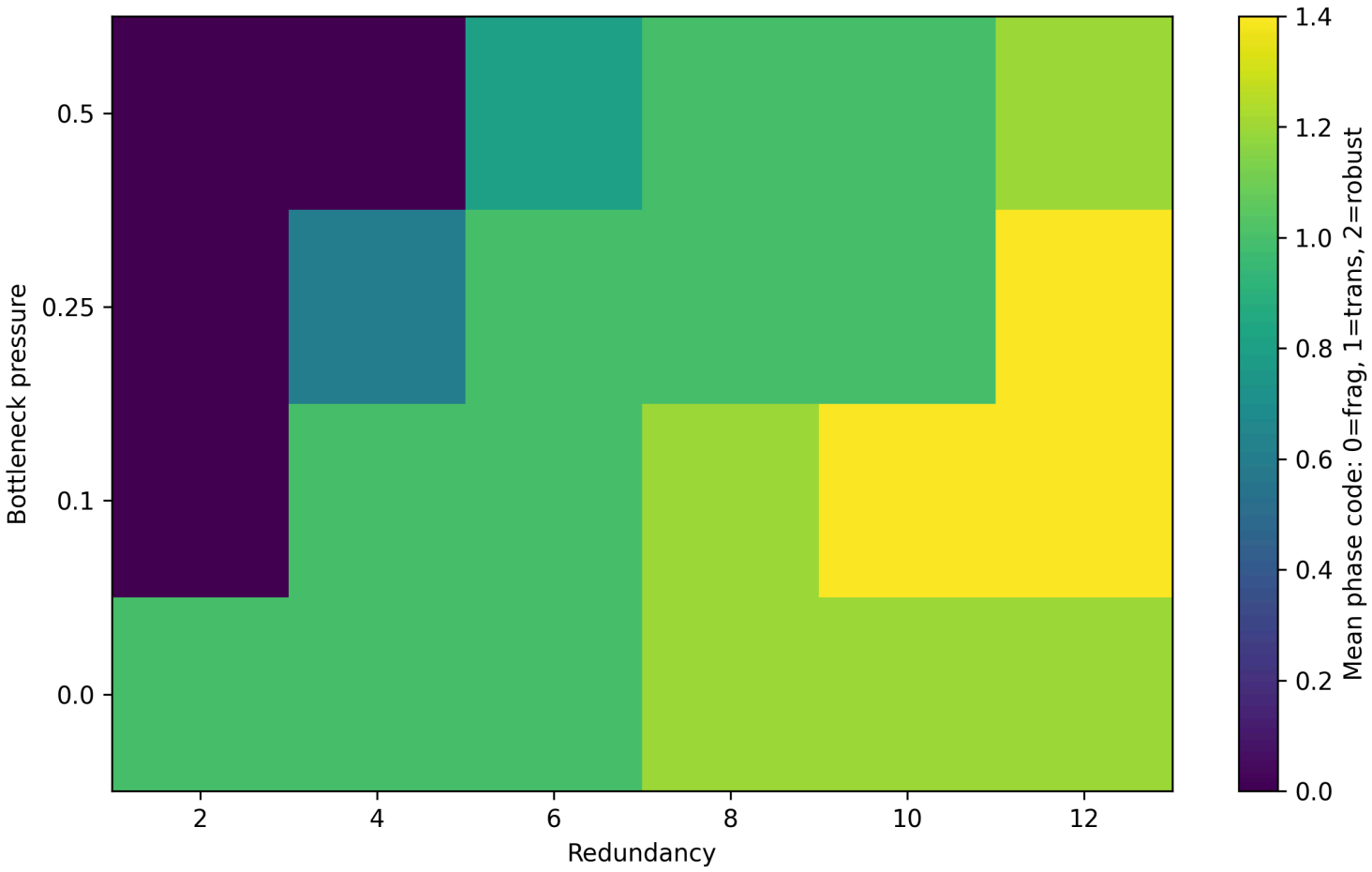

Figure 10

Recursive Transport Phase Map

Transport systems occupy distinct phase regions defined by redundancy and bottleneck pressure. Increasing redundancy systematically shifts systems toward robust transport behavior.