Graph Topology vs. Correlation Structure: Does Correlation Reveal Hidden Geometry?

A Falsifiable Experiment Report

Experiment 01 — Relational Dimension Project

January 29, 2026

Abstract

We test the hypothesis that correlation structure between graph nodes reveals geometric information that pure topological connectivity cannot capture. Specifically, we predict that graphs with long-range correlations (correlations between topologically distant nodes) will exhibit lower effective dimensionality in correlation space than in topological space. We formalize this as five falsifiable predictions with pre-specified thresholds. After testing 150 configurations across random geometric graphs and lattices with three correlation patterns, we find that **4 of 5 predictions fail**, including the core hypothesis. The measured compression ratio for long-range correlations ($\delta = 0.246$) falls just below the pre-specified threshold of 0.25. We discuss methodological limitations, particularly the instability of dimension estimation, and propose refinements for future work.

1 Introduction

Graph-based representations are ubiquitous in modeling complex systems, from social networks to neural connectivity. A fundamental question is whether the *correlation structure* between nodes—how their states or activities co-vary—contains geometric information beyond what the graph's *topological structure* (edges and paths) reveals.

1.1 Core Hypothesis

When nodes have correlation patterns that deviate from their topological connectivity, the correlation-based dimensional embedding will differ from the topology-based embedding. Specifically, strong correlations between topologically distant nodes create "shortcuts" in correlation space, potentially allowing lower-dimensional representations.

1.2 Why This Matters

If correlation structure reveals latent geometry invisible to topology, this has implications for:

- Dimensionality reduction in network analysis

- Detecting hidden structure in complex systems

- Understanding the relationship between functional and structural connectivity

2 Methods

2.1 Graph Types

We study two graph families:

Random Geometric Graphs (RGG)   Nodes are placed uniformly at random in $[0,1]^2$, with edges connecting pairs within radius r. The radius is chosen to achieve average degree ≈ 6.

Square Lattices   Regular 2D grids with nearest-neighbor connectivity. Lattices provide a known-geometry baseline where we expect $d \approx 2$.

2.2   Correlation Patterns

For each graph, we construct three types of correlation matrices:

Nearest-Neighbor (NN) $C_{ij} = \rho$ if nodes i, j are adjacent, else $C_{ij} = 0$ (plus identity diagonal). This serves as a baseline where correlation matches topology.

Long-Range (LR) $C_{ij} = \rho \cdot \exp(-d_{\text{euclidean}}(i,j)/\lambda)$, where $\lambda = 0.3 \times$ diameter. This creates correlations based on geometric proximity rather than topological distance.

Random (RAND) k random node pairs receive correlation ρ, where k matches the number of significant correlations in the LR pattern. This controls for correlation density without geometric structure.

All correlation matrices are projected to the nearest positive semidefinite matrix via eigenvalue clipping.

2.3   Dimension Extraction

Topological Dimension (d_{topo})

1. Compute all-pairs shortest path distances D_{topo}

2. Apply Isomap embedding with $n_{\text{neighbors}} = 8$

3. Find minimum k where reconstruction error $< 0.1\times$ initial error

Correlation Dimension (d_{corr})

1. Convert correlation to distance: $D_{ij} = \sqrt{2(1 - C_{ij})}$

2. Apply same Isomap procedure

Both methods are validated by requiring Isomap and MDS estimates to agree within 0.5 dimensions.

2.4   Compression Ratio

The key metric is the compression ratio:

$$\delta = \frac{d_{\text{topo}} - d_{\text{corr}}}{d_{\text{topo}}}$$

- $\delta > 0$: Correlation dimension is lower (compression)

- $\delta \approx 0$: Dimensions agree

- $\delta < 0$: Correlation dimension is higher (expansion)

2.5 Test Matrix

- Graph sizes: $N \in \{50, 100, 200\}$ for RGG; sides $\in \{7, 10, 14\}$ for lattices

- Correlation strength: $\rho = 0.8$

- Replications: 10 per configuration

- Total: 150 configurations

3 Pre-Registered Predictions

We state five predictions with explicit pass/fail thresholds, determined *before* running the experiment.

Table 1: Pre-registered predictions with falsification criteria

ID	Prediction	Pass Criterion	Fail Criterion
P1	Baseline Agreement	$|\delta_{\mathrm{NN}}| < 0.2$	$|\delta_{\mathrm{NN}}| \geq 0.2$
P2	LR Compression	$\delta_{\mathrm{LR}} > 0.25$	$\delta_{\mathrm{LR}} \leq 0.15$
P3	Scaling	$\delta_{200}/\delta_{50} > 1.3$	ratio ≤ 1.0
P4	Control	$\delta_{\mathrm{rand}} < 0.15$ AND $\delta_{\mathrm{LR}} > 0.25$	$\delta_{\mathrm{rand}} \geq \delta_{\mathrm{LR}}$
P5	Dose-Response	$r(\alpha, \delta) > 0.8$	$r < 0.5$

P1 (Baseline Agreement) When correlation follows topology exactly (NN pattern), topological and correlation dimensions should agree. This validates our measurement approach.

P2 (Core Hypothesis) Long-range correlations should produce significant compression ($\delta > 0.25$). This is the central claim—that correlation shortcuts enable lower-dimensional embedding.

P3 (Scaling) The compression effect should increase with graph size, ruling out finite-size artifacts.

P4 (Control) Random correlations should not produce systematic compression, demonstrating that geometric structure matters.

P5 (Dose-Response) Compression should increase monotonically with correlation strength α, showing a smooth causal relationship.

4 Results

4.1 Summary

4.2 Detailed Results

P1 Failure: High Baseline Variance The mean absolute compression ratio for NN correlations was $|\delta| = 0.368$, far exceeding the 0.2 threshold. Individual measurements ranged from -0.75 to $+0.67$. This high variance indicates that even when correlation matches topology, our dimension estimates are unstable.

Table 2: Prediction outcomes

ID	Description	Threshold	Measured	Result
P1	Baseline Agreement	< 0.20	0.368	**FAIL**
P2	LR Compression	> 0.25	0.246	**FAIL**
P3	Scaling Behavior	> 1.30	3.112	**PASS**
P4	Random Control	rand < 0.15, LR > 0.25	0.11, 0.25	**FAIL**
P5	Dose-Response	$r > 0.80$	0.175	**FAIL**
	Predictions passed:			**1/5**

P2 Near-Miss: Compression Below Threshold Long-range correlations produced mean $\delta = 0.246$, just below the 0.25 threshold. While this suggests some compression effect may exist, it fails to meet our pre-specified criterion.

P3 Success: Scaling Confirmed The ratio $\delta_{200}/\delta_{50} = 3.11$ strongly exceeds the 1.3 threshold. Compression increases with graph size:

- $N = 50$: $\delta = 0.108 \pm 0.46$

- $N = 200$: $\delta = 0.336 \pm 0.49$

P4 Failure: LR Threshold Not Met While random correlations showed low compression ($\delta_{\mathrm{rand}} = 0.11 < 0.15$), the LR compression ($\delta_{\mathrm{LR}} = 0.246$) failed to exceed 0.25, causing overall failure.

P5 Failure: No Dose-Response The correlation between strength α and compression δ was $r = 0.175$ ($p = 0.74$), indicating no significant relationship. The dose-response curve was essentially flat.

4.3 Baseline Calibration

The known-geometry baseline (10×10 lattice with NN correlations) revealed a calibration issue:

- $d_{\mathrm{topo}} = 2.0$ (expected)

- $d_{\mathrm{corr}} = 3.0$ (expected 2.0)

This $\delta = -0.5$ indicates the correlation-to-distance transformation does not preserve dimensionality as expected.

4.4 Method Disagreement

A key finding is the disagreement between Isomap and MDS:

- For RGG $N = 200$: Isomap gives $d = 2.0$, MDS gives $d = 8.1$

- This 6-dimension gap suggests fundamental instability in dimension estimation

5   Discussion

5.1   Interpretation

The core hypothesis—that long-range correlations create geometric shortcuts enabling dimensional compression—is **not supported** by these data. While P2 came close ($\delta = 0.246$ vs. threshold 0.25), we must respect the pre-registered threshold.

However, the P1 failure suggests a deeper problem: our measurement approach may lack the precision to detect the hypothesized effect. The high variance in baseline measurements ($|\delta| = 0.37$ when it should be ≈ 0) indicates substantial noise.

5.2   Why Did P3 Pass?

The scaling prediction passed despite other failures. This could indicate:

1. A real but weak effect that becomes detectable only at larger scales

2. Systematic bias in dimension estimation that scales with N

3. Improved Isomap/MDS stability at larger sample sizes

5.3   Methodological Limitations

1. **Dimension estimation instability**: Isomap and MDS often disagreed by 2–6 dimensions

2. **Threshold-based detection**: Using a fixed error threshold (0.1) may be too sensitive to noise

3. **Correlation-to-distance transformation**: The $D = \sqrt{2(1 - C)}$ mapping may not preserve geometric structure

4. **Small graph sizes**: $N \leq 200$ may be insufficient for stable manifold learning

5.4   Alternative Interpretations

The data are consistent with alternative hypotheses:

- Correlation structure provides *different* (not lower-dimensional) geometric information

- The effect exists but requires stronger correlations or larger graphs

- Isomap/MDS are inappropriate for this type of distance matrix

6   Conclusion

We tested whether correlation structure reveals hidden low-dimensional geometry in graphs. Our pre-registered predictions were largely falsified:

- **Core hypothesis (P2)**: Not supported ($\delta = 0.246 < 0.25$)

- **Baseline calibration (P1)**: Failed, indicating measurement issues

- **Scaling (P3)**: Confirmed, suggesting possible weak effect

- **Controls (P4, P5)**: Failed

The experiment demonstrates the value of pre-registered predictions: without explicit thresholds, one might interpret $\delta = 0.246$ as "substantial compression." The falsifiable framework forces intellectual honesty.

6.1 Recommendations for Future Work

1. Use continuous dimension estimators (e.g., explained variance ratios) rather than threshold-based detection

2. Require stronger Isomap/MDS agreement before accepting estimates

3. Test alternative manifold learning methods (UMAP, t-SNE with fixed perplexity)

4. Increase graph sizes ($N \geq 500$) for more stable estimation

5. Consider fractional dimension estimators

Data Availability

All code, data, and analysis artifacts are available in the repository:

- Source code: `experiments/01-topology-vs-correlation/src/`

- Raw results: `experiments/01-topology-vs-correlation/output/metrics.json`

- Figures: `experiments/01-topology-vs-correlation/output/*.png`

Acknowledgments

A   Detailed Results by Configuration

Table 3: RGG results by size and correlation type

Type	N	d_{topo}	d_{corr}	δ	Std
NN	50	3.50	2.75	-0.05	0.56
NN	100	3.55	2.50	0.21	0.27
NN	200	5.05	3.55	0.17	0.49
LR	50	3.50	2.75	0.11	0.42
LR	100	3.45	2.50	0.25	0.26
LR	200	5.05	3.20	0.34	0.37
RAND	50	3.50	2.75	0.05	0.52
RAND	100	3.45	2.55	0.17	0.30
RAND	200	5.05	3.65	0.11	0.44

B   Lattice Results

Lattice graphs showed deterministic results (zero variance) due to fixed structure:

Note: The 10×10 lattice shows $d_{\text{topo}} = 2.0$ (correct) but $d_{\text{corr}} = 3.0$ (incorrect), indicating a calibration issue in the correlation-based dimension estimation.

Table 4: Lattice results

Type	Side	d_{topo}	d_{corr}	δ
NN	7	3.0	2.5	0.17
NN	10	2.0	3.0	-0.50
NN	14	3.0	2.0	0.33
LR	7	3.0	2.5	0.17
LR	10	2.0	3.0	-0.50
LR	14	3.0	2.5	0.17