Compression Scaling Laws:
How Does Dimensional Compression Scale with Graph Size?

A Falsifiable Experiment Report

Experiment 02 — Relational Dimension Project

January 30, 2026

Abstract

Building on Experiment 01's finding that compression ratio scales with graph size (ratio 3.11 for $N = 200/50$), we investigate the functional form of compression scaling. We test logarithmic ($\delta = a \log N + b$) and power law ($\delta = cN^{\alpha}$) models across graph sizes $N = 50$ to $N = 1000$, with 60 total configurations. We formalize six falsifiable predictions with pre-specified thresholds. Results show that logarithmic scaling fits better ($R^2 = 0.81$) than power law ($R^2 = 0.63$) for $N \leq 500$, but a methodological transition to sparse methods at larger N introduces a discontinuity that invalidates extrapolation tests. **All 6 predictions failed**, primarily due to this methodological artifact. However, within the consistent-method range ($N = 50$–500), scaling behavior is evident and approximately logarithmic. We discuss implications and propose methodological refinements.

1 Introduction

Experiment 01 established that long-range (LR) correlations produce measurable compression of correlation-based distances relative to topological distances. The compression ratio $\delta = (d_{\text{topo}} - d_{\text{corr}})/d_{\text{topo}}$ increased from 0.108 at $N = 50$ to 0.336 at $N = 200$, yielding a scaling ratio of 3.11—the only prediction that passed in Experiment 01.

1.1 Research Question

This experiment asks: **What is the functional form of compression scaling, and does it predict behavior at unseen graph sizes?**

Understanding the scaling law has both theoretical and practical implications:

- **Theoretical:** Logarithmic scaling would suggest information-theoretic limits; power-law scaling would suggest geometric origins
- **Practical:** A validated scaling law enables prediction of compression at scales too costly to measure directly

1.2 Key Findings from Experiment 01

The effect appears real but requires larger N to manifest clearly.

2 Methods

2.1 Improved Dimension Estimation

We extend Experiment 01's threshold-based dimension detection with continuous (fractional) dimension estimation:

Table 1: Experiment 01 results motivating this study

N	δ (RGG-LR)	δ (RGG-NN)	Ratio to $N = 50$
50	0.108	0.070	1.00
100	0.296	0.233	2.74
200	0.336	0.410	3.11

1. Compute reconstruction error curve $e(k)$ for $k = 1, \ldots, k_{\max}$
2. Find dimension where error drops below threshold $\tau \cdot e(1)$ (where $\tau = 0.1$, same as Exp01)
3. Use **linear interpolation** between integer dimensions for fractional estimate

This maintains consistency with Experiment 01's error-based approach while providing smoother dimension estimates that should reduce variance.

2.2 Sparse Methods for Large N

For graphs with $N > 500$, full distance matrices become computationally expensive. We employ:

Landmark MDS Select 200 random landmarks, embed with classical MDS, project remaining points via inverse-distance-weighted interpolation.

Sparse Isomap Use k-nearest neighbor graph ($k = 15$) instead of full distance matrix for geodesic estimation.

Critical Note: As results will show, this methodological transition introduces artifacts that compromise cross-scale comparisons.

2.3 Test Matrix

Table 2: Test matrix for scaling analysis

N	Replications	Method	Purpose
50	10	Full	Anchor to Exp01
100	10	Full	Anchor to Exp01
200	10	Full	Anchor to Exp01
300	10	Full	New data point
500	10	Full	Training set boundary
750	5	Sparse	Test set
1000	5	Sparse	Extrapolation test

Total: 60 configurations. Correlation type: LR only (the pattern showing scaling signal in Exp01).

2.4 Model Fitting

We fit two candidate scaling models:

Logarithmic Model

$$\delta(N) = a \log(N) + b$$

Rationale: Many information-theoretic scaling laws are logarithmic.

Power Law Model

$$\delta(N) = cN^{\alpha}$$

Rationale: Geometric phenomena often exhibit power-law scaling.

Models are fit on $N \leq 500$ (training set) using least-squares regression. Predictive validity is tested on $N > 500$.

3 Pre-Registered Predictions

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

Table 3: Pre-registered predictions with falsification criteria

ID	Prediction	Pass Criterion	Rationale
P1	Monotonic Scaling	Spearman $r > 0.9$	Compression should increase with N
P2	Logarithmic Form	$R^2 > 0.85$	Log model fits well
P3	Power Law Form	$R^2 > 0.85$, $\alpha > 0$	Power model fits well
P4	Non-Saturation	$\delta_{1000} > \delta_{500} + 0.05$	Effect continues growing
P5	Predictive Validity	Relative error $< 20\%$	Model extrapolates
P6	Variance Reduction	$\mathrm{Std}(\delta)_{N=100} < 0.15$	Improved method reduces noise

P1 (Monotonic Scaling) If compression is real, larger graphs should show more compression. Spearman correlation captures monotonicity regardless of functional form.

P2 & P3 (Functional Form) At least one model should fit well. If both fail, the scaling may be more complex (e.g., saturating, piecewise).

P4 (Non-Saturation) The effect should not plateau prematurely. A saturating effect would suggest finite-size artifacts rather than true scaling.

P5 (Predictive Validity) The key test: can we predict $\delta(1000)$ from data at $N \leq 500$? A valid scaling law should extrapolate.

P6 (Improved Baseline) The continuous dimension estimation should reduce variance compared to Exp01's threshold-based approach (which had Std $= 0.27$ at $N = 100$).

4 Results

4.1 Raw Scaling Data

Critical observation: There is a sign change at $N > 500$ where sparse methods begin. The negative δ values indicate $d_{\text{corr}} > d_{\text{topo}}$, the opposite of compression.

4.2 Model Fitting (Training Set: N $\leq$ 500)

Logarithmic Model

$$\delta = 0.107 \log(N) - 0.258$$

$R^2 = 0.812$

Table 4: Compression ratio by graph size

N	δ (mean)	δ (std)	Method
50	0.108	0.317	Full
100	0.298	0.185	Full
200	0.337	0.509	Full
300	0.326	0.386	Full
500	0.386	0.547	Full
750	-0.151	0.302	Sparse
1000	-0.097	0.431	Sparse

Power Law Model

$$\delta = 0.022 N^{0.487}$$

$R^2 = 0.630$

The logarithmic model provides better fit, approaching but not exceeding the 0.85 threshold.

4.3 Prediction Evaluation

Table 5: Prediction outcomes

ID	Description	Threshold	Measured	Result
P1	Monotonic Scaling	$r > 0.90$	-0.29	**FAIL**
P2	Logarithmic Form	$R^2 > 0.85$	0.81	**FAIL**
P3	Power Law Form	$R^2 > 0.85$	0.63	**FAIL**
P4	Non-Saturation	$\Delta\delta > 0.05$	-0.48	**FAIL**
P5	Predictive Validity	error $< 20\%$	594%	**FAIL**
P6	Variance Reduction	Std < 0.15	0.19	**FAIL**
Predictions passed:				**0/6**

4.4 Analysis of Failures

P1 Failure: Sign Reversal The Spearman correlation is *negative* ($r = -0.29$) because δ increases for $N = 50$–500 but then *decreases* (becomes negative) for $N = 750$–1000. This is clearly a methodological artifact.

P2 Near-Miss The log model achieved $R^2 = 0.81$, just below the 0.85 threshold. Within the $N \leq 500$ range, logarithmic scaling is a reasonable description.

P3 Failure: Power Law Insufficient Power law fit ($R^2 = 0.63$) is substantially worse than logarithmic, suggesting the relationship is not geometric power-law.

P4 & P5 Catastrophic Failures Both predictions failed catastrophically due to the methodological transition. The model predicted $\delta(1000) \approx 0.48$ but measured $\delta(1000) = -0.10$, a 594% error with wrong sign.

P6 Failure: Variance Not Reduced The standard deviation at $N = 100$ was 0.19, slightly exceeding the 0.15 threshold. The continuous dimension estimation did not substantially reduce variance.

5 Discussion

5.1 The Methodological Discontinuity Problem

The dominant finding is that switching from full to sparse methods at $N = 500$ introduces a discontinuity that invalidates cross-method comparisons. Within the full-method range ($N =$ 50–500):

- Compression *does* increase with N, consistent with Exp01
- Logarithmic scaling provides reasonable fit ($R^2 = 0.81$)
- Variance remains high, suggesting need for more replications

The sparse methods (Landmark MDS, Sparse Isomap) appear to systematically bias dimension estimates in the opposite direction, producing negative δ.

5.2 Interpretation Within Consistent Range

Restricting to $N = 50$–500 (full methods only):

- Spearman $r = 0.90$ (exactly at threshold)
- Log model $R^2 = 0.81$ (close to threshold)
- Clear upward trend in compression

The scaling hypothesis is *plausible but not validated* within this range.

5.3 Why Did Sparse Methods Fail?

Several factors may explain the discontinuity:

1. **Landmark sampling bias:** Random landmark selection may not capture the manifold structure
2. **k-NN graph artifacts:** The sparse neighbor graph may distort geodesic distances differently for correlation vs. topological metrics
3. **Scale mismatch:** Sparse methods may be calibrated for different regimes than full methods

5.4 Recommendations

1. **Use consistent methods:** Full Isomap/MDS is computationally feasible up to $N = 1000$ (only 8MB per distance matrix)
2. **Increase replications:** 20+ replications to reduce standard error
3. **Add intermediate sizes:** $N \in \{400, 600, 800\}$ to verify continuity
4. **Calibrate sparse methods:** Before using sparse methods, validate on sizes where both methods can be applied

6 Conclusion

We tested whether compression scaling follows logarithmic or power-law forms and whether the scaling law predicts behavior at larger graph sizes. Our pre-registered predictions were falsified:

- **All 6 predictions failed**
- **Primary cause:** Methodological discontinuity when switching to sparse methods at $N > 500$
- **Within consistent range ($N \leq 500$):** Logarithmic scaling is plausible ($R^2 = 0.81$)

The experiment demonstrates both the value and challenge of falsifiable predictions. The pre-registered design forced us to confront the methodological failure directly rather than selectively reporting favorable subsets. The scaling hypothesis remains plausible but requires methodologically consistent replication at larger scales.

6.1 Key Takeaways

1. Scaling signal *is* present in $N = 50$–500 range
2. Logarithmic scaling fits better than power law ($R^2 = 0.81$ vs. 0.63)
3. Sparse methods introduce systematic bias incompatible with full methods
4. Methodological consistency is essential for cross-scale studies

Data Availability

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

- Source code: `experiments/02-scaling-laws/src/`
- Raw results: `experiments/02-scaling-laws/output/metrics.json`
- Figures: `experiments/02-scaling-laws/reports/*.png`

Acknowledgments

This experiment was conducted as part of the Relational Dimension research project, using a falsifiable science methodology where predictions and thresholds are specified before data collection.

A Detailed Scaling Data

B Model Parameters

B.1 Logarithmic Model (Training Set)

$$\begin{aligned} \delta &= a \log(N) + b \\ a &= 0.1066 \\ b &= -0.2577 \\ R^2 &= 0.8123 \end{aligned}$$

Table 6: Full results by graph size

N	Replications	δ (mean)	δ (std)	δ (min)	δ (max)
50	10	0.108	0.317	−0.42	0.58
100	10	0.298	0.185	0.02	0.54
200	10	0.337	0.509	−0.31	0.92
300	10	0.326	0.386	−0.18	0.85
500	10	0.386	0.547	−0.25	0.95
750	5	−0.151	0.302	−0.51	0.15
1000	5	−0.097	0.431	−0.62	0.38

B.2 Power Law Model (Training Set)

$$\begin{aligned} \delta &= cN^{\alpha} \\ c &= 0.0218 \\ \alpha &= 0.4869 \\ R^2 &= 0.6297 \end{aligned}$$

C Prediction at N=1000

Using the logarithmic model (better fit):

$$\delta_{\text{predicted}}(1000) = 0.1066 \cdot \log(1000) - 0.2577 = 0.479$$

Measured: $\delta_{\text{measured}}(1000) = -0.097$
Relative error: $\frac{|0.479-(-0.097)|}{|-0.097|} = 594\%$
This catastrophic failure is attributable to the methodological transition, not the scaling model itself.