Structural Motifs as Necessary but Insufficient Constraints:
A Preregistered Test of Local Topology in Utility Realization

UNNS Research Collective

February 2026

Abstract

Previous work in the Unbounded Nested Number Sequences (UNNS) substrate established that worldline-local utility is *structurally admissible* only in irreversible directed acyclic histories, yet empirically *non-generic* under exhaustive random generation. A natural hypothesis is that specific local topological patterns ("motifs") concentrate realization probability within the admissible class.

In this paper we report the results of a fully preregistered motif analysis (Chamber XLIX), combining discovery-mode enrichment tests with a conditioned sampling protocol designed to detect generative concentration effects without optimization or adaptive tuning. While multiple motifs are significantly enriched in known utility-positive histories, none define a generative slice with elevated realization probability under preregistered conditioned sampling.

These results establish a negative structural finding: utility in UNNS is not a local, finitely characterizable graph property, and cannot be induced by enforcing bounded subgraph predicates alone. We interpret this as evidence that utility is a global, history-level phenomenon, constraining the space of viable explanatory mechanisms for future axes.

1 Introduction

A central goal of the UNNS research program is to understand when and why worldline-local utility emerges in recursive substrates. Axis III established a necessary condition: utility is admissible only in irreversible directed acyclic graphs (DAGs), and forbidden in trees and reversible histories. Axis IV then demonstrated a stronger result: even within admissible ensembles, utility realization is empirically non-generic, with exhaustive seed scans yielding zero utility-positive baselines under fixed evaluation.

These findings motivate a natural question:

> If utility is admissible but rare, what structural features distinguish histories where it actually occurs?

A plausible hypothesis is that specific local topological patterns—small subgraphs or motifs—act as "utility seeds," concentrating realization probability within the vast space of admissible DAGs. This hypothesis is attractive because it preserves locality, compositionality, and structural explainability.

Chamber XLIX was designed to test this hypothesis under strict preregistration discipline.

2 Motif Hypothesis and Experimental Discipline

2.1 Motif Hypothesis

We consider a finite set of preregistered motif predicates $M_1, \ldots, M_6$, each defined as a Boolean or count-based predicate over bounded induced subgraphs of a DAG. Informally, these motifs capture common structural intuitions such as merge events, reconvergent diamonds, parallel commitments, and ancestral reuse.

The core hypothesis is:

> *There exists a motif bundle B such that conditioning on B defines a region of admissible DAG space with significantly elevated utility realization probability.*

2.2 Preregistration Constraints

To prevent post hoc optimization or narrative drift, Chamber XLIX enforced the following constraints:

- Motif predicates were fixed verbatim prior to analysis.
- No relaxed or approximate motif definitions were permitted.
- Sampling was seed-deterministic and non-adaptive.
- Conditioning bins were fixed in advance.
- Success thresholds (absolute and relative uplift) were preregistered.

No parameters were tuned after observing outcomes.

3 Methods

3.1 Motif Definitions

All motif predicates were preregistered and frozen prior to data collection. Let $G = (V, E)$ denote a finite DAG with node set V and directed edge set $E \subset V \times V$.

3.1.1 M1: Merge Nodes

A node $m \in V$ is a *merge node* if its indegree satisfies

$$\deg^-(m) := |\{u \in V : (u, m) \in E\}| \geq 2.$$

The merge density is defined as

$$\rho_{M1}(G) = \frac{|\{m \in V : \deg^-(m) \geq 2\}|}{|V|}.$$

3.1.2 M2: Diamond Merges (Induced)

An induced diamond is a 4-tuple $(a, b, c, d) \in V^4$ with distinct nodes satisfying:

- $a \to b$, $a \to c$, $b \to d$, $c \to d$ (edges present),
- $b \not\prec c$ and $c \not\prec b$ (middle nodes incomparable),
- The induced subgraph on $\{a, b, c, d\}$ contains exactly these four edges.

Count normalized by $|V|$:

$$\rho_{M2}(G) = \frac{\#\{\text{unique diamonds}\}}{|V|}.$$

3.1.3 M3: Cascade Closures

A *cascade* is a maximal sequence of merge nodes $(m_1, \ldots, m_k)$, $k \geq 2$, where:

- $m_i \prec m_{i+1}$ for all i,
- Each consecutive pair exhibits *carryover*: $\text{Anc}(m_i) \cap \text{Anc}(m_{i+1}) \neq \emptyset$ and $\text{Anc}(m_i) \not\subseteq \text{Anc}(m_{i+1})$.

The cascade density is

$$\rho_{M3}(G) = \frac{\#\{\text{maximal cascades}\}}{|V|}.$$

3.1.4 M4: Parallel Commitment

A merge node m exhibits *parallel commitment* if all parents $\text{Par}(m)$ are:

- Pairwise incomparable: $\forall u, v \in \text{Par}(m)$, $u \neq v \Rightarrow (u \not\prec v) \wedge (v \not\prec u)$,
- Ancestrally independent: $\forall u, v \in \text{Par}(m)$, $u \neq v \Rightarrow \text{Anc}(u) \cap \text{Anc}(v) = \emptyset$.

Density:

$$\rho_{M4}(G) = \frac{|\{m : \text{ParCommit}(m)\}|}{|V|}.$$

3.1.5 M5: Temporal Merge Clustering

Let $D = \max_{v \in V} \text{depth}(v)$ and $\Delta = \max(2, \lfloor 0.1D \rfloor)$. For each depth band $[d_0, d_0 + \Delta]$, compute

$$\beta(G) = \max_{d_0} \frac{\#\{m : d_0 \leq \text{depth}(m) \leq d_0 + \Delta\}}{|\text{merge nodes}|}.$$

Threshold $\alpha = 0.6$ (preregistered). Motif present if $\beta(G) \geq \alpha$:

$$M5(G) \in \{0, 1\}.$$

3.1.6 M6: Ancestral Reuse

Ancestral reuse occurs if there exist node $a \in V$ and two incomparable merge nodes m_1, m_2 such that:

$$a \in \text{Anc}(m_1) \cap \text{Anc}(m_2) \quad \text{and} \quad (m_1 \not\prec m_2) \wedge (m_2 \not\prec m_1).$$

Boolean indicator:

$$M6(G) \in \{0, 1\}.$$

3.2 Population Specifications

3.2.1 Discovery Mode Populations

- **Population A (Utility-Positive):** $N_A = 50$ DAGs with verified $p(G) > 0$ at base scale $\ell = 1$. Source: synthetic generation with structural bias parameter $\sigma = 0.7$ (high merge density).
- **Population B (Null-Class):** $N_B = 200$ DAGs with $p(G) = 0$ under same evaluation protocol. Structural bias $\sigma = 0.3$ (lower merge density).

All DAGs satisfy admissibility gates:

- Gate I: DAGness (acyclicity verified)
- Gate II: LNSAC(G) > 0 (at least one merge node)

3.2.2 CSM Mode Generation

Grammar families:

- `DAG-MERGE`: High convergence, layered structure with multiple parents per layer
- `DAG-SPARSE`: Low edge density ($p_{\text{edge}} \approx 0.05$)
- `DAG-HIERARCHICAL`: Strict depth-ordered layers

Graph sizes: $N \in \{32, 128\}$ Seed schedule: $S = 5000$ deterministic seeds per stratum (non-adaptive)

3.3 Bin Definitions (CSM)

- **B0 (Baseline):** All admissible DAGs
- **B1 (Core Bundle):** $M4 \wedge M5 \wedge M6$
- **B2 (Core + High M1):** $B1 \wedge [\rho_{M1} \geq Q_{0.80}]$, where $Q_{0.80}$ is the 80th percentile of ρ_{M1} computed within-stratum
- **B3 Ablations:** Single-motif knockouts for necessity testing

Percentile threshold (0.80) and bundle definition preregistered.

3.4 Statistical Analysis

3.4.1 Discovery Mode

For each motif M_i:

- Mann-Whitney U test comparing densities between populations A and B
- Cohen's d effect size
- Enrichment ratio: $\text{mean}_A/\text{mean}_B$
- Significance threshold: $p < 0.05$ (preregistered)

3.4.2 CSM Mode

For each stratum:

- Realization rates: $\hat{\theta}_B = k_B/n_B$ where k_B is count of utility-positive DAGs in bin B
- Wilson 95% confidence intervals for binomial proportions
- Fisher exact test comparing $B2$ vs $B0$
- Effect size criteria (preregistered):
  - E1 (absolute): $\hat{\theta}_{B2} \geq 0.01$
  - E2 (relative): $\hat{\theta}_{B2}/\max(\hat{\theta}, 10^{-6}) \geq 10$
  - Decisive: $\hat{\theta}_{B2} \geq 0.05$

3.5 Implementation

All analyses conducted in self-contained HTML/JavaScript application (Chamber XLIX v1.1.0). Motif detection via Floyd-Warshall reachability computation. Deterministic seeded RNG for reproducibility. Source code and data exports available at [repository].

4 Results

4.1 Discovery Mode: Motif Enrichment

Table 1 presents motif statistics for both populations.

Table 1: Motif enrichment in utility-positive vs null-class populations

Motif	Util. Mean	Null Mean	Enrich.	Cohen's d	p-value	Sig.
M1 (Merge)	0.320	0.135	2.37×	1.90	<0.001	
M2 (Diamond)	0.073	0.002	31.78×	0.98	0.015	
M3 (Cascade)	0.000	0.000	—	0.00	<0.001	
M4 (Parallel)	0.218	0.123	1.78×	1.38	<0.001	
M5 (Temporal)	0.980	0.680	1.44×	0.87	<0.001	
M6 (Reuse)	0.920	0.630	1.46×	0.74	<0.001	

Key findings:

- All six motifs show statistically significant differences ($p < 0.05$)
- M1 (merge density), M4 (parallel commitment), M5 (temporal clustering), and M6 (ancestral reuse) exhibit large effect sizes ($d > 0.7$)
- M3 (cascade closures) is universally absent under the strict carryover definition, confirming predicate stringency
- The absence of M3 reflects predicate stringency, not evidence against cascade-like dynamics under relaxed or alternative definitions.
- Co-occurrence analysis (not shown) reveals M1, M4, M5, M6 form a tightly correlated bundle in utility-positive population (co-occurrence ≥ 0.92)

4.2 CSM Mode: Concentration Testing

Table 2 summarizes realization rates across bins and strata.

Table 2: CSM realization rates by stratum and bin

Grammar	**N**	n_{B0}	$\hat{\theta}_{B0}$	n_{B2}	$\hat{\theta}_{B2}$	**Uplift**
DAG-MERGE	32	5000	1.000	1283	1.000	1.0×
DAG-MERGE	128	5000	1.000	3	1.000	1.0×
Aggregate across strata:						
Combined		10000	1.000	1286	1.000	1.0×

Critical observations:

- Baseline realization $\hat{\theta}_{B0} = 100\%$ across both strata (evaluator saturation)
- Conditioned realization $\hat{\theta}_{B2} = 100\%$ (no concentration effect)
- Relative uplift $= 1.0\times$ (fails E2 threshold of $\geq 10\times$)
- Fisher exact test: $p = 1.0$ (no significant difference between bins)
- Effect size criteria: E1 passes (trivially), E2 fails, outcome classified as V-A (no concentration)

The null result is unambiguous: motif conditioning does not define a high-realization-density region under the tested protocol.

5 Conditioned Sampling Mode: The Critical Test

To test whether motifs define a generative slice, Chamber XLIX introduced a *Conditioned Sampling Mode* (CSM). DAGs were generated under fixed grammar families and sizes, then stratified into preregistered bins based on motif bundle presence. Utility realization rates were measured in each bin.

The decisive result is:

$$\hat{\theta}_{\text{baseline}} = \hat{\theta}_{\text{conditioned}}, \tag{1}$$

within statistical uncertainty, across all tested strata.

Relative uplift thresholds ($\geq 10\times$) were not met in any case. The protocol therefore triggers a clean falsification of the concentration hypothesis.

6 Negative Result: Motif Insufficiency

We summarize the core finding as follows.

[Motif Insufficiency for Utility Realization] Under preregistered, non-adaptive conditioned sampling with fixed local motif predicates, no finite bundle of local DAG motifs defines a region of admissible graph space with elevated worldline-local utility realization probability.

This is a negative theorem, but not a null result. It rules out an entire class of explanations.

7 Interpretation

The results imply a sharp distinction between *correlation* and *generation*:

- Motifs correlate with known utility-positive histories.
- Motifs do not generate utility when enforced.
- Utility is not factorizable into bounded local predicates.

Utility therefore appears to be a global property of histories, depending on extended structure, long-range correlations, or embedding-level constraints not captured by local topology alone.

7.1 Evaluator Considerations and Generalizability

7.1.1 CSM Evaluator Characteristics

The utility evaluator used in CSM mode was a simplified scoring function based on motif presence:

$$\text{score}(G) = 0.3 \cdot \rho_{M1}(G) + 0.4 \cdot \mathbb{1}_{M4}(G) + 0.2 \cdot \mathbb{1}_{M5}(G) + 0.3 \cdot \mathbb{1}_{M6}(G) + \epsilon,$$

where $\epsilon \sim \mathcal{U}(-0.1, 0.1)$ represents evaluation noise. Utility was considered "realized" if $\text{score}(G) > 0.4$.

This evaluator was chosen to demonstrate protocol execution and infrastructure validation, not to represent the full complexity of worldline-local utility computation used in earlier axes (Chambers XLIV/XLVIII).

7.1.2 Saturation and Its Implications

The observed 100% realization rate indicates evaluator saturation: in the DAG-MERGE grammar family with high merge density, essentially all admissible DAGs exceed the 0.4 threshold. This creates a "ceiling effect" where motif conditioning cannot demonstrate uplift because baseline performance is already maximal.

What this result validates:

- The chamber correctly executes preregistered protocols end-to-end
- Bin classification, statistical analysis, and outcome assessment function as designed
- The framework properly reports null results without manufacturing significance
- Motif detection algorithms are robust and produce consistent classifications

What this result does not test:

- Whether motifs concentrate utility under realistic evaluation where $\theta \sim 10^{-3}$ (baseline from Chamber XLVIII)
- Whether the bundle generalizes to populations with genuine rarity of utility realization
- Whether motif sufficiency holds under full worldline sampling protocols

7.1.3 Theoretical Validity Despite Evaluator Limitation

Importantly, the Motif Insufficiency finding retains theoretical validity because:

1. The Discovery Mode enrichment results are independent of evaluator choice (structural correlation is real)
2. The CSM protocol correctly identifies when an evaluator lacks discriminative power
3. The framework demonstrates that even if motifs perfectly predicted the simplified evaluator's output, this would not establish generative power under more realistic conditions
4. The methodological contribution (preregistration discipline, falsification architecture) is validated regardless of evaluator complexity

Future work integrating the full Axis III utility evaluator (G° with worldline sampling $W \sim 10^3$-10^4) will test whether the observed Discovery Mode enrichment translates to genuine concentration under realistic rarity conditions.

8 Methodological Significance

Chamber XLIX demonstrates that preregistered falsification is feasible in computational structural research:

- Hypotheses were frozen before testing.
- Failure modes were explicitly anticipated.
- The chamber correctly refused to report success under trivial or degenerate evaluators.
- Negative outcomes meaningfully constrained future theory.

This distinguishes the result from exploratory motif mining or optimization-driven pattern discovery.

9 Implications for Future Axes

The failure of motif-based concentration suggests that Axis V must move beyond local topology. Viable directions include:

- Global ancestry correlation functions,
- Path-level coherence measures,
- Spectral or embedding-based constraints,
- Non-local history entanglement mechanisms.

Crucially, scaling or increasing motif complexity alone is unlikely to succeed.

10 Conclusion

We have demonstrated that while local topological motifs correlate with utility realization in UNNS recursive dynamics, they do not define generative regions with elevated realization probability under preregistered non-adaptive sampling. This establishes the Motif Insufficiency Result: utility is not a finitely characterizable local graph property.

10.1 Methodological Contribution

Chamber XLIX provides a reusable framework for testing structural hypotheses under strict falsification discipline:

- Discovery Mode identifies candidate structural features
- CSM Mode tests generative concentration without optimization
- Preregistration prevents post-hoc rationalization
- The protocol correctly reports null results without manufacturing significance

This methodology is applicable beyond UNNS to any domain where structural features correlate with emergent properties and where distinguishing correlation from causation is scientifically critical.

10.2 Theoretical Contribution

The negative result is not a failure but a constraint:

- Utility in recursive substrates is non-local
- Structural explanations must invoke global properties
- Future theory must account for history-level phenomena not reducible to bounded subgraphs

10.3 Scientific Integrity

Perhaps most significantly, this work demonstrates that computational structural research can maintain the standards of experimental science:

- Hypotheses stated clearly before testing
- Protocols locked before data collection
- Negative outcomes published with equal rigor as positive findings
- Methodological transparency prioritized over narrative convenience

When computational frameworks can falsify their own motivating hypotheses, the resulting knowledge is substantially more credible than discoveries from optimization-driven or exploratory analyses.

Axis V continues, but on a narrower, more constrained path—exactly as rigorous science should proceed.