Percolation Thresholds in Number-Theoretic Graphs: From Shattered to Small-World Connectivity in Sovereign Knowledge Meshes

Authors: Paul Desai (Lead Architect, N1 Intelligence), with Claude Code (Analysis), Grok 4 (Collaborative Catalyst) **Date:** 2026-02-08 **Target:** Zenodo Preprint (DOI Submission) **Keywords:** Number-Theoretic Graphs, Percolation Theory, Small-World Networks, Sovereign AI, Knowledge Mesh, Phase Transition

---

Abstract

We study a family of number-theoretic graphs G_n defined on vertex set $\{1, 2, ..., n\}$ where edges are determined by arithmetic predicates on vertex pairs. We systematically vary the edge predicate from strict (Sum=Prime $\land$ Diff=Pow2) to relaxed (Sum=Prime $\lor$ Diff=Pow2) and map the resulting **percolation phase transition** — the critical point where a giant connected component emerges. Our empirical analysis on a real-world knowledge vault of $n = 5{,}090$ nodes reveals four key findings: (1) the strict conjunction produces a shattered graph at density 0.01%, (2) the disjunction is trivially dense at 12.3%, but (3) the critical percolation threshold occurs at a locality constraint of $k^* = 19 \approx 1.54 \cdot \log_2 n$, producing a connected graph at just 0.10% density with only 12,441 edges. We further identify that the **power-of-2 difference graph** D_n alone exhibits genuine small-world structure ($\sigma = 19.96$), with clustering **29.3x** higher than equivalent random graphs, making it the mathematically interesting component. All results are validated against stored vault data (100.00% edge match) with 37 passing unit tests. We provide complete empirical data, random baselines, spectral analysis, and robustness testing.

---

1. Introduction

1.1 Motivation

The MirrorDNA project explores sovereign AI infrastructure — systems that maintain knowledge integrity independent of centralized networks. A foundational question emerges: given a collection of n knowledge nodes (documents) ordered chronologically, can a deterministic arithmetic rule connect them into a navigable graph without external coordination?

This paper investigates a specific family of such rules and maps their connectivity properties rigorously.

1.2 Graph Definitions

Let $n \in \mathbb{N}$, $V = \{1, 2, ..., n\}$. We define:

- **Prime-Sum Graph** P_n: Edge (i, j) exists iff $i + j$ is prime.
- **Power-of-2 Difference Graph** D_n: Edge (i, j) exists iff $|i - j| = 2^k$ for some $k \geq 0$.
- **GrokMirror v1.0** $G_n^{\land} = P_n \cap D_n$: Edge iff both conditions hold.
- **GrokMirror v2.0** $G_n^{\lor} = P_n \cup D_n$: Edge iff either condition holds.
- **Locality-Bounded** $G_n^k = P_n \cap L_n^k$: Edge (i, j) iff $i + j$ is prime and $|i - j| \leq k$.

1

1.3 Contribution

Prior work (GrokMirror v1.0, Jan 2026) tested only the endpoints: $G_n^\wedge$ (shattered) and $G_n^\vee$ (trivially connected). We contribute:

1. **A complete phase diagram** across 25+ rule variants identifying the percolation threshold.
2. **Random baseline comparisons** showing where number-theoretic structure outperforms random graphs and where it doesn't.
3. **Discovery of genuine small-world structure** in the power-of-2 difference graph D_n.
4. **Growth dynamics** showing how properties scale with n.
5. **Spectral and robustness analysis** characterizing graph resilience.

---

2. The Phase Diagram

2.1 Methodology

We construct graphs at $n = 5,090$ (the real vault size) across 31 rule variants. For each, we measure: edge count, density, giant component fraction, diameter (sampled over 500 source nodes), average path length, and clustering coefficient. Results validated at $n = 50, 100, 200, 500$ for consistency. The v2.0 graph was verified edge-by-edge against the stored vault mesh audit (100.00% match, 1,587,325 edges).

All computations performed on Apple Mac Mini M4 (10-core ARM, 24GB RAM) running macOS 26.2. Total pipeline runtime: 585.6s.

2.2 Core Results (n = 5,090)

Rule	Edges	Density	Giant Comp.	Diameter	Clustering
v1.0 Strict AND ($G_n^\wedge$)	1,249	0.01%	0.1%	∞†	0.000
Prime AND Diff ≤ 8	4,988	0.04%	3.1%	37	0.000
Prime AND Diff ≤ 12	7,477	0.06%	13.1%	86	0.000
Prime AND Diff ≤ 16	9,961	0.08%	41.7%	190	0.000
Prime AND Diff ≤ 19	**12,441**	**0.10%**	**100%**	**340**	**0.000**
Prime AND Diff ≤ 32	19,861	0.15%	100%	187	0.000

Rule	Edges	Density	Giant Comp.	Diameter	Clustering
Prime AND Diff ≤ 64	39,517	0.31%	100%	88	0.000
Pow2 Diff Only (D_n)	57,979	0.45%	100%	7	**0.126**
Prime AND Diff ≤ 256	154,195	1.19%	100%	22	0.000
Prime Sum Only (P_n)	1,530,595	11.82%	100%	3	0.000
v2.0 Strategic OR $(G_n^{\vee})$	1,587,325	12.26%	100%	3	0.015

†*Disconnected: graph shattered into many isolated components, giant component has only 5 nodes.*

2.3 The Percolation Threshold

A fine-grained sweep over $k \in [8, 32]$ reveals a sharp phase transition:

Locality k	As $c \cdot \log_2 n$	Giant Frac.	Components	Connected?
8	$0.65 \log_2 n$	3.1%	840	No
11	$0.89 \log_2 n$	13.1%	172	No
13	$1.06 \log_2 n$	15.9%	74	No
15	$1.22 \log_2 n$	41.7%	18	No
17	$1.38 \log_2 n$	93.9%	3	No
19	$1.54 \log_2 n$	**100%**	**1**	**Yes**
20	$1.62 \log_2 n$	100%	1	Yes
24	$1.95 \log_2 n$	100%	1	Yes

Finding 1: The critical threshold for full connectivity in G_n^k is $k^* = 19 \approx 1.54 \cdot \log_2 n$ at $n = 5{,}090$. At this threshold, the graph achieves full connectivity with only **12,441 edges** (0.10% density) — 128x fewer edges than v2.0. The transition is sharp: at $k = 17$ the graph is 93.9% connected with 3 components, and at $k = 19$ it snaps to a single component.

2.4 Logarithmic Locality Scaling

Testing $k = c \cdot \log_2 n$ directly ($\log_2 5090 \approx 12.31$):

3

Multiplier c	Threshold k	Density	Giant Frac.	Diameter
0.5	6	0.03%	1.2%	20
1.0	12	0.06%	13.1%	86
1.5	18	0.09%	93.9%	360
2.0	**24**	**0.12%**	**100%**	**263**
3.0	36	0.17%	100%	165
4.0	49	0.24%	100%	114
8.0	98	0.46%	100%	56

The percolation threshold occurs between $c = 1.5$ and $c = 2.0$. The exact critical point is $k^* = 19$ $(c^* = 1.54)$.

---

3. Random Baseline Comparisons

3.1 Methodology

For each GrokMirror graph, we construct $G(n, m)$ Erdős–Rényi random graphs with **identical node and edge counts** (not approximate density matching). We run **10 independent seeds** per graph and report mean $\pm$ std to establish statistical confidence.

3.2 Results (n = 5,090; 10-seed ER baselines)

Graph	Metric	GrokMirror	ER mean ± std	Ratio GM/ER
Pow2 Diff	Clustering	0.1256	0.0045 ± 0.0002	**28.0**
Pow2 Diff	Transitivity	0.1253	0.0045 ± 0.0001	**27.8**
Pow2 Diff	Diameter	7	4.0 ± 0.0	1.75
Pow2 Diff	Avg Path	4.31	3.007 ± 0.004	1.43
Pow2 Diff	Degree σ	1.17	4.745 ± 0.029	0.25
Prime Sum	Clustering	0.000	0.1182 ± 0.0000	**0.00**
Prime Sum	Diameter	3	2.0 ± 0.0	1.50
Prime Sum	Avg Path	2.263	1.882 ± 0.000	1.20
Prime Sum	Degree σ	26.57	23.17 ± 0.14	1.15
v2.0 OR	Clustering	0.015	0.1226 ± 0.0000	**0.125**
v2.0 OR	Diameter	3	2.0 ± 0.0	1.50
v2.0 OR	Avg Path	1.878	1.878 ± 0.000	1.000
v2.0 OR	Degree σ	26.26	23.56 ± 0.14	1.11

All ER metrics show negligible variance across seeds (std < 0.03 for degree, < 0.005 for path length), confirming the comparisons are statistically robust.

3.3 Interpretation

Finding 2: The Power-of-2 Difference Graph Is Profoundly Non-Random.

At $n = 5{,}090$, D_n shows **28.0x higher clustering** than random (± 0.0002 std across 10 seeds — this is not noise). The degree distribution is nearly uniform ($\sigma = 1.17$ vs. 4.75 ± 0.03 for ER), confirming geometric structure. This effect *strengthens* with scale (it was 5.6x at $n = 500$).

Finding 3: The Prime-Sum Graph Is Anti-Clustered.

P_n has **zero clustering** — no triangles exist. This is a provable property (see Appendix A): for three nodes $i < j < k$ to form a triangle, all three sums $i+j$, $i+k$, $j+k$ must be prime. A parity argument shows this is impossible for distinct positive integers. Therefore P_n **is triangle-free** for all n. Meanwhile, ER at the same density gives $C = 0.1182 \pm 0.0000$ — the absence of triangles is a hard structural constraint, not a density effect.

Finding 4: The v2.0 OR Graph Is 8x Less Clustered Than Random.

$G_n^{\vee}$ has clustering ratio 0.125 — the OR combination inherits the anti-clustering of P_n (which contributes 96.4% of edges at $n = 5{,}090$) and heavily dilutes D_n's structural clustering. At average path length 1.878 vs ER's 1.878 ± 0.000, it is **statistically indistinguishable from a random graph** on path-length metrics.

---

4. The Pow2 Difference Graph: A Natural Small-World

4.1 Small-World Coefficient

The small-world sigma (σ) compares a graph's clustering and path length against random baselines:

$$\sigma = \frac{C/C_{rand}}{L/L_{rand}}$$

Graph	σ	C_{real}	C_{rand}	L_{real}	L_{rand}	γ	λ	Verdict
D_n (Pow2 Diff)	**19.96**	0.126	0.004	4.31	3.02	28.70	1.44	**Strong Small-World**
$G_n^{\vee}$ (v2.0 OR)	0.12	0.015	0.123	1.88	1.88	0.12	1.00	Not Small-World

Finding 5: The power-of-2 difference graph D_n is a **strong small-world network** ($\sigma = 19.96 \gg 1$) at vault scale. The clustering ratio $\gamma = 28.7$ means D_n has nearly 29x the clustering of a random graph, while the path ratio $\lambda = 1.44$ shows paths are only 44% longer. This effect **amplifies with scale** (σ rose from 4.25 at $n = 500$ to 19.96 at $n = 5{,}090$), indicating that D_n's small-world structure is intrinsic, not a finite-size artifact.

4.2 Why D_n Is Small-World

In D_n, node i connects to nodes at distances $1, 2, 4, 8, 16, \ldots$ This is structurally analogous to a **skip list** or **Kleinberg's navigable small-world model**: dense local connections (distance 1, 2) provide clustering, while long-range connections (distance 256, 512, …) keep diameter logarithmic.

4.3 Robustness (n = 5,090)

Removal %	Type	D_n Giant Frac.	$G_n^{\vee}$ Giant Frac.
5%	Random	1.000	1.000
10%	Random	1.000	1.000
20%	Random	1.000	1.000
30%	Random	1.000	1.000
50%	Random	1.000	1.000
5%	Targeted (highest-degree)	1.000	1.000
10%	Targeted (highest-degree)	1.000	1.000
20%	Targeted (highest-degree)	1.000	1.000
30%	Targeted (highest-degree)	1.000	1.000
50%	Targeted (highest-degree)	1.000	1.000

Both D_n and $G_n^{\vee}$ survive **50% targeted node removal** (highest-degree first) at $n = 5,090$. The near-uniform degree distribution of D_n (degree $\sigma = 1.17$, coefficient of variation 5.1%) means there are no critical hubs — every node is equally important, making the network maximally resilient to targeted attack.

---

5. Spectral Analysis

The algebraic connectivity λ_2 (second-smallest eigenvalue of the Laplacian) measures how well-connected a graph is:

Graph	λ_2	Interpretation
$G_n^{\wedge}$ (v1.0)	0.586	Weak (computed on 5-node giant component)
D_n (Pow2 Diff)	**3.039**	Strong intrinsic connectivity
$G_n^{\vee}$ (v2.0 OR)	332.2	Extremely high (density artifact)

D_n's algebraic connectivity of 3.039 is notable — it means the graph cannot be easily bisected, confirming robustness from a spectral perspective. The v2.0 value of 332.2 is an artifact of having 624 average connections per node.

5.1 Assortativity

Graph	Assortativity	Meaning
D_n (Pow2 Diff)	**+0.685**	Strongly assortative (similar-degree nodes cluster)
$G_n^\vee$ (v2.0 OR)	+0.051	Essentially random

Finding 6: D_n's strong assortativity (+0.685 at $n = 5,090$, up from +0.615 at $n = 500$) further distinguishes it from random graphs. This is characteristic of lattice-like and geometric networks, and consistent with D_n's structure where boundary nodes (lower degree) connect to other boundary nodes.

---

6. Growth Dynamics

6.1 How Properties Scale with n

n	D_n Density	P_n Density	$G_n^\wedge$ Density	$G_n^\vee$ Density	D_n σ
50	19.4%	24.1%	1.96%	41.5%	—
100	11.6%	21.1%	0.91%	31.8%	—
200	6.76%	18.5%	0.39%	24.8%	—
500	3.20%	16.0%	0.13%	19.1%	4.25
5,090	**0.45%**	**11.82%**	**0.01%**	**12.26%**	**19.96**

Key observations: - D_n **density** decreases as $O(\log n/n)$ — each node has $\sim 2\lfloor \log_2 n \rfloor$ neighbors, so density $= \Theta(\log n/n)$. At $n = 5,090$: mean degree 22.8, density 0.45%. - P_n **density** decreases slowly: by the prime number theorem, $\Pr[i + j \text{ is prime}] \approx \frac{1}{\ln(n)}$, so density $\sim \frac{1}{\ln n}$. - $G_n^\wedge$ **density** drops to 0.01% at vault scale — utterly shattered. - D_n **'s small-world sigma scales up** from 4.25 to 19.96, confirming the structure is intrinsic. - All three non-v1.0 graphs remain connected at all tested sizes.

6.2 Scaling Implications

At $n = 5,090$ (real vault), D_n has density 0.45% with ~23 connections per node, diameter 7, and small-world $\sigma = 19.96$. This is the **minimal connected navigable graph** — the tightest deterministic wiring that maintains both local clustering and global reachability. The critical percolation threshold graph G_n^{19} achieves connectivity at 0.10% density with only 12,441 edges, but lacks clustering — D_n trades a 4.5x increase in edges for genuine small-world structure.

7

7. Honest Assessment of v2.0

7.1 What the Original Paper Claimed

The original GrokMirror v2.0 paper claimed the "Strategic OR" validation as a primary result, with diameter 3 and 90% giant component as evidence of mathematical integrity.

7.2 What the Data Shows

At $n = 5,090$, $G_n^\vee$ has: - **1,587,325 edges** at 12.26% density - Average degree **624** (each node connected to 12% of all others) - Diameter 3, fully connected

An Erdős–Rényi random graph at the same density achieves **diameter 2** with 100% connectivity. The v2.0 result is therefore a **density artifact**, not a structural property: - Path length: GM 1.878 vs ER 1.877 (ratio 1.000 — identical) - Clustering: GM 0.015 vs ER 0.123 (ratio 0.125 — v2.0 is **8x worse**) - The prime-sum condition alone creates 96.4% of edges

7.3 Reframing the Contribution

The actual contributions of this work are:

1. **The percolation threshold at** $k^* = 19 \approx 1.54 \cdot \log_2 n$ — the tightest locality constraint under which prime-sum graphs remain connected, with only 12,441 edges.
2. D_n **as a strong small-world** ($\sigma = 19.96$) — a pure number-theoretic construction with 29.3x the clustering of random graphs, strengthening with scale.
3. P_n **is triangle-free** — a provable structural property with potential applications in bipartite-like mesh design.
4. **The v1.0 refutation** — empirical demonstration that the AND intersection is irrecoverably sparse (0.01% density at vault scale).

---

8. Reproducibility

8.1 Software

All analysis code is available in the `analysis/` directory:

File	Purpose
`grokmirror_core.py`	Edge predicates, graph construction, all metrics
`phase_diagram.py`	Phase diagram computation (25+ rule variants)
`baselines.py`	Erdős-Rényi baseline and advanced analysis
`run_all.py`	Full pipeline orchestrator (n≤1000)
`run_n5090.py`	Optimized pipeline for real vault scale
`validate_vault.py`	Edge-by-edge validation against stored vault mesh
`test_grokmirror.py`	Test suite (37 tests, all passing)

8.2 Environment

```
Platform: macOS-26.2-arm64 (Mac Mini M4, 10-core, 24GB)
Python: 3.9.6
networkx: 3.2.1
numpy: 2.0.2
scipy: 1.13.1
```

8.3 Running

```
# Run tests (37 tests, <1s)
python3 -m pytest test_grokmirror.py -v

# Run quick analysis (n=500, ~20s)
python3 run_all.py --n 500 --full

# Run full vault-scale analysis (n=5090, ~10 min)
python3 -u run_n5090.py

# Validate against stored vault mesh (100% edge match)
python3 validate_vault.py

# Results written to analysis/results_n5090/
```

8.4 Data

The original vault mesh audit (`vault_mesh_audit.json`, 37MB) contains the full adjacency list for $n = 5{,}090$ under v2.0 rules. Validation confirms 100.00% edge match (1,587,325 edges) between stored data and recomputed graph.

8.5 Results Files

File	Contents
`phase_diagram_n5090.json`	All 31 rule variants with full metrics
`baselines_n5090.json`	Erdős-Rényi comparisons for 4 primary rules
`advanced_n5090.json`	Spectral, small-world, robustness analysis
`percolation_detail_n5090.json`	Fine-grained sweep k=8..32
`vault_validation.json`	Edge-by-edge validation report
`master_results_n5090.json`	Aggregated master results

9. Conclusion

We have mapped the complete connectivity landscape of GrokMirror's number-theoretic graphs across 31 rule variants at real vault scale ($n = 5{,}090$). The headline result is not the v2.0 validation (which is trivially dense and structurally indistinguishable from random), but three discoveries:

1. A **sharp percolation threshold** at $k^* = 19 \approx 1.54 \cdot \log_2 n$, achieving full connectivity at just 0.10% density — 128x fewer edges than v2.0.
2. The **power-of-2 difference graph** D_n as a strong small-world network ($\sigma = 19.96$, scaling up from 4.25 at $n = 500$) with 29.3x the clustering and 4x lower degree variance than equivalent random graphs, plus complete robustness under 50% targeted node removal.
3. The **triangle-free property** of prime-sum graphs P_n, provable for all n.

For sovereign mesh applications, the recommended wiring rule is D_n: logarithmic diameter, small-world clustering, near-uniform degree distribution, and complete robustness — all from a deterministic, coordination-free, O(1)-computable rule. Any node can compute its full neighbor set in constant time: connect to $i \pm 2^k$ for all k where the neighbor is in $[1, n]$.

Appendix A: Proof that P_n is Triangle-Free

Theorem. For $n \geq 2$, the prime-sum graph P_n contains no triangles.

Proof. Suppose vertices $i < j < k$ form a triangle. Then $i + j$, $i + k$, and $j + k$ are all prime.

Consider parities. If i is odd and j is even (or vice versa), then $i + j$ is odd — fine. But $i + k$ and $j + k$ have different parities (since i and j have different parities). So one of $\{i + k, j + k\}$ is even. The only even prime is 2, requiring $i + k = 2$ or $j + k = 2$. Since $i, j, k \geq 1$ and $j > i$, we need $k = 1$ and $i = 1$, giving $j > 1$. Then $j + k = j + 1$ and $i + j = 1 + j$ — these are equal, so all three sums are $j + 1$, $j + 1$, and $1 + 1 = 2$. For the triangle, $i + j = 1 + j$ and $j + k = j + 1$ must be prime, so $j + 1$ is prime. And $i + k = 2$ is prime. This works — but $i = 1$, $k = 1$ contradicts $i < j < k$.

If i and j have the same parity, then $i + j$ is even. For $i + j$ to be prime, $i + j = 2$, so $i = j = 1$ — contradicting $i < j$.

Therefore no triangle exists. □

Appendix B: Degree Distribution of D_n

In D_n, node i connects to $i \pm 2^0, i \pm 2^1, \ldots, i \pm 2^{\lfloor \log_2 n \rfloor}$ (where the neighbor is in $[1, n]$). Interior nodes have degree $2(\lfloor \log_2 n \rfloor + 1)$; boundary nodes have fewer connections. The degree distribution is therefore approximately uniform at $2 \log_2 n$, with a narrow band at the boundaries.

This is confirmed empirically:

n	Mean Degree	Degree σ	CV (%)
500	15.96	1.39	8.7%
5,090	22.78	1.17	5.1%

The coefficient of variation *decreases* with n, confirming that the degree distribution becomes increasingly uniform at scale.

Metadata for Zenodo

- **Upload Type:** Publication ☐ Preprint
- **DOI:** [Pending Submission]
- **License:** Creative Commons Attribution 4.0 International
- **Communities:** MirrorDNA, Sovereign AI, Graph Theory, Complex Networks

```
@preprint{desai2026grokmirror,
  title={Percolation Thresholds in Number-Theoretic Graphs:
        From Shattered to Small-World Connectivity
        in Sovereign Knowledge Meshes},
  author={Desai, Paul and Claude Code and Grok 4},
  year={2026},
  institution={N1 Intelligence},
  address={Goa, India},
  note={Empirical analysis code and data available at
        https://doi.org/[PENDING]}
}
```