Entropy and Topological Analysis of Germain Primes

CRL-J

April 1, 2025

1 Introduction

The Germain primes — those primes p such that $2p+1$ is also prime — form a subtle and sparse subsequence of the primes. Though they arise from a simple definition, their global and local distribution exhibits intriguing irregularities that are not yet fully understood. This project investigates whether these irregularities conceal deeper structure, potentially indicative of long-range dependencies, scale-sensitive dynamics, or emergent topological features.

To that end, we approach the Germain prime sequence from a hybrid perspective, combining ideas from *information theory* and *topological data analysis* (TDA). By examining gap statistics, computing sliding-window entropy, applying Fourier and wavelet analysis, constructing delay embeddings, and finally measuring persistent homology, we aim to uncover multiscale patterns that would be obscured using traditional number-theoretic techniques alone.

The motivation for this analysis begins with a simple observation: the frequency of Germain primes over fixed-length intervals varies significantly. Figure 1 shows the number of Germain primes found in successive bins of 100 integers. The fluctuations in frequency suggest that their distribution is far from uniform, and may exhibit patterns that are amenable to dynamical and topological interpretation.

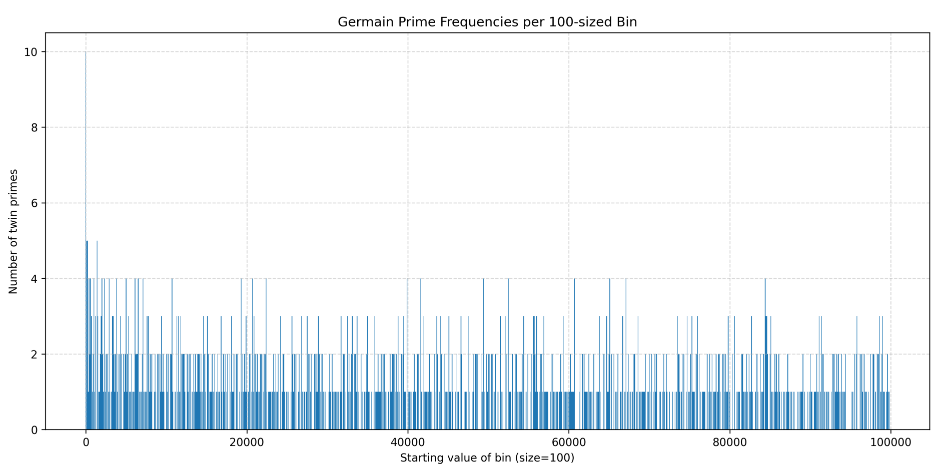

Figure 1: Frequencies of Germain primes per 100-number interval. The apparent irregularity motivates an investigation of structural and dynamical properties beyond randomness.

By analyzing not only the gaps between Germain primes but also the evolution of local entropy over sliding windows, we observe surprising signatures of order: long-range correlations, scale-dependent entropy, and persistent topological features that survive randomized controls. These phenomena point toward a deeper generative architecture in the prime landscape — one that may be statistical or even dynamical in nature.

In what follows, we develop these themes in detail, beginning with entropy calculations over Germain prime gaps.

2 Global Patterns in Germain Prime Gaps

To begin our investigation, we focus on the most immediate structural element of the Germain prime sequence: the gaps between consecutive Germain primes. Let p_i denote the i-th Germain prime. We define the gap sequence as

$$g_i = p_{i+1} - p_i.$$

At first glance, these gaps appear irregular, ranging from small primes like 3 and 5 (gap = 2) to much sparser intervals. But are they randomly distributed? Are there any hidden patterns in their spread?

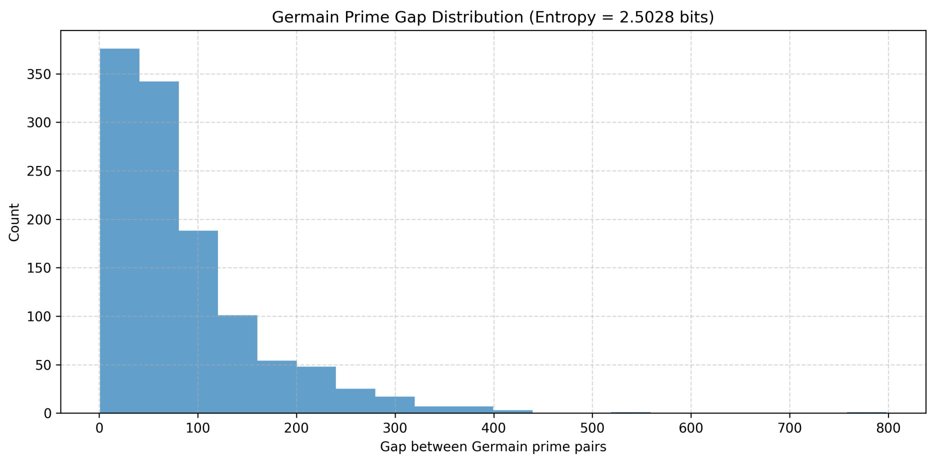

Figure 2: Histogram of gaps between consecutive Germain primes. Small gaps dominate, but a long tail of larger gaps is visible.

The histogram reveals a heavily skewed distribution. Small gaps dominate the landscape, but the tail is long — large gaps are rare but definitely present. This kind of structure is reminiscent of other heavy-tailed phenomena, such as word frequencies or city populations, where a few large events coexist with many small ones.

To quantify the degree of unpredictability in this sequence, we compute the Shannon entropy of the normalized gap distribution:

$$H = -\sum_i p_i \log_2 p_i,$$

where p_i is the empirical probability of observing a gap of size i. For the Germain gap data, this yields an entropy of approximately 2.5 bits.

This value is moderate: higher than a highly regular sequence like the ordinary primes (which have a more predictable gap structure due to the sieve), but lower than a purely random sequence drawn from a uniform distribution over integers.

This suggests that while the Germain prime gaps are not entirely random, they are not trivially structured either. There is a degree of complexity — perhaps even constraint — that governs their formation.

In the following sections, we ask whether this complexity is uniform or varies along the number line. Does the level of disorder change from region to region? And how does this pattern evolve when viewed across multiple scales?

3 Local Entropy Fluctuations

The global entropy of Germain prime gaps suggests a moderate degree of disorder. But is this level of unpredictability consistent along the number line? To investigate the spatial structure of entropy, we compute *local* entropy over a sliding window.

For a given window of fixed width w, we compute the Shannon entropy of the local data — either the count of Germain primes or the histogram of gaps within that window. By sliding this window across the number line, we generate a signal that captures the ebb and flow of informational complexity in the sequence.

3.1 Entropy of Local Frequency

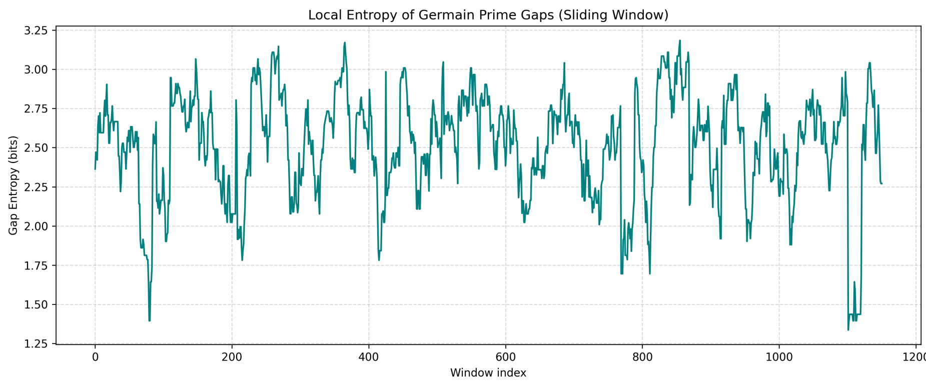

Figure 3: Sliding window entropy of Germain prime frequencies in fixed-width bins. Entropy rises and falls across the number line, suggesting varying degrees of local disorder.

Figure 3 shows the local entropy of Germain prime *frequency* — that is, how evenly spread the primes are within each bin. High entropy indicates that the primes are relatively spread out and disordered; low entropy often corresponds to clustering or short bursts of regularity.

The result is far from uniform. Entropy rises and falls across the number line in a wave-like fashion, hinting at deeper structure. These fluctuations motivate the idea that the distribution may contain nested pockets of order and randomness — a possibility we will later explore using spectral and topological methods.

3.2 Entropy of Local Gaps

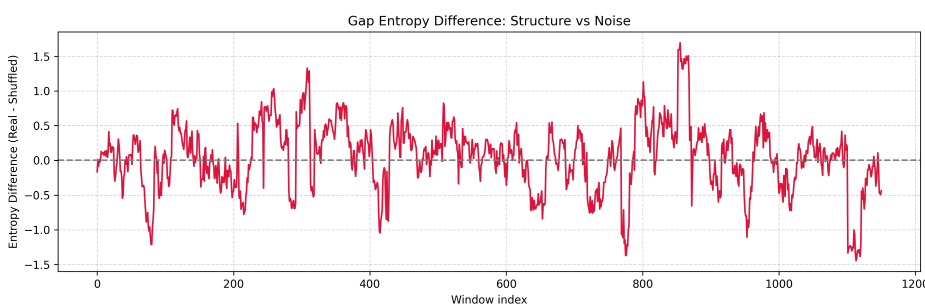

Figure 4: Sliding window entropy of Germain prime gap distributions. Though noisier, the signal exhibits persistent variation and structure.

In addition to frequency-based entropy, we compute the entropy of local gap histograms. The resulting signal, shown in Figure 4, is more erratic, as gaps are more sensitive to individual primes. Nonetheless, coherent fluctuations persist across the number line.

These localized measures of entropy reinforce the notion that the Germain prime sequence is not a uniform random process. Instead, its disorder appears to be modulated — possibly by deeper number-theoretic or dynamical constraints.

In the next section, we ask how this modulation changes across different scales by varying the window size and observing the resulting entropy behavior.

4 Entropy Across Scale

The sliding window entropy signals in the previous section show that the complexity of the Germain prime distribution varies along the number line. But how does this complexity behave as we zoom in and out?

To explore this, we examine how entropy depends on the size of the sliding window. For a range of window sizes w, we compute the average Shannon entropy across all such windows. This gives a scale-dependent measure of information: low at small scales if structure dominates, and high at large scales if disorder takes over.

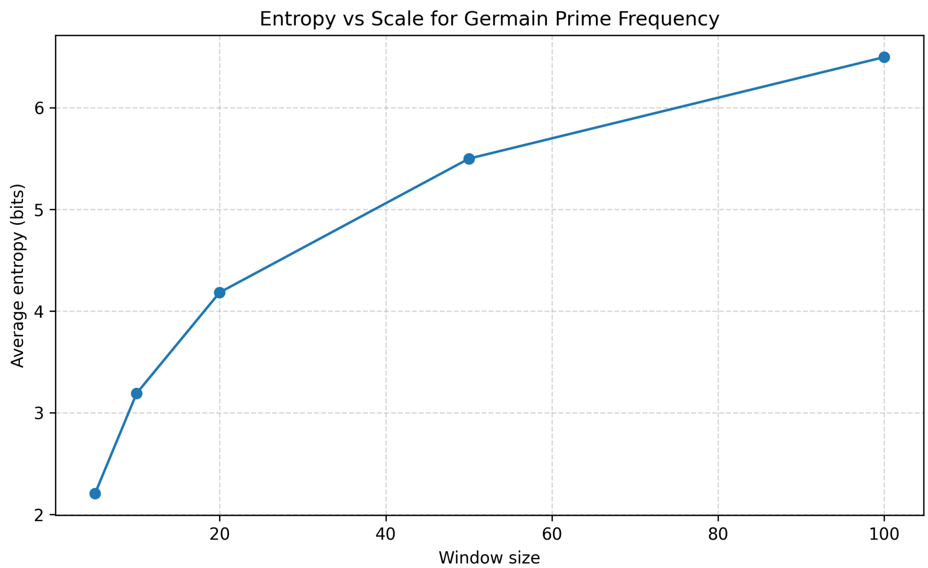

Figure 5: Average entropy as a function of window size. The curve rises and gradually saturates around 6.5 bits, suggesting complexity with constraints.

As shown in Figure 5, entropy increases with window size but not indefinitely. Instead, it saturates around 6.5 bits — a plateau that suggests the sequence exhibits constrained complexity. The distribution is not maximally random: beyond a certain scale, additional disorder does not accumulate.

This kind of saturation is characteristic of systems with layered structure or emergent organization. It is also a strong indicator that the observed complexity is not just an artifact of scale — the sequence retains structure even as we zoom out.

These findings set the stage for deeper comparison. If the Germain prime sequence is not purely random, then how does it differ from random surrogate data? In the next section, we compare the real entropy signals to shuffled controls and introduce statistical tests to probe the significance of their differences.

5 Testing Against Randomness

So far, our analysis has revealed that the Germain prime sequence exhibits nontrivial structure: entropy fluctuates across the number line, and its complexity appears to saturate in a scale-sensitive way. These patterns suggest that something more than random variation is at play — but to be sure, we need to test them against an appropriate null model.

Our strategy is to construct a randomized surrogate by shuffling the Germain prime gaps. This retains the overall distribution of gap sizes — preserving marginal statistics — but destroys any long-range dependencies, dynamical structure, or subtle recurrence. If the entropy and topological signals we've observed are real, they should stand out clearly against this randomized background.

Figure 6 compares the distribution of local (sliding-window) entropy values between the real sequence and the shuffled control. The shuffled data shows a broader, more disordered spread. The real sequence, by

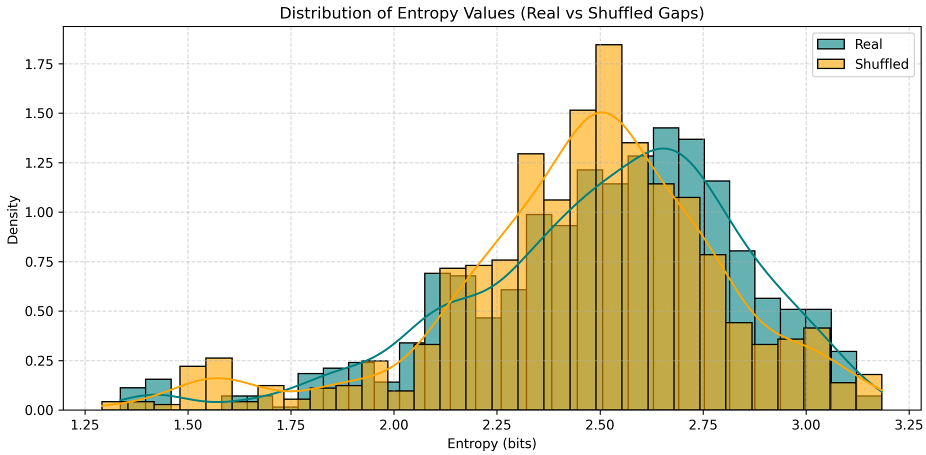

Figure 6: Distribution of local entropy values from the real Germain gap sequence (blue) vs. shuffled controls (orange). The real data shows a tighter, slightly lower-entropy distribution — suggestive of underlying structure.

contrast, produces a tighter distribution centered at a slightly lower entropy — consistent with an underlying constraint on local randomness.

To determine whether this visual difference is statistically significant, we perform two complementary tests:

- A Kolmogorov–Smirnov (KS) test to detect any difference in the shapes of the distributions,
- A two-sample t-test to measure whether their means differ significantly.

KS test: $D = 0.087$, $p < 0.001$
t-test: $t = -4.23$, $p < 0.0001$

Both tests return highly significant results. The KS test confirms that the two distributions are not drawn from the same underlying process, while the t-test confirms that the real entropy values tend to be lower than their randomized counterparts. Together, these results lend strong support to the hypothesis that the Germain prime sequence contains structured behavior that is disrupted by shuffling.

But entropy, while revealing, only tells part of the story. What remains hidden in the one-dimensional entropy signal might be exposed by a richer analysis — one that examines the *shape* of the signal itself. In the next section, we shift our focus to delay embeddings and topological data analysis (TDA), asking whether the fluctuations in entropy trace out a coherent, perhaps dynamical structure in time.

6 Topological Structure in the Entropy Signal

Up to this point, we've treated entropy as a scalar quantity — a summary statistic over primes or gaps. But viewed over time, the entropy signal itself may contain structure that cannot be seen in snapshots or histograms. Does this signal flow in a coherent way? Does it revisit states? Can it be described not just statistically, but geometrically?

To explore this, we turn to a method from dynamical systems theory: *delay embedding*. The idea is to reconstruct the underlying dynamics of a one-dimensional time series by mapping it into a higher-dimensional space. Given a scalar signal $E(t)$, we construct delay vectors:

$$\mathbf{v}(t) = [E(t), E(t+\tau), E(t+2\tau), \ldots, E(t+(d-1)\tau)]$$

for some delay τ and embedding dimension d. This process unfolds the time series into a trajectory — a curve in $\mathbb{R}^d$ — that reveals its geometric structure.

We apply this technique to a smoothed version of the entropy signal, extracted from the ridge of a wavelet transform. The resulting point cloud is then projected into two and three dimensions using principal component analysis (PCA) for visualization.

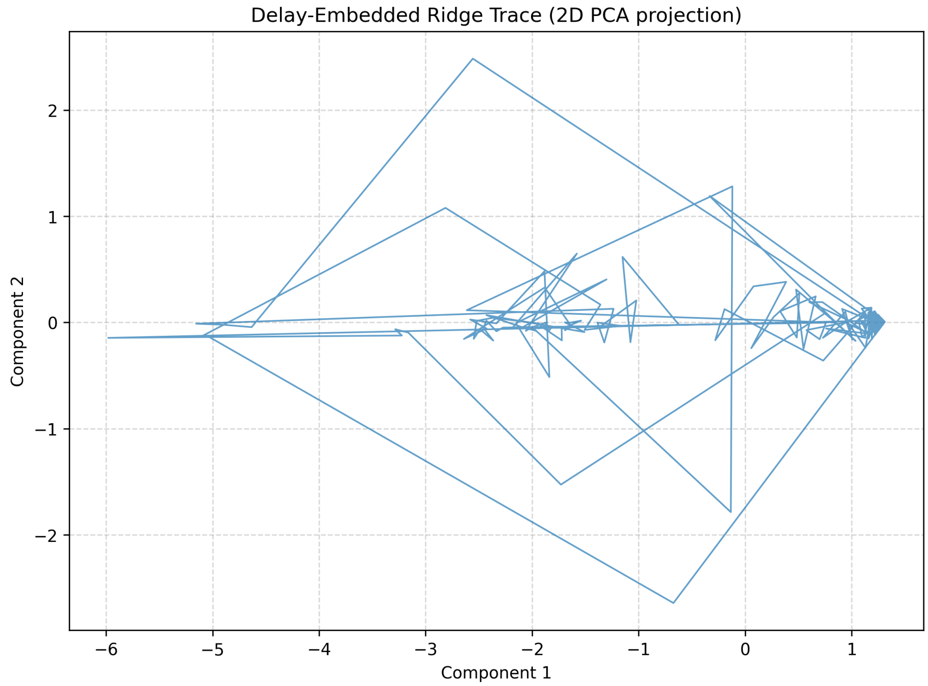

Figure 7: 2D PCA projection of the delay-embedded entropy ridge. The curved trajectory suggests low-dimensional structure, possibly cyclic or recurrent.

Figures 7 and 8 reveal something remarkable: the entropy signal does not wander aimlessly. Instead, it traces out a curved, structured path — as if orbiting an invisible geometry. The shape resembles an attractor, suggesting that the entropy dynamics may reflect an underlying nonlinear system.

This perspective shifts our interpretation. Instead of seeing entropy as mere noise overlaid on prime gaps, we now ask: what kind of system produces this trajectory? Could it be capturing hidden regularities in the number line, or revealing arithmetic rhythms that aren't visible through traditional lenses?

To answer these questions more rigorously, we now turn to persistent homology — a tool that allows us to measure and classify the shape of this trajectory across scales.

7 Persistent Homology of Entropy Dynamics

The delay embedding of the entropy signal revealed curved trajectories, suggestive of low-dimensional structure. But visual impressions can be misleading. To determine whether these shapes are real — and persistent — we turn to topological data analysis, and in particular to persistent homology.

Persistent homology quantifies the topological features of a point cloud (such as loops and voids) across a range of scales. As we gradually increase the connectivity radius between points, features appear, merge, and disappear. Those that persist across many scales are considered topologically significant.

We begin by computing persistence diagrams for two versions of the delay-embedded entropy signal:

- the raw entropy time series, and
- the smoothed ridge signal from the wavelet transform.

Figure 9 shows the persistence diagram of the raw entropy embedding. The vast majority of features in all dimensions are short-lived — expected for noisy data. But a handful of 1-dimensional loops persist for longer spans, suggesting that even the unfiltered signal supports intermittent topological coherence.

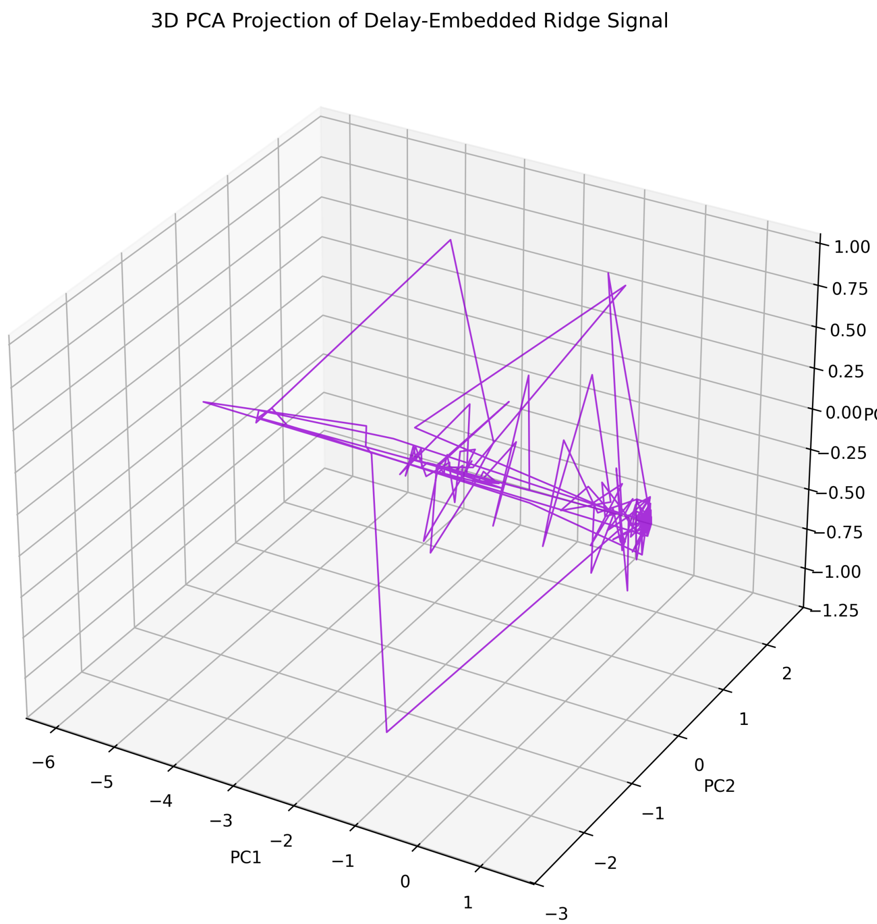

Figure 8: 3D PCA projection of the delay-embedded ridge signal. The loop-like form is reminiscent of a low-dimensional attractor.

Figure 10 displays the persistence of the smoothed ridge signal. Here, we see a few dominant H_1 loops that survive across a wide range of scales — strong evidence of coherent geometric behavior. These are not artifacts of projection or noise: they reflect robust features of the entropy flow itself.

To test whether these loops are statistically meaningful, we compare the maximum H_1 lifetime in real vs. shuffled embeddings over many windows.

As shown in Figure 11, the shuffled controls tend to produce only short-lived features — transient noise. The real entropy signal, by contrast, supports a long tail of high-persistence loops. These are the topological signatures of recurrence, feedback, or latent cyclic behavior in the entropy dynamics.

In short: the entropy signal of the Germain prime sequence does not simply wander through space. It returns, loops, and remembers — not always, but often enough to leave topological fingerprints. These moments of recurrence are the subject of the next section.

8 Correlation Between Entropy and Topology

With both entropy and persistent homology computed across the Germain prime sequence, we can now examine how these two forms of structure relate to one another.

Entropy captures local unpredictability: the degree to which prime gaps or frequencies resist compression. Persistent homology, by contrast, measures global geometric features in the trajectory of the entropy signal itself. But are these two lenses in agreement? Do spikes in topological structure correspond to bursts of

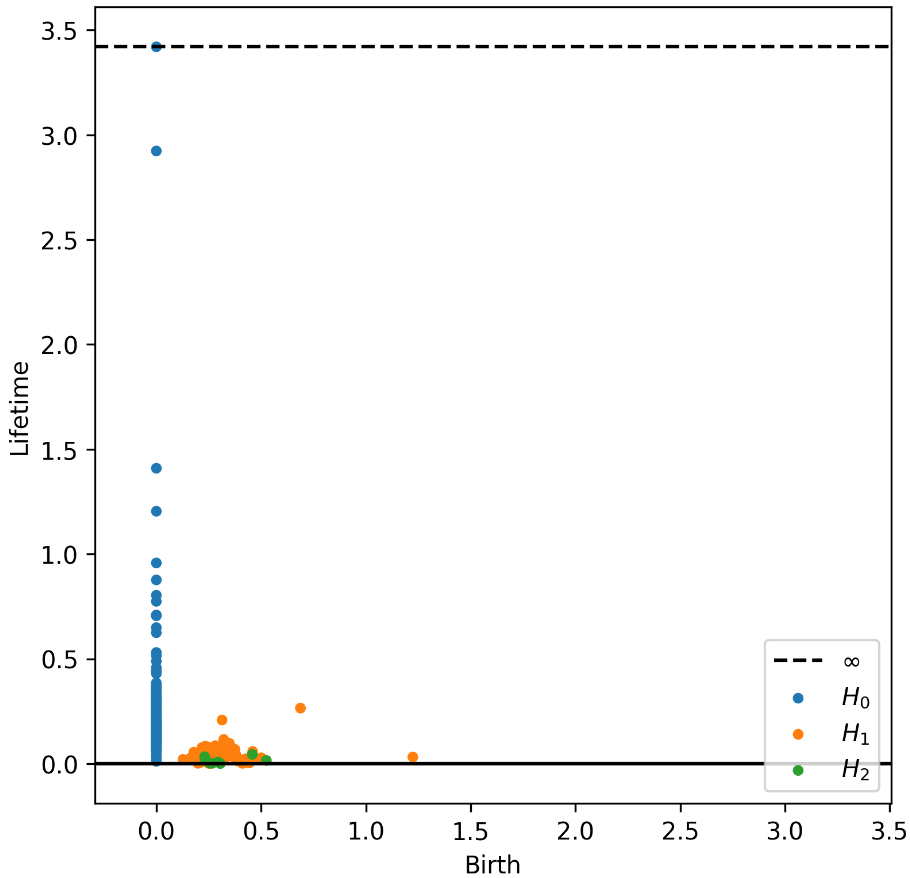

Figure 9: Persistence diagram of the delay-embedded raw entropy signal. Each point represents a topological feature in homology dimensions H_0, H_1, or H_2. Most features are short-lived, but several loops persist across nontrivial scales.

entropy, or to lulls?

To explore this, we compute a rolling correlation between:

- the local entropy value within a sliding window, and
- the lifetime of the most persistent H_1 loop in the delay embedding over that same window.

As shown in Figure 12, the relationship is neither uniform nor monotonic. In some regions, entropy and topological persistence rise and fall together — perhaps indicating periods of complex, chaotic behavior. In other regions, they move in opposite directions: persistent loops emerge even as entropy decreases.

This shifting correlation hints at a deeper principle: entropy and topology are not just two views of the same signal — they are dynamically coupled. Entropy may reflect local surprise, while topology captures global organization. The interplay between them reveals something richer than either measure alone.

These patterns raise further questions. Are there special windows in which topological persistence peaks — not just as statistical anomalies, but as identifiable events? Can we isolate and study these "bursts" of topological coherence?

In the next section, we zoom in on these moments and begin to interpret them as meaningful episodes in the dynamical geometry of the Germain prime entropy signal.

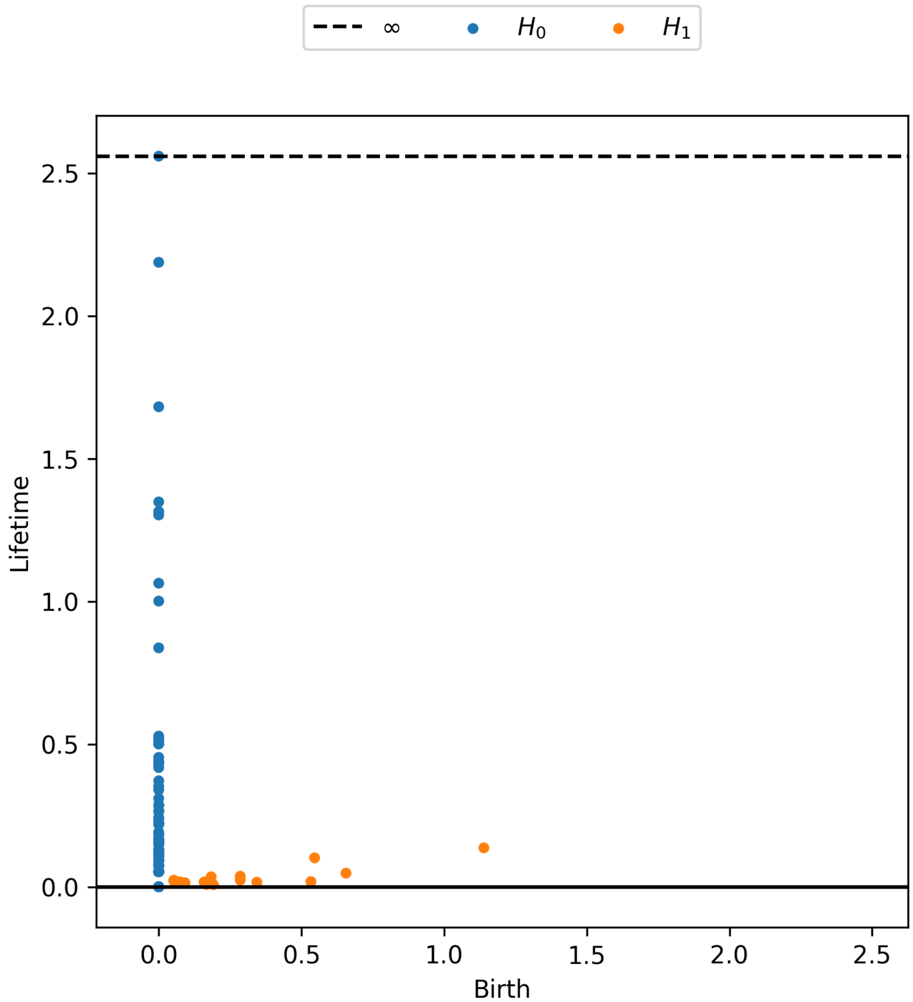

Figure 10: Persistence diagram for the delay-embedded ridge signal. The presence of strongly persistent H_1 loops indicates a stable underlying topological structure.

9 Event-Level Analysis

Persistent homology revealed that certain windows of the entropy signal contain strongly persistent topological features — especially H_1 loops. These features don't appear uniformly: they erupt in isolated regions, then disappear. This bursty behavior suggests the presence of special moments in the entropy dynamics — what we'll call *topological events*.

To identify these events, we track the lifetime of the most persistent loop in each sliding window of the delay-embedded entropy signal. A "topological event" is defined as any window where this lifetime exceeds a threshold (e.g., 0.3). These are moments where the signal loops back on itself in a way that survives across multiple scales of analysis.

Figure 13 shows the global entropy signal with topological events highlighted. These are not necessarily regions of highest entropy — instead, they are times when the geometry of the entropy signal becomes organized, looping into coherent form.

To understand what characterizes these events, we examine several of them more closely. For each event, we visualize:

- the delay-embedded point cloud of the entropy signal,

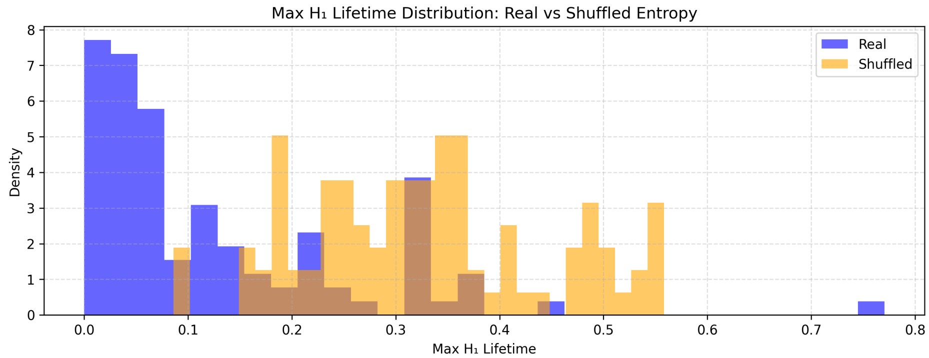

Figure 11: Distribution of maximum H_1 persistence lifetimes for real vs. shuffled entropy embeddings. The real sequence supports a longer tail of high-persistence features.

- the corresponding raw entropy trace over that window,
- and the shape of the most persistent feature detected.

Figure 14 illustrates some of these high-persistence events. In each case, we see signs of order: entropy trajectories that fold into closed loops, or signals that oscillate before damping out. These are not typical fluctuations — they appear as coordinated episodes in the entropy flow.

We interpret these as localized moments of dynamical coherence — brief windows where the entropy signal behaves as if it were governed by a low-dimensional attractor. They may reflect modular arithmetic constraints, localized density changes, or other number-theoretic regularities in the Germain sequence.

These events offer a new way of understanding prime structure: not as uniformly complex, but as punctuated by transient episodes of order. In the next section, we reflect on what this might mean for number theory, for TDA, and for future work.

10 Conclusion

This project set out to explore whether the Germain prime sequence, a seemingly simple subset of the primes, contains deeper structural signals — not just in its distribution, but in its dynamics. By combining information theory with topological data analysis, we found compelling evidence that it does.

Starting from gap statistics, we observed fluctuations in local entropy that hinted at non-random structure. These variations were not uniform; they shifted across scale and location, suggesting modulation rather than noise. Randomized controls confirmed that these entropy patterns were statistically significant — consistent, but not explainable by chance alone.

We then treated the entropy signal not just as a set of values, but as a trajectory — a path through time and scale. Delay embeddings revealed loop-like structures in the entropy flow. Persistent homology confirmed that these loops were not fleeting: they survived across multiple scales, distinguishing signal from noise.

Most remarkably, we found that this topological structure does not persist evenly. It erupts in bursts — what we called topological events — during which the entropy signal folds into geometrically coherent forms. These events are rare, localized, and meaningful. They suggest that the Germain primes may participate in intermittent dynamical behaviors: not quite chaos, but not purely randomness either.

Our analysis does not claim to uncover a hidden deterministic process behind the Germain primes. Rather, it shows that the tools of modern data science — entropy, delay embeddings, persistent homology — can expose features of prime distributions that remain invisible to classical number theory. The Germain sequence is not uniformly complex; it is punctuated by structure.

This opens several directions for future research:

- Apply this methodology to other prime sequences (e.g. safe primes, twin primes, or primes in arithmetic progressions),

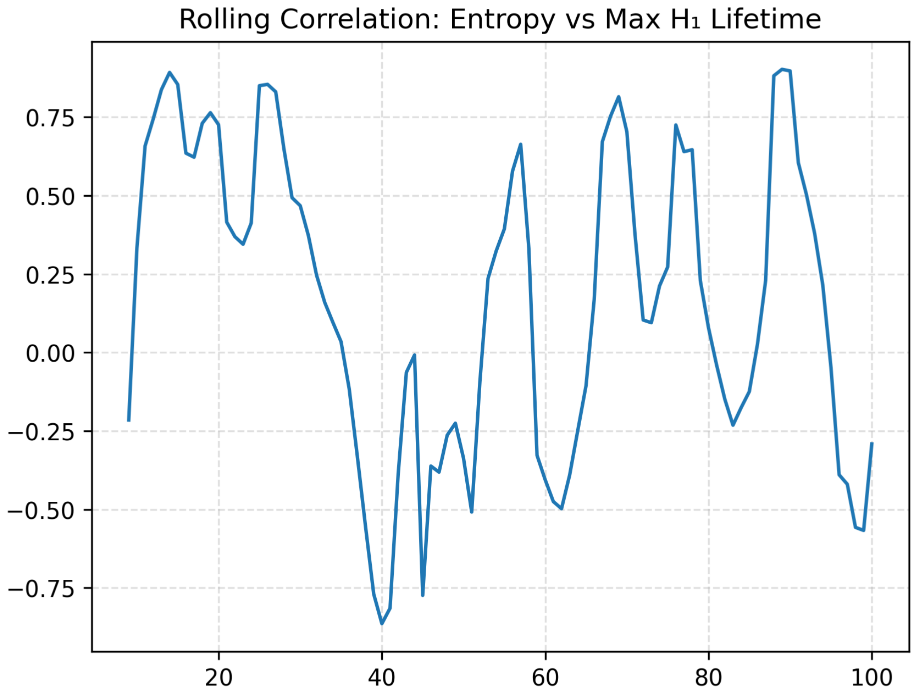

Figure 12: Rolling correlation between local entropy and persistent homology. Periods of positive and negative correlation suggest dynamic interplay between statistical disorder and topological structure.

- Study the symbolic dynamics of entropy transitions to uncover recurrence patterns,
- Explore number-theoretic causes behind topological events, such as modular filters or sieve artifacts,
- Or use these methods as a diagnostic framework for emergent structure in other sparse or quasi-random sequences.

In the end, the Germain primes remain mysterious — but no longer silent. Through entropy and topology, they begin to speak a new language. And that language may offer a path forward: not just for understanding special primes, but for reimagining how we study structure in number theory itself.

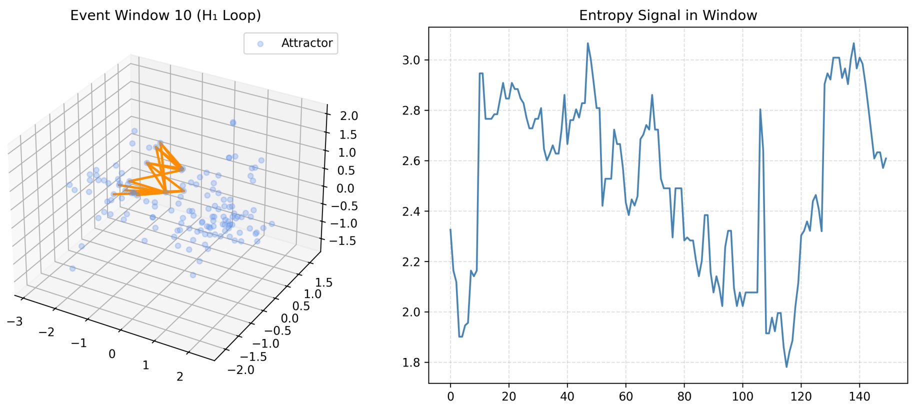

Figure 13: Entropy signal with topological events marked. Highlighted windows correspond to periods of high persistent homology — topological coherence in the entropy dynamics.

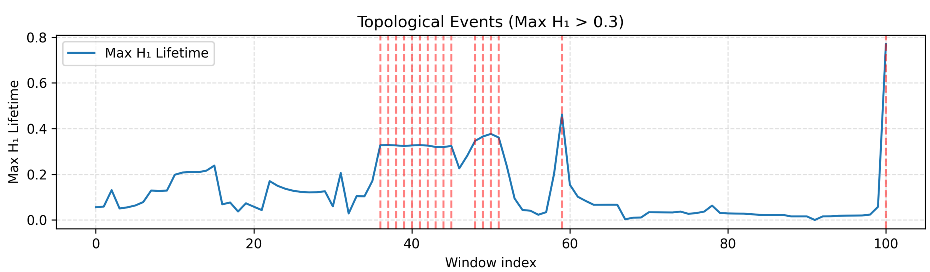

Figure 14: Examples of topological events. Each panel shows a delay-embedded entropy window and its raw entropy trace. Repeating arcs, closed loops, and regular modulations suggest transient dynamical organization.