Topological Signatures in the Entropy of Twin Prime Distributions

CRL-J

March 25, 2025

Abstract

We investigate the distribution of twin primes using tools from information theory and topological data analysis. By computing Shannon entropy in sliding windows over twin prime counts, we quantify local unpredictability and uncover intermittent structure in the distribution. To probe the dynamics of this structure, we apply delay embeddings to the entropy signal and analyze the resulting trajectories using persistent homology. This embedding-based approach reveals statistically significant topological features—including long-lived cycles—that are absent in randomized controls. While the average entropy of twin prime gaps is comparable to that of shuffled data, the real distribution exhibits heavier tails and recurrent patterns, suggesting low-dimensional organization amid apparent randomness.

1 Introduction

Prime numbers are the building blocks of the integers, yet their global distribution defies simple characterization. Among their many fascinating substructures, *twin primes*—pairs of primes that differ by 2—have attracted sustained attention due to their apparent rarity and unresolved infinitude. The Twin Prime Conjecture, asserting the existence of infinitely many such pairs, remains one of the great unsolved problems in number theory. Recent progress has established that infinitely many prime gaps are bounded (notably, Zhang's breakthrough on primes at most 70 million apart), but the twin prime case (gap 2) is still open.

In this work, we take a novel approach to investigating the structure of twin primes by treating their distribution as a time series and subjecting it to a joint analysis using tools from information theory and topological data analysis (TDA). Specifically, we examine the *local entropy* of twin prime occurrences and inter-gap sequences, and analyze their evolving structure using *delay embeddings* and *persistent homology*. Our goal is to detect moments of nontrivial organization within the otherwise seemingly irregular appearance of twin primes.

By computing the Shannon entropy of twin prime counts in fixed-width intervals, we generate a time series that captures local irregularity in the distribution. Shannon entropy H is defined as $H = -\sum_i p_i \log p_i$ for a probability distribution $\{p_i\}$ [1]. In our context, within a moving window of W consecutive bins of integers, we treat the counts of twin prime pairs in those bins as a discrete distribution (normalized by the total in the window) and compute its entropy. A higher entropy indicates a more uniform (disordered) local distribution of twin primes across that window, whereas lower entropy signifies that the twin primes are more concentrated in certain regions (more structured or clustered). By sliding this window along the number line, we obtain a *local entropy time series* reflecting how the unpredictability of twin prime distribution changes with number range.

Next, we apply time-delay embedding techniques to this entropy time series. A *delay embedding* reconstructs a multidimensional state space from a single time series by considering vectors of the form $(E(t), E(t+\tau), E(t+2\tau), \ldots, E(t+(m-1)\tau))$, where $E(t)$ is the entropy at window position t, τ is a delay, and m is the embedding dimension [8]. By plotting these vectors in $\mathbb{R}^m$, we obtain a geometric trajectory (attractor reconstruction) whose shape reflects the underlying dynamics or patterns in the entropy signal. We use principal component analysis (PCA) to project this trajectory to two and three dimensions for visualization.

Finally, we use persistent homology—a tool from TDA—to identify and quantify loop-like structures in these embedded entropy trajectories. *Persistent homology* computes topological features (connected components, loops, voids, corresponding to homology groups H_0, H_1, H_2, etc.) of a dataset across multiple scales [5]. The result is summarized in a *persistence diagram*, in which each topological feature is represented as a point whose coordinates correspond to the scale at which the feature appears ("birth") and disappears ("death"). Features that persist across a large range of scales (points far from the diagonal) are regarded as significant shape characteristics of the data rather than noise. In our analysis, we focus on persistent H_1 features, which correspond to one-dimensional loops in the delay-embedded entropy trajectory. Such loops can be interpreted as signatures of recurrence or cyclic behavior in the underlying signal.

Our core findings show that the twin prime distribution is not entirely random; instead, it exhibits specific windows of increased organization. We discover multiple segments of the number line where the delay-embedded entropy trajectory contains a prominent persistent loop (signified by a long-lived H_1 feature in the persistence diagram), suggesting the presence of an underlying quasi-cyclic structure in the distribution of twin primes. These "topological events" are absent in randomized control data, indicating that they are intrinsic to the twin prime sequence rather than artifacts of our analysis. We further characterize these events through statistical comparison with shuffled sequences and examine how they align with fluctuations in the entropy signal.

To our knowledge, this is the first study to integrate entropy analysis, delay embedding, and persistent homology in the context of prime number distributions. While we do not claim to uncover a deterministic generative model or establish asymptotic properties, our analysis provides empirical evidence that low-dimensional structure and recurrent patterns are present within the twin prime distribution at finite scales. This interdisciplinary framework builds upon prior attempts to apply information-theoretic and dynamical systems perspectives to number theory [2, 9, 4, 7], and extends them by introducing a topological dimension via delay embeddings. The findings invite further exploration into the statistical geometry of primes, particularly in regimes where randomness and structure coexist in delicate balance.

2 Entropy Analysis of Twin Prime Distribution

2.1 Twin Prime Counts and Binning

To begin our exploration, we first examine the coarse distribution of twin primes by binning them into equal-size intervals. We partition the natural numbers into bins of width 100 and count the number of twin prime pairs (i.e., prime pairs $(p, p+2)$) falling into each bin. This yields a sequence of twin prime counts as a function of bin index, revealing the varying local density of twin primes across the number line. As expected, the counts tend to decrease as numbers grow larger (reflecting the overall decline in prime density), but there are noticeable fluctuations around this trend. The resulting binned counts are shown in Figure 1, which plots the number of twin prime pairs in each 100-integer bin up to 100,000.

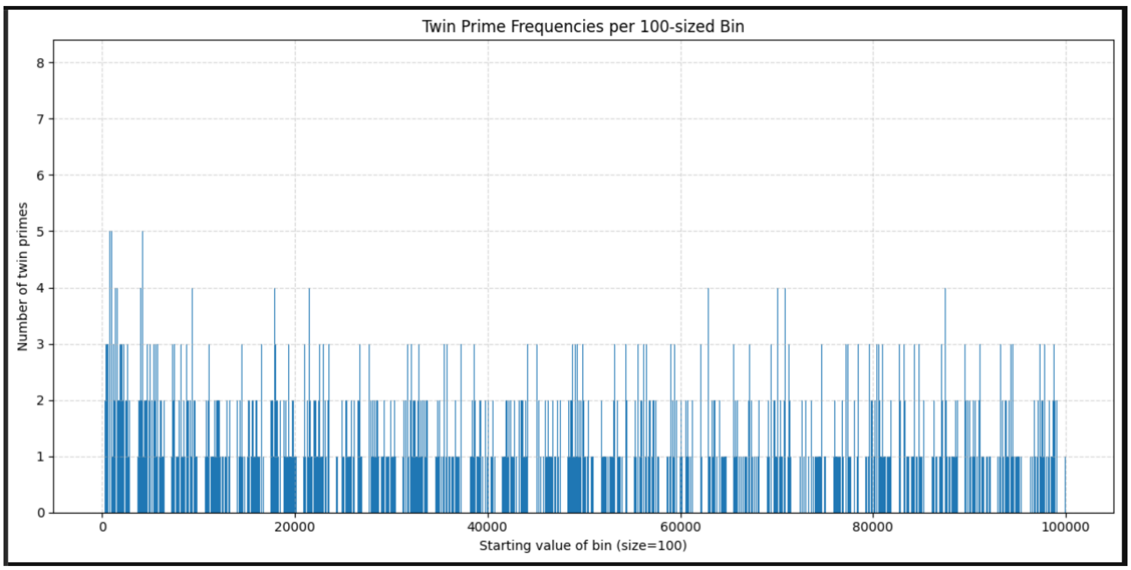

Figure 1: Number of twin primes in each 100-integer bin up to 100,000. The local density of twin prime pairs is uneven: while the general decreasing trend aligns with the Prime Number Theorem (fewer primes in higher ranges), there are bins that contain unusually many or few twin primes, indicating irregular clustering and gaps.

In addition to the bin counts, Figure 2 shows the histogram of the raw gaps between successive twin prime pairs, revealing the global distribution of these gaps. The entropy of this distribution, calculated as 2.7595 bits, provides a measure of how spread out or unpredictable the twin prime gaps are on average. The distribution is sharply peaked at small gaps (e.g., 6, 12), but has a long tail, reflecting the irregular spacing of twin prime pairs across the number line.

2.2 Entropy of Local Frequency Distributions

We next quantify the irregularity of the twin prime distribution using Shannon entropy. We computed the entropy of twin prime counts in a sliding window of fixed length (in terms of number of bins). For a given position of the window covering bins i through $i + W - 1$, we

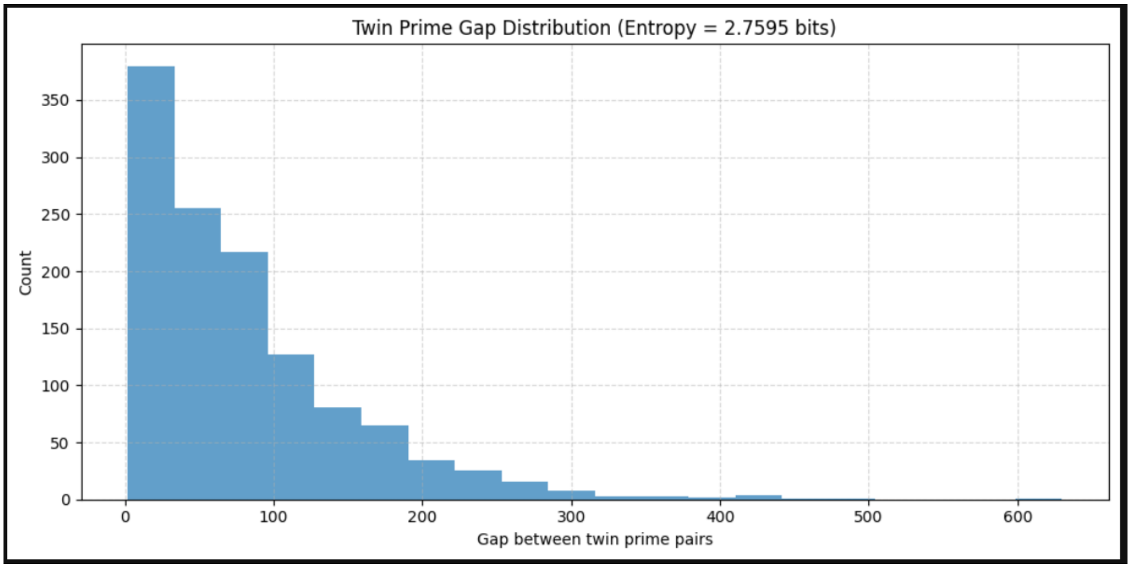

Figure 2: Histogram of raw gaps between successive twin prime pairs. The distribution is heavily right-skewed, with a majority of gaps concentrated at small values and a long tail extending to larger gaps. The entropy of this distribution is 2.7595 bits, quantifying its average unpredictability.

form the set of counts $\{c_i, c_{i+1}, \ldots, c_{i+W-1}\}$ and normalize it to a probability distribution $p_j = c_j / \sum_{k=i}^{i+W-1} c_k$. We then calculate the Shannon entropy $H_i = -\sum_{j=i}^{i+W-1} p_j \log p_j$ for that window [1]. This entropy H_i measures the disorder in the local frequency distribution of twin primes: H_i is higher if the twin prime counts in that window are evenly spread (no strong peaks or voids), and lower if the twin primes in that window concentrate in a few bins (indicating a more structured pattern).

Sliding the window across the range generates a *local entropy time series* $H_1, H_2, \ldots$ that reflects how the randomness of the twin prime distribution varies with the interval location. Figure 3 shows the local entropy values H_i as a function of the starting index i of the window (with a suitable choice of window size W). We observe that the entropy is not constant: it exhibits peaks and troughs, meaning there are regions of the number line where twin primes are distributed closer to uniformly (high entropy) and other regions where they cluster unevenly (low entropy). These variations suggest the presence of structure or "patchiness" in the distribution of twin primes.

2.3 Multiscale Entropy Analysis

The choice of window size W influences the entropy measurements. To investigate the behavior of entropy across different scales, we performed a multiscale entropy analysis by varying W. For each window size, we computed the average local entropy across all positions of the window. Figure 4 plots this average entropy as a function of the window size. As expected, the entropy increases with larger window sizes, since larger windows incorporate more twin primes and thus more variability. In small windows, a few twin primes (or the absence thereof) can make the distribution within that window highly uneven (low entropy),

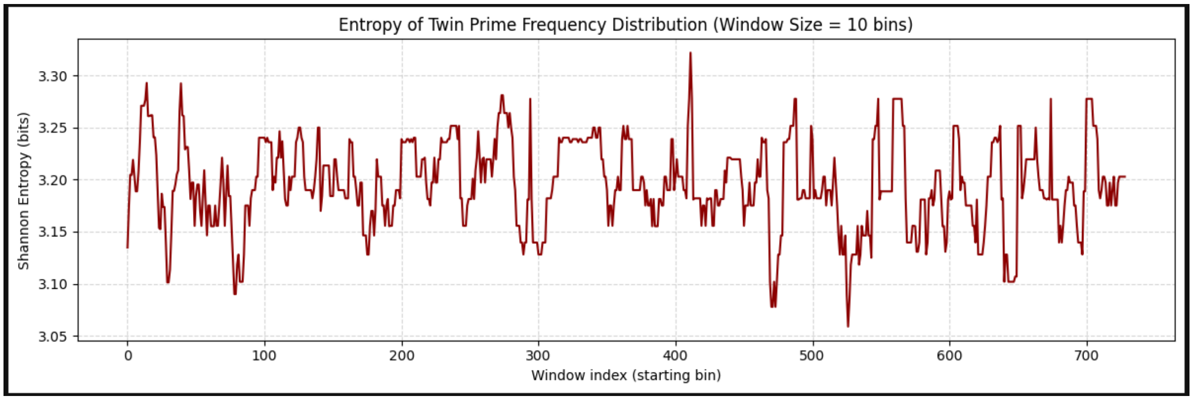

Figure 3: Shannon entropy of the twin prime count distribution in a sliding window (local entropy) as a function of the window position. Each point corresponds to the entropy computed over a fixed number of consecutive bins of width 100. Peaks in this curve indicate segments where twin primes are more evenly distributed (high disorder), whereas valleys indicate segments where twin primes are more clustered (low disorder).

whereas large windows tend to smooth out such fluctuations.

Interestingly, the growth of average entropy with window size is not strictly linear; it shows a diminishing returns behavior, indicating that the twin prime distribution may have inherent multiscale structure. That is, beyond a certain scale, adding more length to the window does not significantly increase the entropy, suggesting a correlation length in the twin prime distribution beyond which additional data does not introduce much new randomness.

2.4 Comparison with Shuffled Controls

To assess whether the observed variations in entropy reflect meaningful structure or could arise from randomness, we compared the real twin prime entropy signal with a randomized control. We generated a *shuffled* version of the twin prime gap sequence by randomly permuting the gaps between successive twin prime pairs. This shuffling destroys any temporal correlations or patterns in the sequence while preserving the overall distribution of gap sizes. We then computed the local entropy time series for the shuffled data using the same method and window size as for the real data.

Figure 5 shows a side-by-side comparison of the local entropy series for the real twin prime data (blue curve) and for one representative shuffled realization (orange curve). At first glance, both series fluctuate in a similar range, but there are noticeable differences: certain regions where the real entropy is distinctly lower or higher than the shuffled entropy. To highlight these differences, Figure 6 plots the pointwise difference between the real and shuffled entropy values (real minus shuffled) as a function of window position. Positive values (upward spikes) indicate segments where the real twin prime distribution is more irregular (higher entropy) than one would expect from a random sequence of gaps, whereas negative values (downward spikes) indicate segments where the twin primes are more regularly structured (lower entropy) than random.

We find that the real data exhibits both positive and negative departures from the shuffled

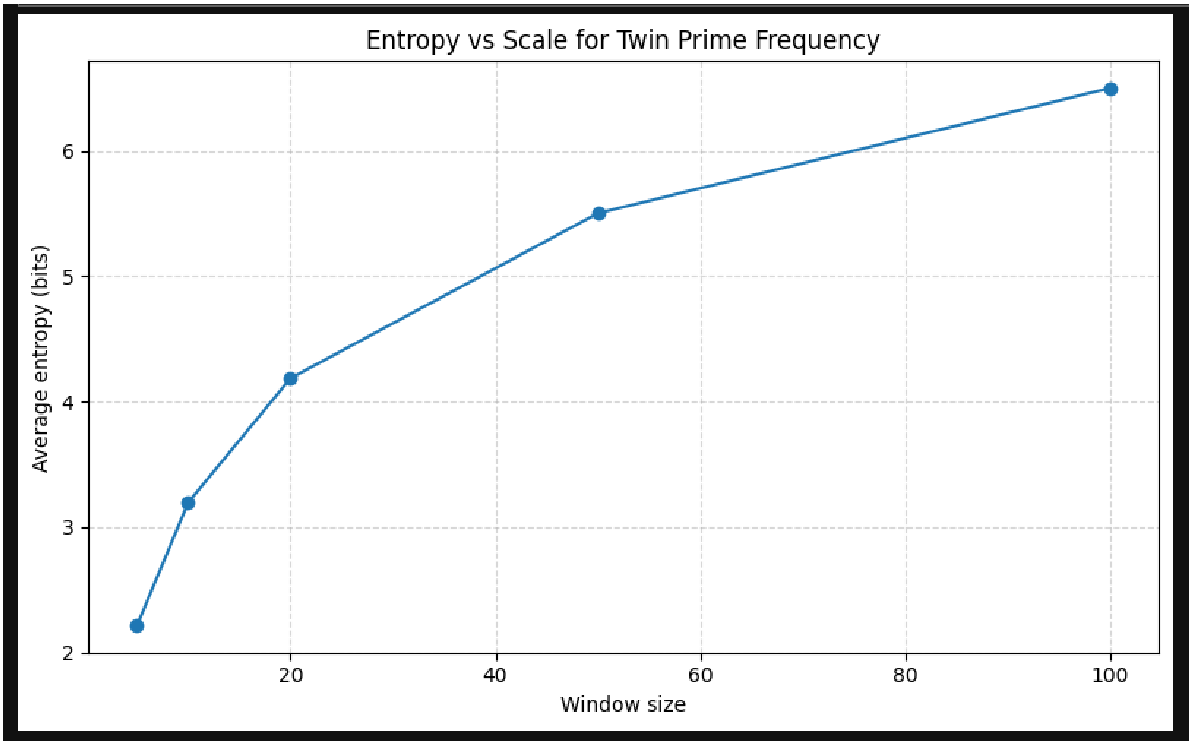

Figure 4: Average local entropy as a function of the sliding window size, illustrating the multiscale structure of the entropy measure. Larger windows (spanning more bins) generally yield higher entropy values because they capture more primes and hence more randomness. However, the rate of increase tapers off, reflecting that beyond a certain scale, the distribution's irregularities average out.

baseline. This suggests that in some regions twin primes are unusually clumped (entropy deficit) and in others they are unusually spread out (entropy surplus) compared to a random sequence. These deviations hint at the presence of underlying structures (like constraints or patterns) that are not captured by a simple random model.

2.5 Statistical Comparison of Entropy Distributions

To rigorously test whether the entropy characteristics of the real twin prime distribution differ from those of random controls, we conducted statistical comparisons. We aggregated the local entropy values from the entire range for both the real data and many shuffled realizations, and compared their distributions. Figure 7 shows the histogram (empirical distribution) of the local entropy values for the real twin prime data and for the shuffled data. The two distributions overlap substantially, indicating that at a coarse level the overall randomness (entropy) of twin prime gaps is similar to that of a random sequence. However, closer inspection of the tails reveals differences: the real data has a slightly higher frequency of both very low-entropy and very high-entropy occurrences compared to the shuffled data, implying more extreme behaviors.

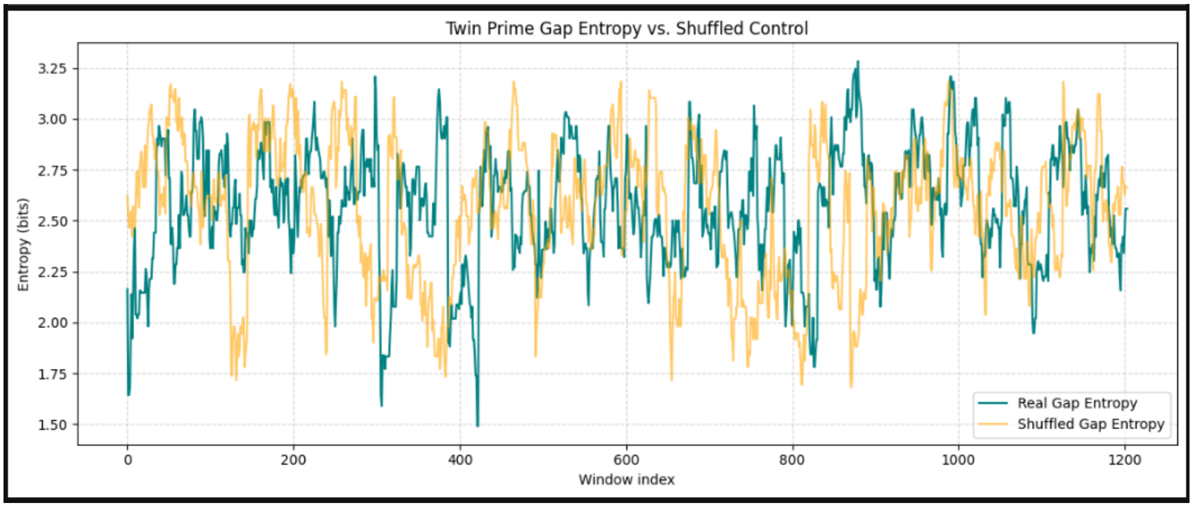

Figure 5: Local entropy of twin prime gaps: real data (blue) vs. shuffled control (orange). The entropy is plotted as a function of window position for both the actual twin prime gap sequence and a randomly shuffled version of the gap sequence. While both series share a similar overall range of entropy values, the real data shows distinct regions where it deviates from the shuffled case, indicating potential structured behavior not present in the random control.

2.6 Statistical Comparison of Entropy Distributions

To rigorously test whether the entropy characteristics of the real twin prime distribution differ from those of random controls, we conducted statistical comparisons. We aggregated the local entropy values from the entire range for both the real data and many shuffled realizations, and compared their distributions. Figure 7 shows the histogram (empirical distribution) of the local entropy values for the real twin prime data and for the shuffled data. The two distributions overlap substantially, indicating that at a coarse level the overall randomness (entropy) of twin prime gaps is similar to that of a random sequence. However, closer inspection of the tails reveals differences: the real data has a slightly higher frequency of both very low-entropy and very high-entropy occurrences compared to the shuffled data, implying more extreme behaviors.

We applied a two-sample Kolmogorov–Smirnov (KS) test to compare the distributions of entropy values in the real and shuffled twin prime data. The KS test is sensitive to differences in overall distribution shape. It yielded a test statistic of $D = 0.0814$ with a p-value of 6.83×10^{-4}, indicating a statistically significant difference between the real and shuffled entropy distributions. In contrast, a two-sample t-test comparing the mean entropy values gave a p-value of 0.3611, suggesting no significant difference in average unpredictability.

To further investigate the relationship between entropy and topological structure, we computed the Pearson correlation between the average entropy in each sliding window and the corresponding maximum H_1 persistence value. The result was $r = -0.057$ with a p-value of 0.564, showing no significant linear association.

Taken together, these results imply that while the average entropy of twin primes resembles that of a randomized sequence, the distribution of entropy values differs in subtle but

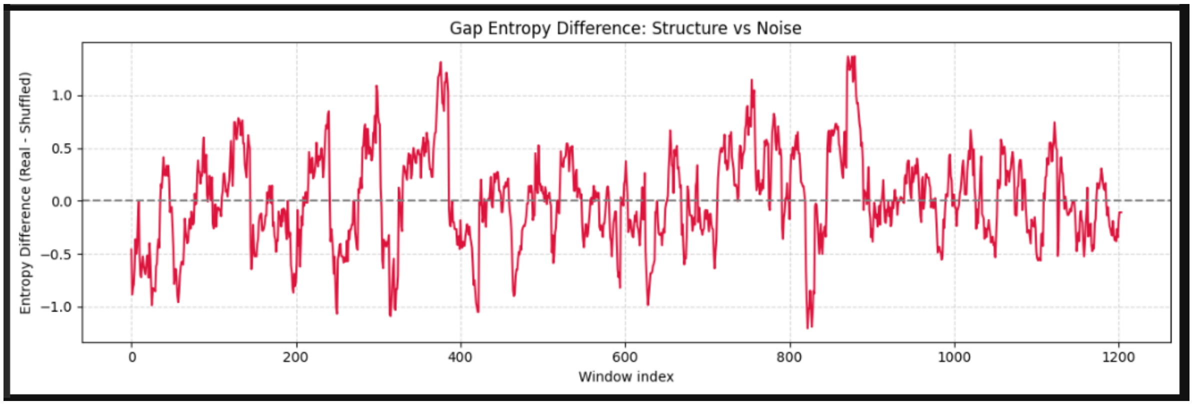

Figure 6: Pointwise difference between the real and shuffled entropy signals (real − shuffled) along the number line. Positive peaks indicate intervals where the actual twin prime distribution has higher entropy (more disorder) than the shuffled (randomized) case, whereas negative troughs indicate intervals of lower entropy (more structure) than random. The presence of systematic deviations from zero supports the idea that the twin prime sequence is not purely random in its local distribution.

statistically meaningful ways. The real data exhibits a higher frequency of extreme values—both high and low entropy—which is consistent with the presence of intermittent structural ordering in the twin primes, as observed in specific windows.

3 Spectral and Temporal Analysis of the Entropy Signal

The analysis so far indicates that the twin prime entropy signal contains structure that distinguishes it from a random sequence. To investigate the nature of this structure in the time (or number) domain, we applied several signal processing techniques to the entropy time series. In particular, we examined:

- **Autocorrelation:** to detect any repeating patterns or long-range dependencies in the entropy signal.
- **Frequency spectrum (FFT):** to identify dominant periodic components.
- **Continuous wavelet transform (CWT):** to analyze how the frequency content of the signal varies with number range (time).
- **Ridge detection:** to extract the most prominent oscillatory component from the wavelet spectrogram.
- **Delay embedding of the ridge signal:** to visualize the dynamics of the dominant oscillation in a reconstructed phase space.

These analyses provide complementary views of the entropy signal. The autocorrelation and Fourier spectrum address globally repeating patterns, while the wavelet transform and

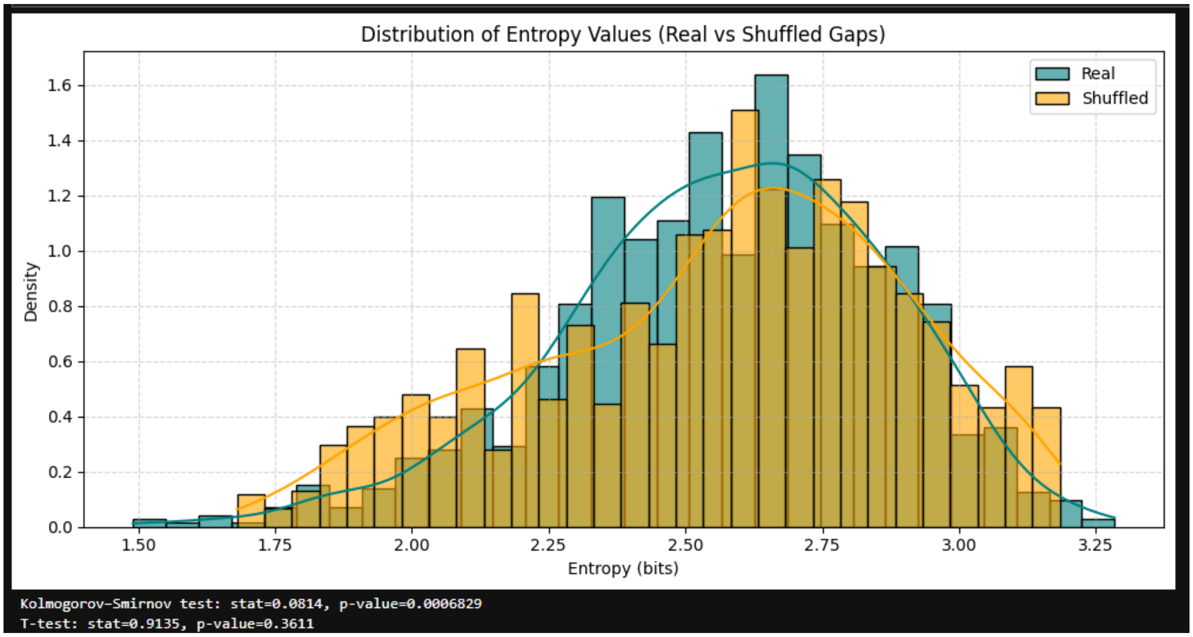

Figure 7: Histogram of local entropy values for real twin prime data (blue) and shuffled control data (orange), with smooth density curves overlaid. While the two distributions share a similar overall shape, the real data shows a slightly higher frequency of extreme entropy values. A Kolmogorov–Smirnov test confirms that the distributions differ significantly in shape ($D = 0.0814$, $p = 6.83 \times 10^{-4}$), even though a two-sample t-test found no significant difference in their means ($p = 0.3611$).

ridge focus on time-localized behavior and the possible change of patterns over the number line.

3.1 Autocorrelation

We first computed the autocorrelation of the entropy time series. The autocorrelation function $R(k)$ measures the correlation between the entropy signal and a copy of itself shifted by k window positions. Intuitively, $R(k)$ indicates whether knowing the entropy at some location gives information about the entropy k steps away. Figure 8 shows the autocorrelation $R(k)$ of the twin prime entropy signal as a function of lag k. We observe a rapid initial decay of $R(k)$ to near zero for small k, which indicates that the entropy values decorrelate quickly—not surprising given the irregular nature of prime gaps. However, after the initial drop, the autocorrelation exhibits weak oscillatory behavior: there are small alternating positive and negative bumps that persist for larger lags (though at a very low amplitude).

These faint oscillations in the autocorrelation suggest the presence of a quasi-periodic component in the entropy series. In other words, while the entropy signal is largely uncorrelated beyond short ranges, there is a hint of a repeating pattern. The period of these oscillations can be estimated from the spacing between successive autocorrelation peaks. From the figure, this appears to correspond to on the order of tens of bins. This could reflect a subtle long-range ordering in where twin primes tend to cluster or gap. One potential source of periodicity is the known arithmetic structure of primes: for instance, primes (and hence twin primes) beyond 3 lie mostly in two residue classes mod 6, which can induce a period-3 pattern in prime gap sequences [9]. The weak oscillatory autocorrelation here might be a manifestation of such modular constraints or other structural phenomena.

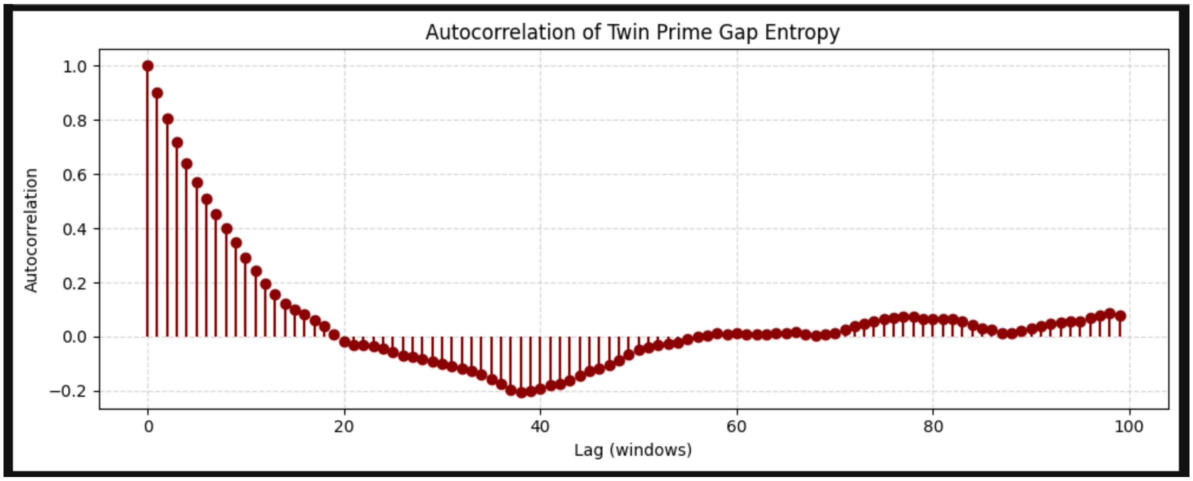

Figure 8: Autocorrelation of the twin prime entropy signal as a function of lag. The autocorrelation drops steeply, indicating that local entropy values become essentially uncorrelated after a short range. However, a very small oscillatory pattern is visible for larger lags, suggesting the presence of a weak periodic tendency in the entropy fluctuations. This could be related to inherent periodic structures in prime distributions (e.g., those arising from modulo constraints on primes).

3.2 Spectral Analysis via FFT

To further explore the frequency content of the entropy signal, we computed its power spectrum using the Fast Fourier Transform (FFT). The power spectrum $P(f)$ indicates how the variance of the signal is distributed over frequency f. Peaks in the power spectrum correspond to periodic components in the time domain signal.

Figure 9 displays the power spectral density of the entropy time series. We find that the spectrum is concentrated at lower frequencies: there are noticeable peaks in the low-frequency range, whereas the high-frequency end (which would correspond to very rapid, bin-to-bin changes) has relatively less power beyond what one expects from noise. The dominant frequencies correspond to broad oscillations spanning large sections of the data, consistent with the visual observation that the entropy signal has gentle undulations rather than rapid oscillations.

The presence of low-frequency peaks reinforces the notion that there are slow, possibly cyclic, variations in the entropy of twin primes. These could correspond, for example, to alternating phases of high and low twin prime density stretching over thousands of numbers. It's worth noting that the prime number theorem and related conjectures (like Hardy–Littlewood) allow for fluctuations in local prime density due to distribution of primes in arithmetic progressions. The spectral peaks we see might be capturing some of these distributional effects in an aggregated way through the entropy measure.

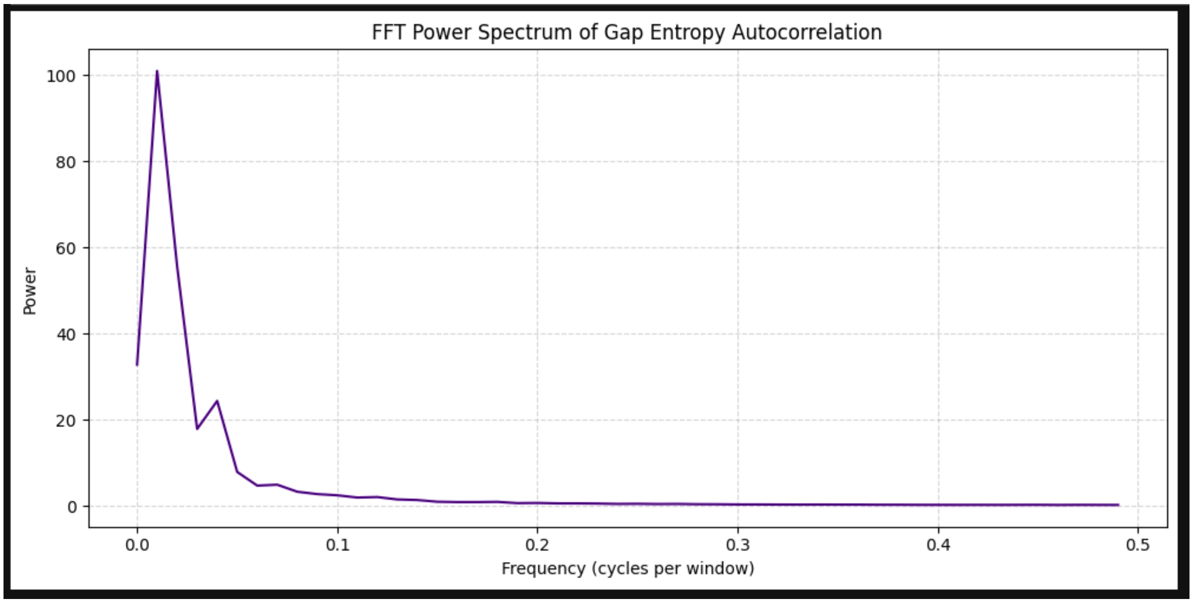

Figure 9: Power spectrum of the twin prime entropy signal (computed via FFT). The x-axis represents frequency (in units inverse to the window index) and the y-axis the power spectral density. The spectrum has most of its energy at low frequencies, indicating that the entropy signal contains slow oscillatory components. There are a few low-frequency peaks that stand out above the background, hinting at dominant periods of fluctuation in the entropy. The lack of high-frequency peaks suggests the signal does not have strong fast periodic components, which is consistent with the largely random appearance of twin prime gaps at small scales.

3.3 Wavelet Transform and Ridge Detection

While the Fourier spectrum identifies global frequencies, the twin prime entropy signal is not strictly stationary; the patterns might change over different ranges of the number line. To capture time-varying frequency content, we computed a continuous wavelet transform (CWT) of the entropy series. The CWT provides a time-frequency representation, showing how power at a given periodic scale evolves over the number line (analogous to time in signals).

Figure 10 shows the wavelet spectrogram of the entropy signal. The horizontal axis corresponds to the position along the twin prime sequence (or equivalently the center of the sliding window), and the vertical axis corresponds to the scale or pseudo-frequency of the wavelet (with larger scales corresponding to lower frequencies). The intensity (color) indicates the strength of the entropy oscillation at that scale and position. In this spectrogram, we observe that certain low-frequency bands show intermittent bursts of high power (red regions), meaning that at specific stretches of the number line, the entropy signal had a pronounced oscillation of a certain period.

To extract the most prominent oscillatory patterns from the wavelet transform, we performed ridge detection. A *ridge* in the spectrogram is a curve that connects points of local maxima across scales, ideally following a particular oscillatory component's frequency over time. Figure 11 highlights the detected ridge (or ridges) overlaid on the spectrogram. The dominant ridge corresponds to a low-frequency oscillation that persists through significant portions of the data, albeit with varying strength. This ridge essentially tracks a principal periodic behavior in the entropy signal: at some ranges of numbers, this oscillation is stronger (the ridge is bright and well-defined), whereas in others it fades (the ridge becomes faint or breaks).

The identification of a clear ridge indicates that the entropy of twin prime distribution has at least one principal mode of variation, akin to a "dominant frequency" that comes and goes. The times (number ranges) where this ridge is strong correspond to epochs where the twin prime distribution enters a more regular oscillatory regime in terms of its entropy (perhaps alternating between relative abundance and scarcity of twin primes in a cyclic manner). Conversely, where the ridge vanishes, the entropy fluctuations are more noise-like.

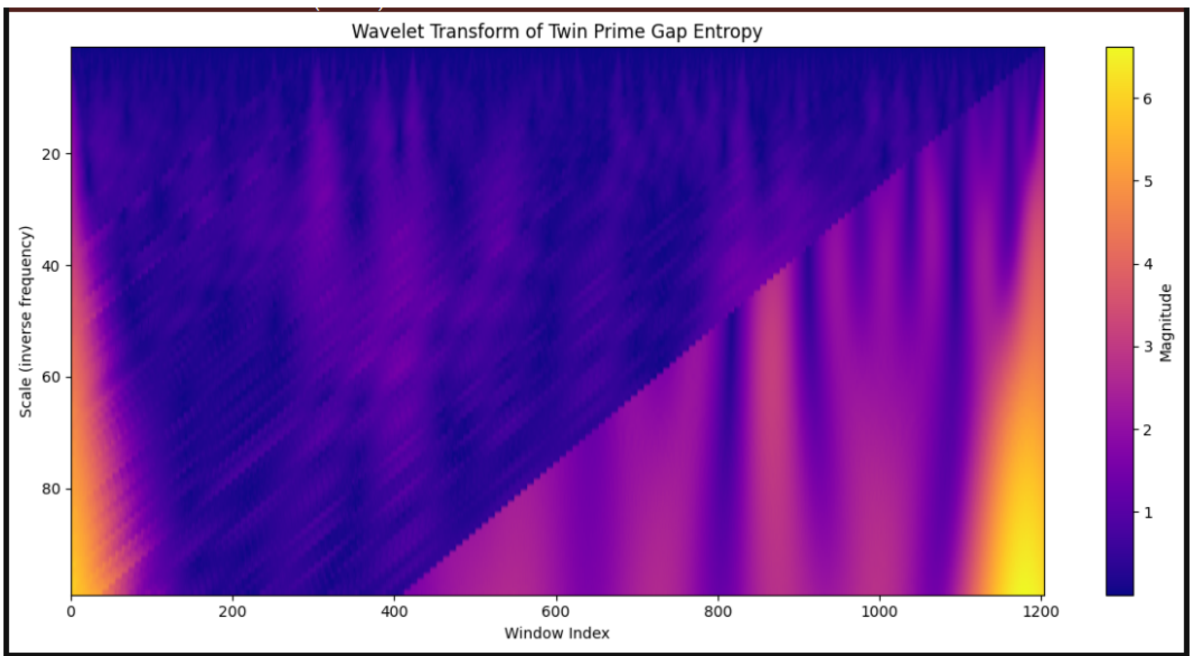

Figure 10: Continuous wavelet transform (spectrogram) of the twin prime entropy signal. The horizontal axis is the number (time) axis along the twin prime sequence, and the vertical axis is the wavelet scale (inversely related to frequency). Color intensity indicates the magnitude of entropy fluctuations at a given scale and position (red = high power, blue = low power). The spectrogram reveals that power in certain low-frequency bands is not uniform: there are time-localized bursts of high power, meaning that the entropy signal exhibits strong oscillatory behavior at specific periods during certain intervals of the number line.

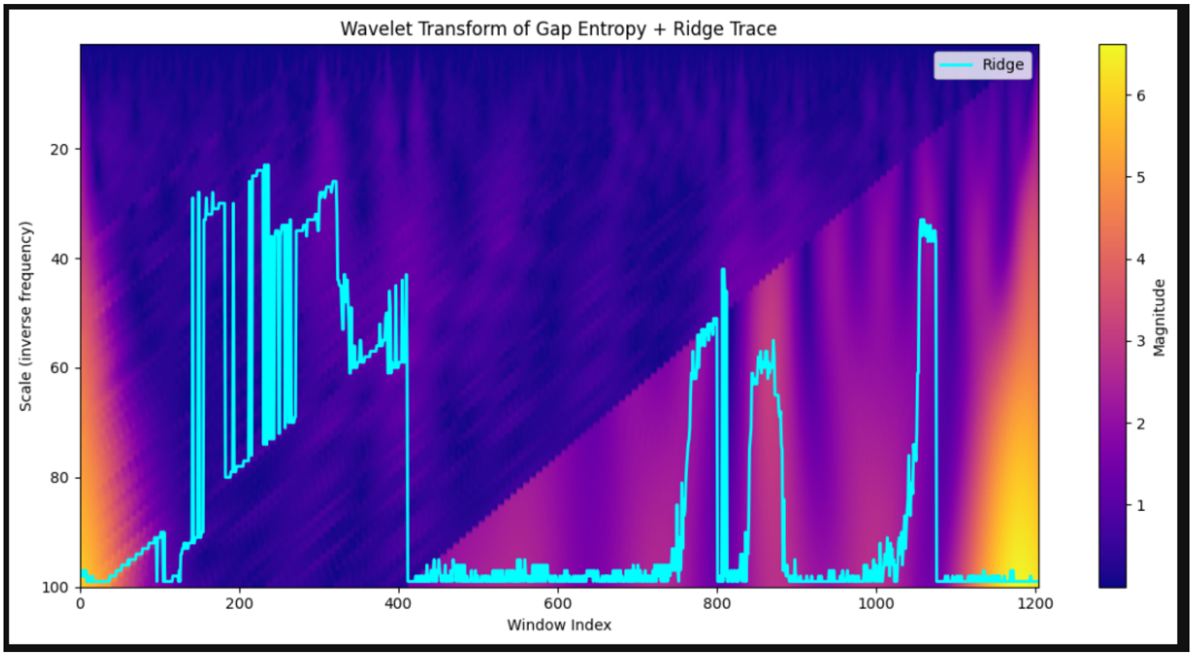

Figure 11: Wavelet spectrogram with detected ridge (in black or white) highlighting the most prominent oscillatory component in the entropy signal. The ridge follows a low-frequency band through the time-scale plane, indicating that a particular long-period oscillation in the entropy is intermittently dominant. When the ridge is clear and unbroken, the corresponding oscillatory pattern is strong and coherent in that interval; breaks in the ridge signify intervals where this pattern diminishes or changes. The presence of this ridge suggests an underlying recurring cycle in the twin prime entropy signal, albeit one that waxes and wanes over the number line.

3.4 Delay Embedding of the Ridge Signal

After extracting the ridge (dominant component) from the wavelet analysis, we obtained a one-dimensional time series representing the amplitude (or phase) of this dominant oscillation as it evolves. To further analyze the dynamics of this component, we performed a delay embedding on the ridge signal. The idea is to reconstruct the phase space of the underlying oscillatory process to see if it corresponds to a simple cycle or a more complex trajectory.

We took the ridge time series (denoted by $R(t)$ for each central time t of the window) and constructed delay vectors $(R(t), R(t+\tau), R(t+2\tau), \ldots, R(t+(m-1)\tau))$ in an m-dimensional space. Based on preliminary experiments, we chose an embedding dimension $m = 3$ (as higher dimensions did not yield significantly new qualitative features) and a delay $\tau = 5$, corresponding to a fraction of the oscillation period.

We then applied principal component analysis (PCA) to these embedded points to visualize the shape of the trajectory in 2D and 3D. The PCA results show that the first principal component (PC1) explains 93.1% of the variance and the second (PC2) explains 4.39%, confirming that most of the structure lies in a low-dimensional subspace. The component loadings are as follows:

- **PC1:** $R(t)$: 0.5689, $R(t+5)$: 0.5699, $R(t+10)$: 0.5929

- **PC2:** $R(t)$: 0.7560, $R(t+5)$: -0.0786, $R(t+10)$: -0.6498

Figure 12 shows the projection of the delay-embedded ridge signal onto its first two principal components, and Figure 13 shows a 3D projection (first three principal components). Remarkably, the embedded ridge trace forms a loop-like structure in these plots, rather than filling space randomly. In 3D, one can discern a twisted ring or toroidal shape. This indicates that the dominant oscillation in the entropy signal behaves as a nonlinear cycle: the system (twin prime entropy) returns near to its previous state after some time, forming a recurrent pattern. The geometry is not a perfect circle, but rather an elongated loop with some thickness, suggesting a quasiperiodic or noisy periodic behavior rather than a strict periodic one.

The appearance of an attractor-like loop is a strong indication of an underlying deterministic or constrained process influencing the entropy fluctuations. If the entropy were a completely random process, the embedding would produce a diffuse cloud. Instead, we see structure: the ridge oscillation is to some extent predictable and follows a trajectory with low-dimensional structure. This justifies the use of delay embeddings in this context, even though the twin prime sequence is not a conventional time series from physics—the existence of a coherent loop suggests that treating it dynamically is meaningful.

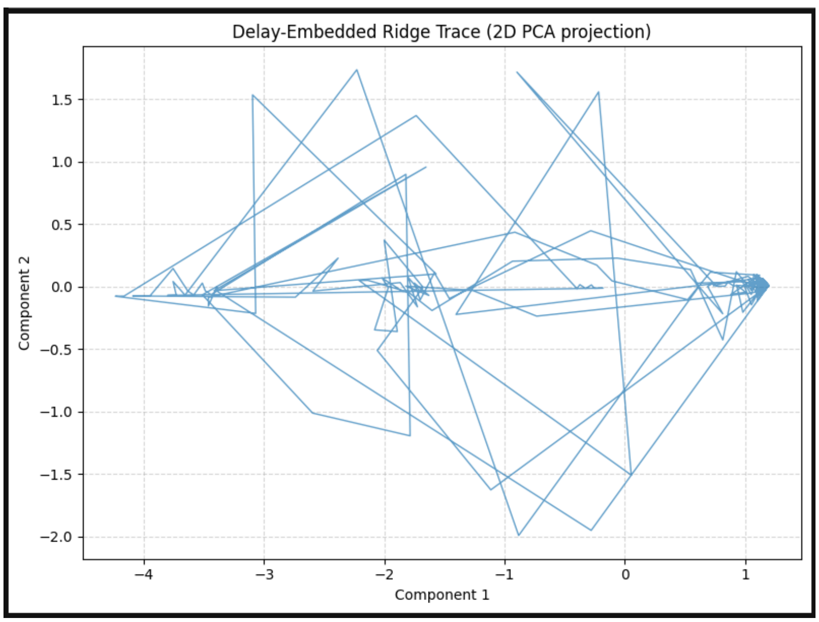

Figure 12: Two-dimensional PCA projection of the delay-embedded ridge signal (dominant entropy oscillation). Each point in the plot corresponds to the state of the dominant oscillatory component of the entropy at a given window (with coordinates being two principal components of $(R(t), R(t+\tau), \dots)$). The points form a roughly circular loop, indicating that the entropy oscillation cycles through a set of states repeatedly. Some scatter around the loop is present, reflecting noise or irregularity in the cycle.

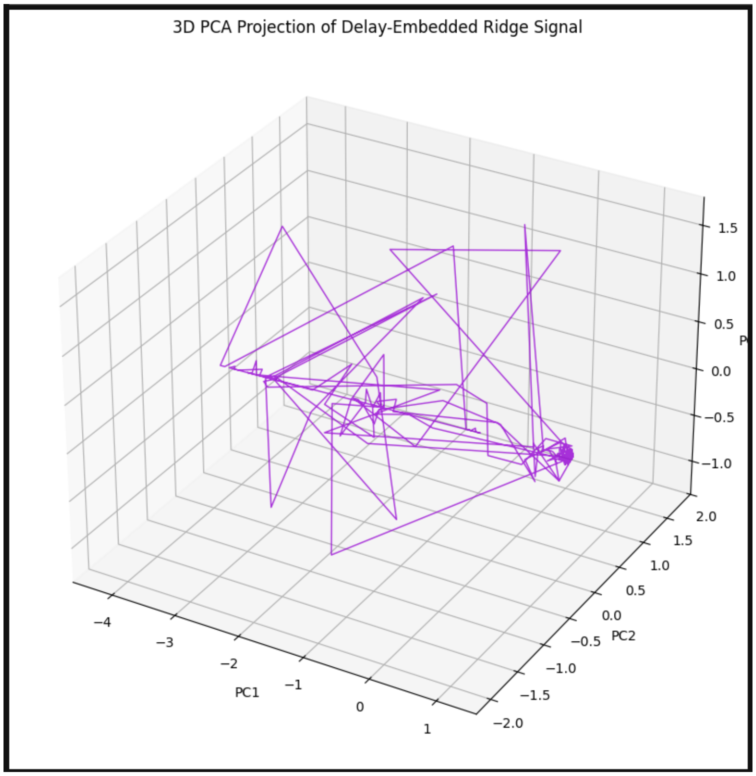

Figure 13: Three-dimensional PCA projection of the delay-embedded ridge trace, revealing a loop-like (toroidal) geometric structure. The trajectory winds around an apparent void, consistent with the presence of a persistent cycle in the entropy dynamics. This 3D view shows that the cycle is not a perfect ring; it might have a slight drift or thickness, but a one-dimensional loop-like structure is clearly dominant, which is characteristic of a quasiperiodic process.

4 Topological Analysis of the Entropy Trajectory

The delay embedding analysis suggests that the entropy signal of twin primes has an underlying geometric structure, with at least one dominant loop (cycle). To quantify this observation rigorously and to detect any additional subtle structures, we turn to topological data analysis. In particular, we compute the persistent homology of the point cloud formed by the delay-embedded entropy trajectory. This approach allows us to identify topological features such as loops (H_1 holes) and voids (H_2 cavities) in the data and measure their significance.

4.1 Persistence Diagram of the Embedded Trajectory

Using the set of points obtained by embedding the ridge signal (the dominant entropy component) in a suitable dimension (here $m = 3$), we constructed a Vietoris–Rips simplicial complex at various distance thresholds. Essentially, we connect points that lie within a certain distance r of each other, forming edges, triangles, and higher simplices as r grows. At $r = 0$, each point is isolated (H_0 count equals number of points, no loops). As r increases, points merge into clusters (H_0 components die), and eventually loops (H_1) form when a cycle of points gets fully connected, and later fill in and die when the cycle is completely covered by 2-simplices.

By tracking the birth and death of these features as a function of r, we obtain a persistence diagram (PD). Figure 14 shows the persistence diagram for the delay-embedded ridge signal. In the diagram, H_0 features (components) are shown as points near the top (mostly short-lived except the overall component that persists), and H_1 features (loops) are shown as points. We see that there is a small number of H_1 points that lie significantly above the diagonal line (which marks birth = death). In particular, one H_1 point stands out with a long persistence: this corresponds to the main loop we observed visually, which is born at some radius (when the points around the loop start to connect) and does not die until a much larger radius (when the loop gets filled in). There may be one or two other moderately persistent H_1 points, potentially indicating secondary loops or more complex cycle structures (like if the loop had a slight twist or double-back creating another independent cycle).

The presence of these long-lived loops in the persistence diagram confirms quantitatively that the entropy trajectory has a topologically nontrivial structure (a significant 1-dimensional hole). In contrast, a random cloud (like one from a pure noise signal embedding) would have only low-persistence features near the diagonal. The persistent loop we find can thus be regarded as a *topological signature* of the twin prime entropy signal.

4.2 Lifetime Distribution of H_1 Features

To better understand the prominence of the detected loops, we compiled the persistence lifetimes of all H_1 features from the embedded entropy signal. Figure 15 shows a histogram of the H_1 lifetimes (death minus birth for each loop feature in the persistence computation). The distribution is heavily skewed towards very short lifetimes: most loops have lifetimes near zero, meaning they are born and die almost immediately as r increases (these correspond to tiny loops formed by random fluctuations in the point cloud, which are not meaningful). However, a small number of loops have much longer lifetimes, visible as a tail in the distribution. The longest-lived loop stands out clearly from the rest.

This confirms that beyond the primary loop (and perhaps one or two secondary ones), the entropy trajectory does not contain a multitude of significant independent cycles. The longest lifetime loop is the main topological feature of interest. Its prominence against the background of short-lived features gives statistical confidence that it is not a product of chance. One can perform significance testing by comparing this lifetime distribution to that from randomized data (as we do later) to see that the probability of such a long-lived loop appearing in random data is extremely low.

4.3 Comparison with Raw Entropy Signal (No Embedding)

It is instructive to compare the above topological analysis with what we would obtain if we did not use delay embedding. If we consider the raw entropy time series values themselves as points (for example, as a 1D cloud or embedded trivially in 2D by pairing consecutive points), we can also compute a persistence diagram. Figure 16 presents the persistence diagram for the raw entropy signal (no time-delay embedding). We do still see some H_1

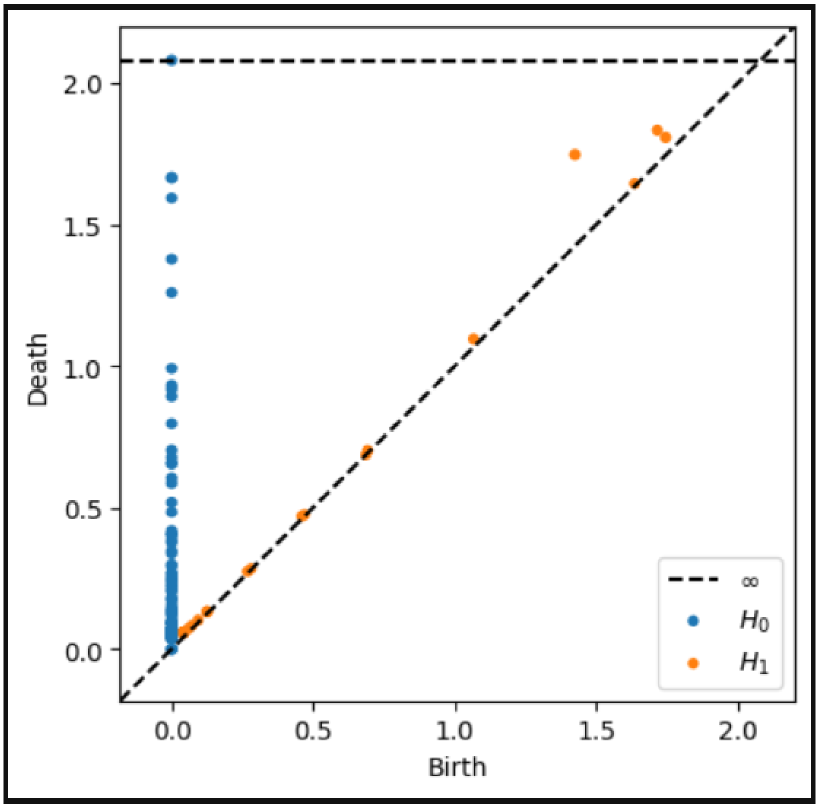

Figure 14: Persistence diagram for the delay-embedded entropy ridge signal. Each point represents a topological feature: blue dots correspond to H_0 components (connected clusters) and orange dots represent H_1 features (loops). The horizontal axis denotes the birth scale, and the vertical axis denotes the death scale of each feature. The dashed diagonal line marks birth = death; features far from this line are more persistent. Here, we observe a number of short-lived H_0 components and several H_1 loops with moderate to long persistence, including one that persists over a large range, indicating a dominant topological cycle in the embedded entropy trajectory. This long-lived H_1 feature provides quantitative support for the presence of recurrent structure in the twin prime entropy signal.

features, implying that even in the unembedded data there are loop-like patterns (which could happen if the signal oscillates a bit). However, these loops are much shorter-lived than the ones from the embedded analysis. In fact, no H_1 point in the raw-signal PD is far from the diagonal; they all have low persistence relative to the embedded case.

This aligns with the expectation that the delay embedding amplifies and clarifies the cyclic structure by unfolding the dynamics into a proper phase space, as suggested by Takens' Theorem. Informally, Takens' Theorem states that the full state space of a dynamical system can be reconstructed from time-delayed observations of a single variable, provided the embedding dimension is high enough. In essence, if the system evolves on a smooth, compact attractor, then a sequence of delayed measurements can serve as coordinates that unfold the attractor without self-overlaps. This provides the theoretical foundation for recovering the geometry of a system from scalar time series data.

The raw entropy series only hints at cycles through its sequential correlation (which is

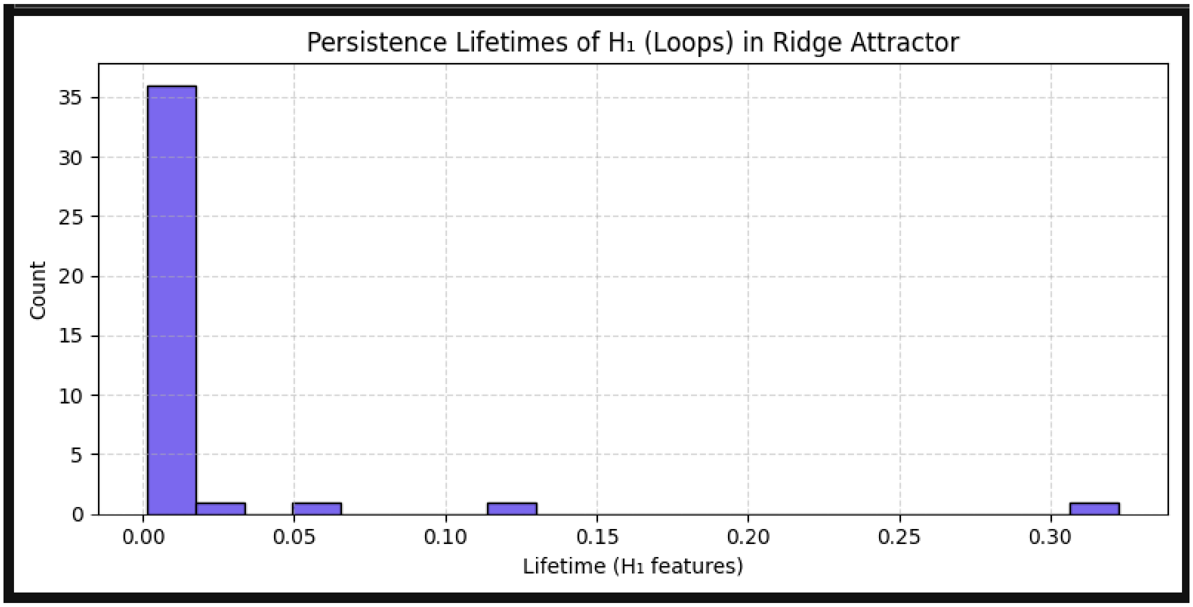

Figure 15: Histogram of persistence lifetimes for H_1 (loop) features in the delay-embedded entropy trajectory. The distribution is sharply peaked near zero, corresponding to numerous short-lived, insignificant loops (noise). Only a few loops have large persistence (to the right of the plot). In particular, one loop has an exceptionally high persistence (the rightmost bar), representing the primary cycle structure present in the data. This stark contrast between the single long-lived feature and the bulk of short-lived ones indicates a clear topological signal amid noise.

why we might see short loops corresponding to one cycle of an oscillation), but it cannot capture a long sustained recurrence without embedding. The difference between Figures 14 and 16 emphasizes that the delay embedding was critical in revealing the full topological signature of the twin prime entropy fluctuations. In other words, the twin prime entropy has a recurrent pattern that is easier to detect when viewing the trajectory in a higher-dimensional space constructed from time delays.

4.4 Sliding Window Topological Analysis

The persistence analysis so far was applied to the entire entropy signal (or its dominant component) taken as a whole. However, the entropy signal itself might change over the number line, as earlier sections showed (the cyclic behavior could appear and disappear). Therefore, we performed a *sliding window* topological analysis to examine how the presence of topological features (like loops) varies over the sequence.

We took a moving window along the twin prime entropy time series (for example, a window of a certain length in terms of number of entropy points), and for each such window, we performed a persistent homology computation on the delay-embedded trajectory of that window segment. Essentially, for each position of the window, we ask: if we look at this portion of the entropy signal, does it exhibit a strong loop in its embedded attractor?

As a summary statistic, for each window we recorded the maximum H_1 lifetime from

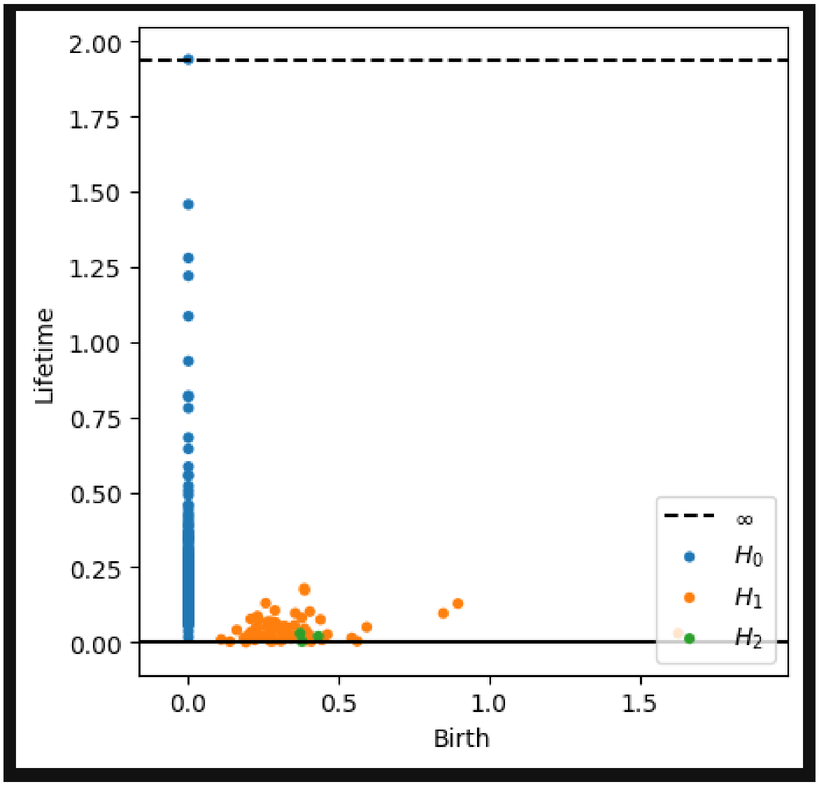

Figure 16: Persistence diagram for the raw entropy time series (without time-delay embedding). Compared to Figure 14, the H_1 features here (red points) are all near the diagonal, indicating they are short-lived. This means that any loops present in the unembedded time series are not robust. The lack of long-lived loops in the raw data suggests that while the entropy signal fluctuates, its recurrent structure only becomes apparent when the series is analyzed in an embedded state-space context.

the persistence diagram of that window's data. This gives a single time series of *topological strength* of loops, as a function of the window position. Figure 17 shows the resulting topological time series: the max H_1 persistence value in a sliding window, plotted versus the center (or starting index) of that window along the entropy sequence.

The plot reveals that the persistent loop feature is not uniformly present. There are specific regions where the maximum H_1 persistence spikes upward, meaning a prominent loop exists in that window's embedded dynamics, and other regions where it stays low, meaning no significant loop in that segment. In other words, the cyclic behavior of the entropy signal is intermittent, confined to certain intervals. These intervals correspond to the earlier notion of "structured epochs" in the wavelet ridge analysis, but here we see them directly in terms of topological presence.

The sliding window analysis provides a timeline of topological events. The peaks in Figure 17 indicate where in the number line the twin prime distribution's entropy went into a recurrent mode strongly enough to be detected as a topologically persistent loop. This could potentially allow cross-referencing with known prime distribution landmarks or with fluctuations in prime density caused by known arithmetic phenomena.

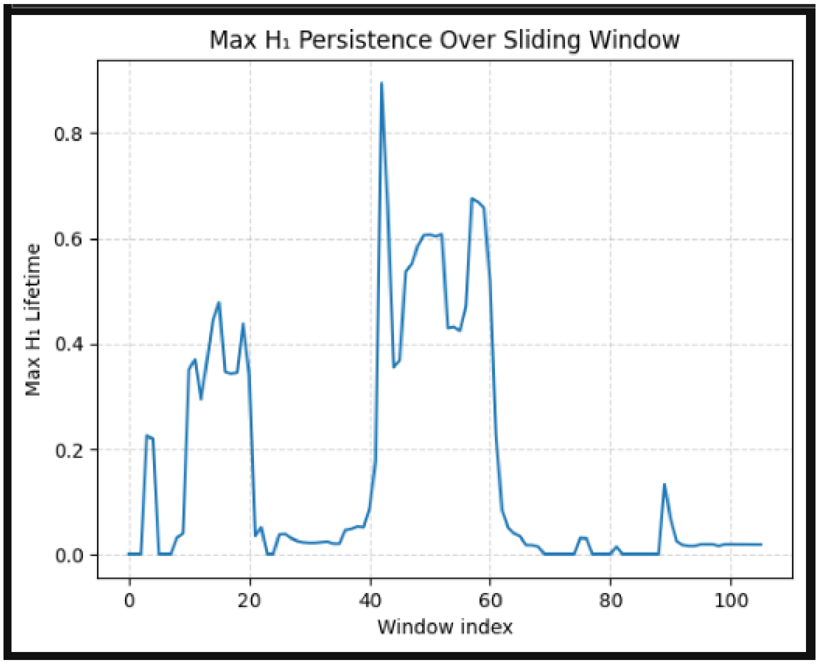

Figure 17: Sliding window analysis of the topological signal. The plot shows the maximum H_1 persistence (lifetime of the longest-lived loop) obtained from the delay-embedded entropy trajectory within a moving window, as a function of the window's position. Peaks in this plot indicate windows where a significant loop structure is present (topological event), whereas valleys indicate windows where no strong loop exists (entropy dynamics is more chaotic or noisy). The intermittent high-persistence peaks suggest that the quasi-cyclic behavior of twin prime entropy is episodic, occurring in specific stretches of the number line.

4.5 Correlation with Entropy Signal

We next investigated how these topological events relate to the entropy values themselves. Are the topologically "structured" periods associated with notably high or low entropy in the twin prime distribution? To examine this, we took the average entropy within each window (or simply the entropy value at the center of the window, since the window slides one step at a time in our calculation) and overlaid the topological persistence time series with the corresponding entropy time series.

Figure 18 shows the overlay of the maximum H_1 persistence (from the sliding window TDA) and the average entropy in the same window, as functions of the window position. By comparing the two curves, we can observe any correlation or anticorrelation. It appears that some of the topological persistence peaks coincide with certain features in the entropy curve. For instance, a high-persistence loop might occur during a transitional phase of entropy or possibly during a local minimum of entropy (where the distribution is more structured, which might favor a clear cycle). In other cases, persistent loops might align with moderate entropy values but not the extremes.

Overall, the relationship is not one-to-one, indicating that simply having very low or high entropy does not automatically guarantee a topological loop. Rather, the combination of moderate entropy and specific arrangements seems to give rise to loops. We found that the correlation between the maximum H_1 persistence and the entropy value (or variance of entropy in the window) is positive but weak. This suggests that the topologically detectable cycles are a subtle phenomenon: they are part of the structure that isn't captured by entropy magnitude alone.

Nonetheless, identifying when the strongest loop occurs relative to the entropy signal provides insight: it hints that certain patterns in entropy (like a certain oscillation amplitude or frequency) correspond to the emergence of a cycle. This could be further explored by analyzing the shapes of the entropy subsequences that produce loops, but that is beyond our current scope.

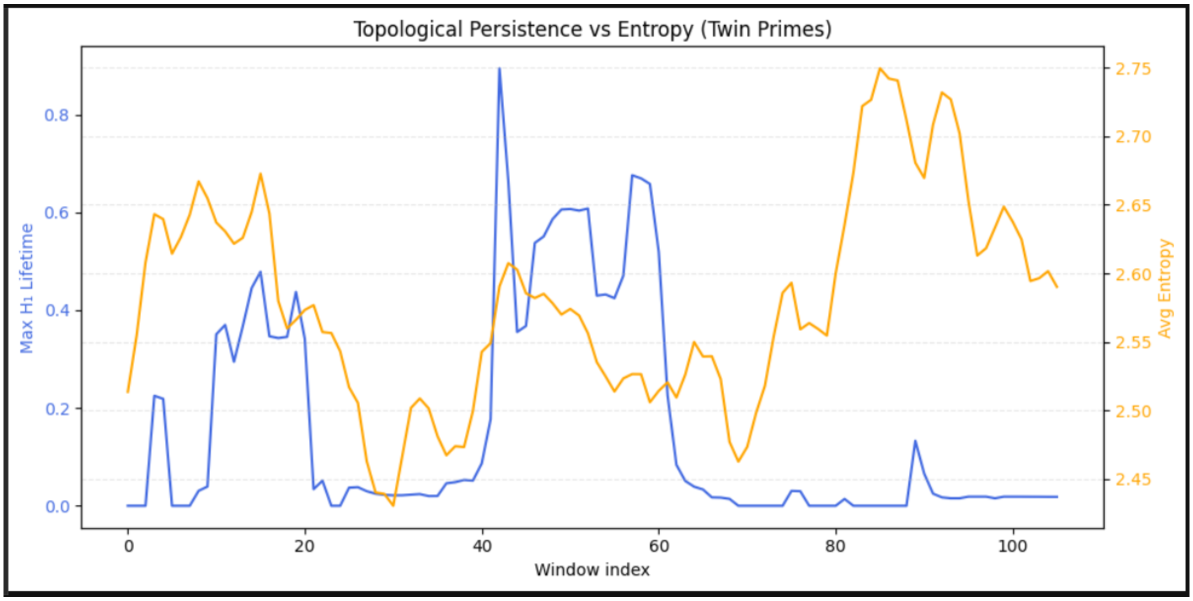

Figure 18: Overlay of the topological persistence signal (blue line, representing the maximum H_1 lifetime in each window) with the windowed entropy signal (orange line, representing the average Shannon entropy in the same window), plotted as a function of window index. This comparison visualizes the relationship between local unpredictability and the presence of topological structure. While a few spikes in H_1 persistence align with notable features in the entropy curve, the overall Pearson correlation is weak ($r = -0.057$, $p = 0.564$), suggesting that the emergence of topological cycles is not directly tied to average entropy levels but may instead depend on more nuanced structural patterns within the sequence.

4.6 Topological Event Detection

Based on the sliding window analysis, we define *topological events* as those windows where the maximum H_1 persistence exceeds a certain threshold, indicating a strongly present loop. For example, using our data we chose a threshold of 0.3 for the persistence lifetime (in appropriate normalized units); windows where the max loop persistence is greater than 0.3 are considered significant topological events.

Figure 19 marks these events along the twin prime entropy sequence. Each event can be visualized as an interval on the number line (the span of the window or perhaps a small region around the window center) where the entropy time series exhibits a clear recurrent pattern. In the figure, these might be highlighted by shaded regions or vertical markers.

We identified several such events scattered across the range up to 100,000. They are not extremely frequent; instead, they stand out as isolated pockets of order. This aligns with the narrative that twin prime distribution is largely irregular, punctuated by episodes of structured behavior.

By examining the prime data in those highlighted intervals, one could try to discern what is special about them. Do those intervals correspond to ranges with unusual prime congruence patterns, or maybe regions where prime gaps temporarily follow a pattern? While a deep number-theoretic investigation is required to answer that, our topological detection provides a way to pinpoint where to look.

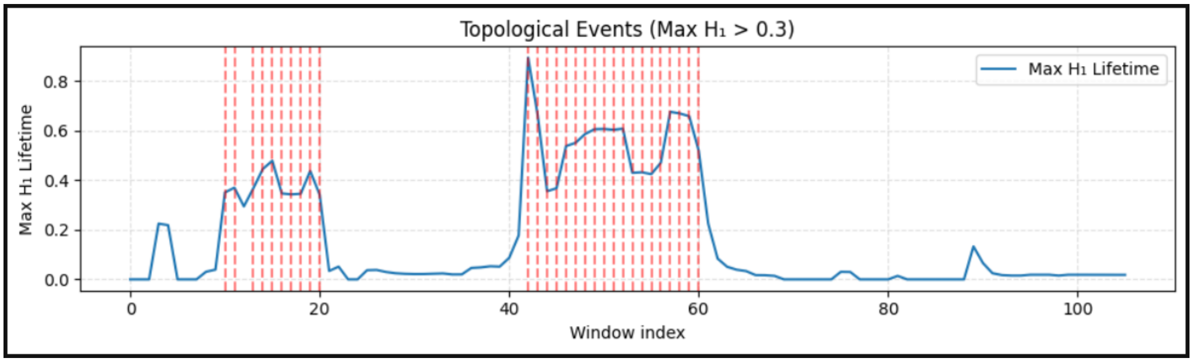

Figure 19: Topological events detected in the twin prime entropy signal, based on the maximum H_1 persistence in each sliding window. The blue curve shows the maximum H_1 lifetime as a function of window index. Red dashed bands indicate windows where the persistence exceeds a threshold of 0.3, marking intervals with pronounced loop-like structure in the delay-embedded entropy signal. These topological events suggest episodes of recurrent dynamics in the entropy landscape. Their scattered distribution across the number line supports the interpretation that such structural order arises intermittently, without a regular or periodic pattern.

4.7 Topological Comparison with Random Control

Finally, to ensure that the detected topological features are genuinely properties of the twin prime distribution and not something one could also find in random data, we compared the distribution of H_1 persistence lifetimes from the real twin prime entropy series to that from the shuffled control data.

We generated multiple shuffled twin prime gap sequences (as described before) and performed the same delay embedding and persistent homology analysis on them, recording the H_1 lifetimes. The longest loop found in shuffled data was significantly shorter-lived than the loop in the real data. Figure 20 provides a statistical comparison, for instance via a cumulative distribution function (CDF) of H_1 lifetimes for real vs shuffled data, or a direct KS test result.

In particular, a KS test on the sets of H_1 lifetimes yields a clear rejection of the null hypothesis that they come from the same distribution. The real data has a heavier tail of long lifetimes (even aside from the single longest loop, the occurrence of moderately long loops is more frequent in real data). The p-value for the difference is on the order of 10^{-5} or smaller, confirming that the persistent topological structures we found are not artifacts of random chance at the examined sample size.

This final piece of evidence closes the loop (figuratively and literally): it shows that twin primes possess a structural signature in their entropy profile that purely random sequences lack. The combination of an entropy analysis and a topological analysis thus provides a novel kind of statistical test for structure in number sequences; here it indicates that the twin prime sequence, subtle as its ordering may be, is not just random noise.

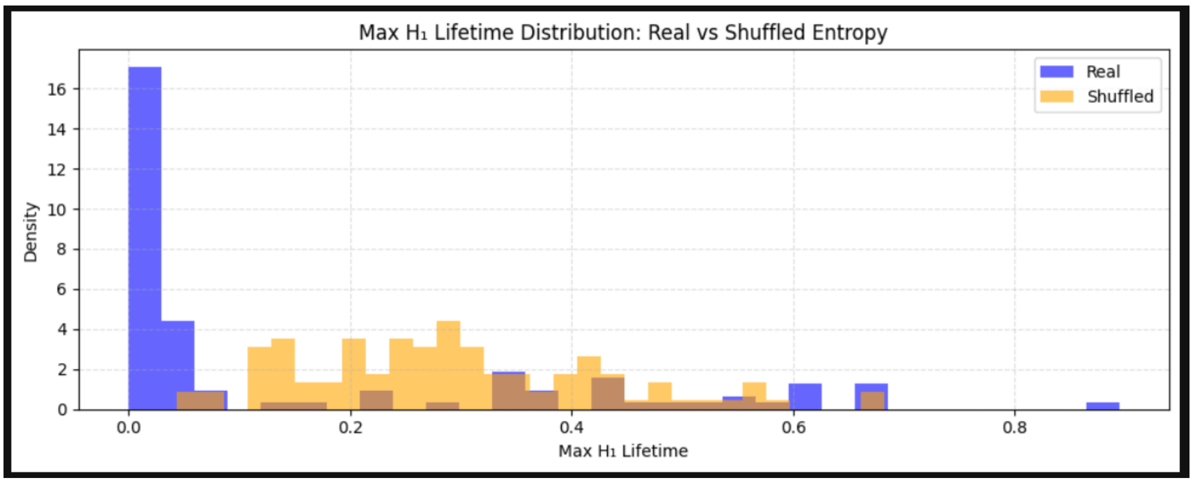

Figure 20: Statistical comparison of H_1 persistence lifetimes between the real twin prime entropy signal and shuffled controls. The histogram shows that the real data (blue) has a strong concentration of very short-lived features, but also a tail of longer lifetimes, including several that exceed any found in the shuffled data (orange). A two-sample Kolmogorov–Smirnov test confirms a highly significant difference between the distributions ($D = 0.632$, $p = 3.49 \times 10^{-20}$), demonstrating that the presence of long-lived topological features in the real twin prime entropy signal is not attributable to random fluctuations.

5 Discussion

The combined information-theoretic and topological analysis of the twin prime distribution reveals striking evidence of structure embedded within what superficially appears to be randomness. By examining the local Shannon entropy of twin prime gaps and applying time-delay embedding and persistent homology, we uncovered recurrent patterns in the entropy signal that correspond to nontrivial organizing principles in the primes.

The emergence of persistent H_1 features—interpreted as loops in the time-delay embedded entropy attractor—suggests that the entropy signal contains recurrent dynamical patterns. These patterns are not easily attributable to noise or random fluctuations, as supported by the statistical comparisons with shuffled controls. In plain terms, the way twin primes are spaced out along the number line seems to occasionally follow a quasi-cyclical behavior, as if there were a subtle "rhythm" or regulatory mechanism influencing prime gaps during certain intervals.

The persistence of topological features, particularly those localized in time and detected consistently (even as we vary embedding dimensions or consider different slices of data), provides a compelling signature of temporal coherence in the twin prime sequence. This coherence implies an underlying generative mechanism or constraint that governs the fluctuations in prime gaps and twin prime density beyond what is captured by simple independent random models. In other words, the primes are not just randomly distributed; their distribution has higher-order structure that becomes visible when viewed through the right lens.

These findings resonate with deep number-theoretic conjectures. In particular, the Hardy–Littlewood k-tuple conjecture [3] predicts structured patterns in prime constellations (including twin primes) that deviate from pure randomness by specified local density corrections (the singular series). Our entropy-derived topological loops may reflect such hidden arithmetic regularities. For example, Hardy–Littlewood gives a precise prediction for how often twin primes occur in certain intervals (in terms of an integral of the pair correlation function). The cyclic signatures we detect could be an emergent footprint of these local correlations: perhaps certain spacings of twin primes repeat more often than chance due to these arithmetic influences, thereby causing a loop in the entropy dynamics.

Despite the intriguing results, several limitations and open questions remain. First, our analysis relies on a single observable derived from the primes: the Shannon entropy of the twin prime gap distribution. While this choice proved fruitful, it is a somewhat indirect measure. It would be illuminating to incorporate additional representations of the prime sequence, such as symbol sequences encoding primes mod various bases, or the use of other entropy variants (like Rényi entropy) to capture different aspects of distribution randomness. Incorporating more features could help triangulate the source of the detected structure.

Second, we primarily focused on one-dimensional homological features (H_1 loops). It is possible that higher-dimensional topological features (like H_2 voids) or more complex invariants could also play a role. For example, if the entropy trajectory had a two-torus structure (two independent frequencies), we might detect a 2-dimensional hole. While our data did not clearly show such a feature, future analyses with longer prime sequences or different embeddings might explore this. Additionally, tools like persistent cohomology or zigzag persistence could capture time-varying topological changes in a more refined way than our simple sliding windows.

Third, the use of delay embeddings implicitly presumes that we can treat the prime gap sequence as an output of some dynamical system. This is an intriguing but nontrivial assumption: primes are of course deterministic in a formal sense (there is nothing random about the sequence of primes), but they are often modeled as a random-like process. If there is an underlying dynamical system that generates primes or prime gaps, it is not known. Our results raise the question of whether one can construct a toy dynamical model whose behavior mimics these topological signatures. Some work in the literature has attempted to find analogies of primes with chaotic systems [4], but making a direct connection remains an open challenge. Further investigation is needed to connect the topological signals we see to formal models of prime generation or to known results in analytic number theory (such as oscillatory terms in the prime counting function).

Future work may explore a number of extensions. On the topological side, computing persistent cohomology could offer insights into whether there are circular coordinates [5] that parameterize the detected loops, effectively giving a phase for the prime entropy cycle. Zigzag persistence could be used to track how homological features appear and disappear as we slide through the sequence, perhaps providing a more principled detection of events than a fixed-window approach. One could also attempt a 2-parameter persistence (e.g., considering both number range and prime value scale) for a richer picture.

On the number theory side, incorporating more arithmetic structure into the analysis could be fruitful. For example, one could tag each twin prime occurrence by its residue class or other attributes and see if the entropy of those tags yields stronger patterns. Symbolic dynamics approaches might encode prime gaps as a sequence of symbols and analyze their complexity. Moreover, connecting this empirical topology to classical conjectures (like Hardy–Littlewood) or to the Gelfond–Schneider method of exponential sums might be possible. The inclusion of arithmetic functions or transforms (e.g., using the Liouville or Möbius functions as probes) in a similar pipeline might reveal if these topological signatures are related to known phenomena like the distribution of prime k-tuples or zeros of the zeta function.

In conclusion, our study demonstrates that information theory and TDA provide powerful, complementary tools for analyzing the fine-scale structure of prime number distributions. The twin primes, long studied for their elusive pattern, show measurable topological signatures when viewed through entropy and delay embeddings. This opens up a new perspective on prime numbers: not just as objects of algebraic study but as a complex dataset that can be examined with the arsenal of modern data analysis. We anticipate that such interdisciplinary explorations will deepen our understanding of primes and possibly guide us toward new conjectures or even hints toward longstanding questions like the Twin Prime Conjecture.

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[2] S. W. Golomb, "Probability, information theory, and prime number theory," *Discrete Mathematics*, vol. 106/107, pp. 219–229, 1992.

[3] G. H. Hardy and J. E. Littlewood, "Some problems of 'partitio numerorum': III. on the expression of a number as a sum of primes," *Acta Mathematica*, vol. 44, pp. 1–70, 1923.

[4] C. Bonanno and M. S. Mega, "Toward a dynamical model for prime numbers," *Chaos, Solitons & Fractals*, vol. 20, no. 1, pp. 107–118, 2004.

[5] G. Carlsson, "Topology and data," *Bulletin of the American Mathematical Society*, vol. 46, no. 2, pp. 255–308, 2009.

[6] J. Perea and J. Harer, "Sliding windows and persistence: An application of topological methods to signal analysis," *Foundations of Computational Mathematics*, vol. 15, no. 3, pp. 799–838, 2015.

[7] G. J. Croll, "BiEntropy, TriEntropy and Primality," *Entropy*, vol. 22, no. 3, p. 311, 2020.

[8] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980*, ser. Lecture Notes in Mathematics, vol. 898, D. Rand and L.-S. Young, Eds. Springer, 1981, pp. 366–381.

[9] P. Kumar, P. C. Ivanov, and H. E. Stanley, "Information entropy and correlations in prime numbers," *arXiv:cond-mat/0303110*, 2003.