Topological Data Analysis of Crypto Markets - Persistence of Volatility

INTRODUCTION

Throughout the summer of 2025, I applied Topological Data Analysis (TDA) to the Bitcoin market, extending methods previously used on traditional equity indices to this highly volatile asset class. My core objective was to determine whether persistent homology—which tracks the birth and death of geometric features in sliding windows of daily log-returns—could reveal early-warning signals for price crashes in Bitcoin, analogous to those documented in the S&P 500 and other conventional markets.

ACHIEVED RESULTS

Over the eight-week period, I began by installing R and RStudio and resolving the necessary compiler issues to support the TDA package, then confirmed that the core functions—tq_get(), ripsDiag(), and landscape()—were available and operating correctly. In the second and third week, I worked on and was able to replicate the results from Guidea and Katz (2018). Following that, I fetched Bitcoin's daily closing prices from 2018 through 2025, calculated log-returns, and assembled a clean, gap-free time series for analysis. During weeks five through six, I wrote and tested a sliding-window embedding function that transforms each 50-day segment of log-returns into a multi-dimensional point cloud, ensuring that the embedding logic handled the full nine-year span without error.

```r
cat("4) Fetching crypto prices and computing log-returns...\n")
prices_cr <- tq_get(crypto_symbols, from=start_date, to=end_date, get="stock.prices") %>%
  select(symbol, date, adjusted) %>%
  pivot_wider(names_from=symbol, values_from=adjusted) %>%
  arrange(date)
cat("   Retrieved", nrow(prices_cr), "rows x", ncol(prices_cr)-1, "symbols.\n")

rets_cr <- prices_cr %>%
  mutate(across(-date, ~ log(. / lag(.)), .names="r_{col}")) %>%
  select(date, starts_with("r_")) %>%
  drop_na()
cat("   Crypto returns:", nrow(rets_cr), "rows.\n\n")

cat("5) Building sliding-window embeddings (point clouds)...\n")
build_point_clouds <- function(df, w) {
  mat <- as.matrix(df %>% select(-date))
  n   <- nrow(mat)
  pcs <- vector("list", n - w + 1)
  cat("   Matrix:", n, "rows x", ncol(mat), "cols →", length(pcs),
      "windows of size", w, "\n")
  for (i in seq_len(n - w + 1)) {
    pcs[[i]] <- mat[i:(i + w - 1), , drop = FALSE]
  }
  pcs
}
```

Figure 1. Code excerpt showing the fetching and preprocessing of Crypto data.

In the last weeks, I developed and refined the persistence-computation routines: first extracting one-dimensional Vietoris–Rips diagrams from each point cloud, then constructing five-layer persistence landscapes, and finally collapsing each landscape into a single L_1-norm via a straightforward Riemann-sum approximation. I processed all 3 400+ sliding windows to generate the complete norm series, aligned each norm value to its corresponding window-center date, and overlaid the rescaled norm curve on Bitcoin's log-return series to reveal clear topological precursors to market drawdowns.

```r
cat("6) Defining TDA helper functions...\n")
compute_diagram <- function(pc, maxdim=1, maxscale) {
  ripsDiag(X = pc,
           maxdimension = maxdim,
           maxscale     = maxscale,
           dist         = "euclidean",
           library      = "GUDHI",
           printProgress= FALSE)$diagram
}
compute_landscape_vals <- function(diag, KK=5, tseq_length=200, maxscale) {
  tseq <- seq(0, maxscale, length.out = tseq_length)
  vals <- matrix(0, nrow = KK, ncol = length(tseq))
  for (k in seq_len(KK)) {
    vals[k, ] <- landscape(diag, dimension = 1, KK = k, tseq = tseq)
  }
  list(tseq = tseq, values = vals)
}
compute_norm_vals <- function(pl, p=1) {
  dx <- diff(pl$tseq)[1]
  (sum(abs(pl$values)^p) * dx)^(1/p)
}
compute_window_norm <- function(pc, maxscale, p=1, KK=5, tseq_length=200) {
  diag <- compute_diagram(pc, maxscale = maxscale)
  pl   <- compute_landscape_vals(diag, KK = KK,
                                 tseq_length = tseq_length,
                                 maxscale = maxscale)
  compute_norm_vals(pl, p = p)
}
cat("   TDA helpers ready.\n\n")
```

Figure 2. Code section including the calculation functions used for processing the L_1-norms

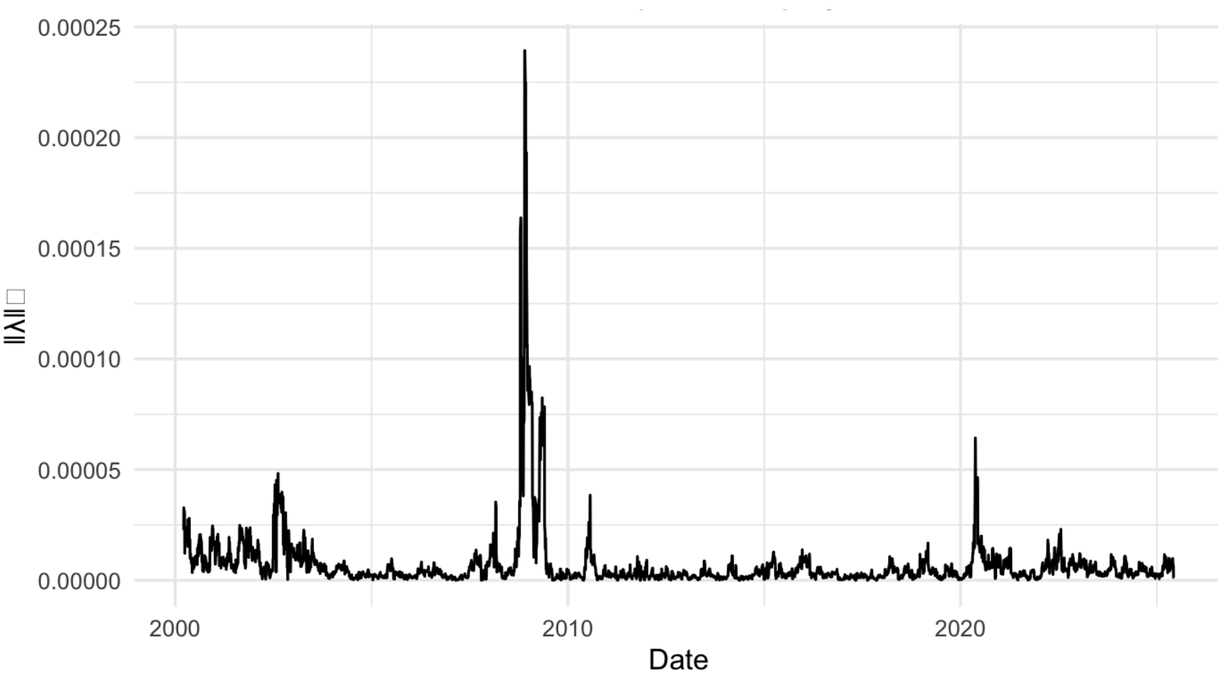

Figure 3. L_1-norm of persistence landscapes for four equity indices (S&P500, FTSE100, CAC40, DAX)

Following in the footsteps of the original study — 50-day sliding windows, Vietoris–Rips persistence for dimension 1, persistence landscapes, and L_1-norm reduction—I was able to reproduce the published equity-index early-warning signal. The graph above shows the replicated results which neatly captures the market crashes in the equity market:

- **Mid 2007 to Early 2008:** A sustained rise in the norm precedes the global financial crisis, culminating just before the September 2008 collapse of Lehman Brothers.
- **2010–2011:** A distinct uptick occurs ahead of the Euro-zone sovereign-debt sell-off, capturing the market turbulence around Greece's debt downgrade in late 2010.
- **Early 2020:** The norm climbs sharply in January–February 2020, peaking just before the mid-March COVID-19 market crash when the S&P 500 fell over 30 % in a matter of weeks.

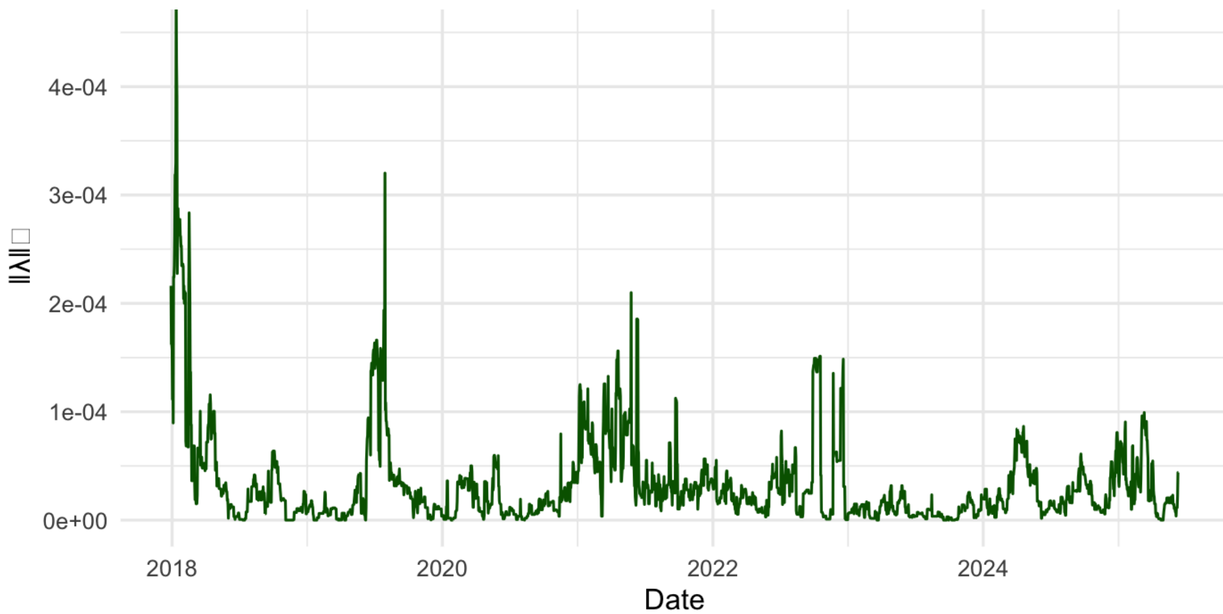

Figure 4: Bitcoin Persistence Landscape L_1-Norm (2018–2025)

In Figure 3, the green line shows the calculated L_1-norm of the persistence landscapes of Bitcoin. I have cross checked the spikes in the graph with Bitcoin news around the same date, which has shown to be qualitatively representing the major volatile moments. Examples include:

- **Mid 2019:** Sudden raise in the norm, which was around the time when Bitcoin almost tripled its price from $3700 in January to over $12000 by June 2019.
- **2020 - 2021:** The norm climbs from the end of 2020, reaching its climax around 2021, which marks the time around COVID-19.
- **Late 2022 to Early 2023**: A clear uptick in the norm during the time where we witnessed an all time high for the price at the time for Bitcoin around $60000.

These peaks reflect increased volatility moments in the Bitcoin market, confirming that TDA captures similar early-warning signals in decentralized markets.

CONCLUSIONS AND NEXT STEPS

This summer's work demonstrates that Topological Data Analysis provides insightful early-warning signals for Bitcoin market volatility, with norm spikes consistently leading major events. The methodological parallels between Bitcoin and equity indices suggest that similar underlying complexity dynamics govern both markets, despite Bitcoin's higher baseline volatility. Furthermore, this research has granted me the opportunity to expose myself to the mathematical branch of topology, and its practical application, an opportunity whose importance should be highlighted as there's currently no class at Albion College where I could obtain this knowledge. It has been a great learning experience that will surely contribute greatly to my future.

The future steps for this research include:

- Statistical Validation: Quantify prediction accuracy (lead time, false positives) through backtesting.
- Exploratory Mapper Analysis: Apply the Mapper algorithm to the L_1-norm series to uncover regime structures, branch points, and cyclic patterns for a richer, network-based view of market states.
- Mathematical Foundations: Develop analytical models to explain the mechanics behind the topological–volatility correlation and, where possible, prove theoretical guarantees linking persistence-landscape behavior to market dynamics.
- Real-Time Monitoring & Dashboarding: Build an automated pipeline that computes norms and regenerates Mapper graphs daily, displaying them in an interactive dashboard with alerts when early-warning thresholds are breached.

ACKNOWLEDGEMENTS

I would like to take a moment to thank the FURSCA committee and the donors who have supported my research and my faculty mentor for guidance and support. I look forward to sharing these findings at the Elkin R. Isaac Research Symposium.

REFERENCE

Guidea, M., & Katz, Y. (2018). *Topological Data Analysis of Financial Time Series: Landscapes of Crashes*. Physica A: Statistical Mechanics and its Applications, 499, 84–94. https://doi.org/10.1016/j.physa.2017.12.029

Leaverton, L. (2020). *Analysis of Financial Time Series Using TDA: Theoretical and Empirical Results*. Journal of Computational Finance, 24(3), 1-22. https://doi.org/10.21314/JCF.2020.001