A Persistent Homology Study of People's Daily Reporting on the United States

Ziyi Wu

Department of Political Science, University of Notre Dame

December 8, 2025

Motivation

- Propaganda literature argues that China portrays Western countries as threatening "others" to mobilize nationalism.
- Yet propaganda can *backfire* when narratives diverge too far from reality (Kao 2025).
- "Honest propaganda" emerges when the state strategically uses factual content (Chen and Xu 2017).
- Whether reporting on the U.S. always follows negative/provocative patterns remains theoretically unclear.

Research Contribution

- Fits into existing literature using topological data analysis (TDA) to study long-term narrative structure shifts.
- Apply persistent homology (PH) to daily feature vectors of U.S.-related *People's Daily* articles (2000–2023).
- Identify major structural regimes in reporting:
  - ➢ Stable from 2000 to mid-2010s.
  - ➢ Distinct structural shift in the Trump years.
  - ➢ Post-2020 simplification and homogenization.

Empirical Preview

- Articles classified along five mutually exclusive dimensions:
  - Scope
  - Reporting style (fact vs. opinion)
  - Topic
  - Sentiment
  - Narrative strategy
- Aggregated to daily feature vectors.
- PH used to compare geometric structure across U.S. presidential-era bins.

Persistent Homology in TDA

- TDA summarizes high-dimensional structure through topology.
- PH tracks the appearance and disappearance of topological features:
  - ➢ H_0: connected components (clusters)
  - ➢ H_1: loops (cyclic patterns)
- Used widely in neuroscience, ML, biology; recent expansion to NLP.

Mapping PH to Media Narratives

- Each persistent H_0 component = stable cluster of narrative configurations.
- Long-lived components = robust patterns of reporting.
- H_1 loops = recurring cycles of narrative–factual combinations.

Why Persistent Homology is Useful Here

1. The method does not require specifying the number of clusters in advance or assuming linear separation. Instead, it allows the structure of the data, how articles actually group together in feature space, to emerge from the analysis.
2. By focusing on clusters that are persistent across scales (i.e., that have large $\varepsilon_d - \varepsilon_b$), the method emphasizes reporting patterns that are not sensitive to arbitrary choices of distance thresholds.
3. Each persistent H_0-cluster and H_1-cluster can be examined in terms of its feature values (e.g., high reliance on factual reporting, particular narrative framing of the United States, or consistent use of comparison between China and foreign countries).
4. Betti curves for H_0 provide a direct way to compare how quickly daily reporting profiles merge into coherent clusters across time periods.

Corpus and Filtering

- People's Daily articles (2000–2023).
- Keyword filter for U.S.-related content.
- Total: 75,721 articles.

Five-Dimensional Classification

Articles classified using Qwen3–30B-Instruct-MLX (8bit) into:

- **Scope**: 5 categories (multilateral / bilateral, etc.).
- **Style**: quote / routine / opinion / factual.
- **Topic**: 8 areas (econ, tariffs, sanctions, etc.).
- **Sentiment**: positive / negative.
- **Narrative**: 5 framing strategies (positive / comparative positive, etc.).

Encoding and Aggregation

- Each article is coded as a 24-dimensional binary vector.
- Aggregated by day: Change into daily feature vectors.

Time Bins		
Time Range	President	Size
2000 to 2004	George W. Bush	1,452
2004 to 2008	George W. Bush	1,453
2009 to 2012	Barack Obama	1,449
2012 to 2016	Barack Obama	1,455
2016 to 2020	Donald J. Trump	1,442
2020 to 2023	Joe Biden	347

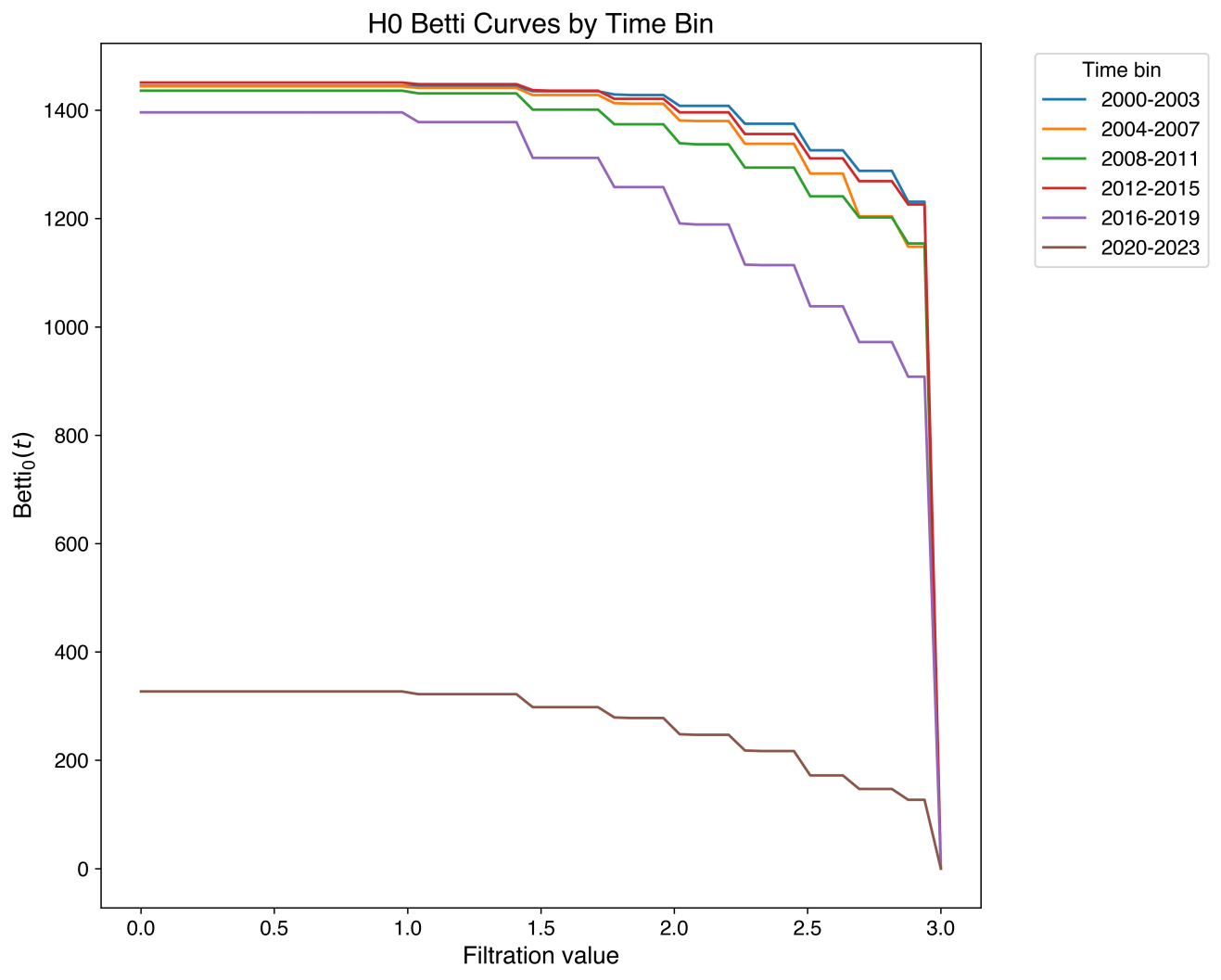
Encoded Vectors

$$\mathscr{X} = \left\{ v \in \{0,1\}^{24} \;\middle|\; \begin{array}{l} \sum_{i=1}^{5} v_i^{\text{scope}} = 1, \\ \sum_{i=1}^{4} v_i^{\text{style}} = 1, \\ \sum_{i=1}^{8} v_i^{\text{topic}} = 1, \\ \sum_{i=1}^{2} v_i^{\text{emotion}} = 1, \\ \sum_{i=1}^{5} v_i^{\text{narr}} = 1 \end{array} \right\}$$

Computing Persistent Homology

- Vietoris–Rips PH via `giotto-tda`.
- Euclidean metric; step-wise filtration function; max edge length = 3.0.
- Wasserstein distances computed with `persim`.

H0 Betti Curves



Interpretation: Betti Curves

- 2000–2015: nearly identical trajectories indicate a stable narrative structure.
- 2016–2019: earlier merging indicates more cohesive and uniform reporting.
- 2020–2023: rapid collapse indicates strong homogenization and fewer observations.

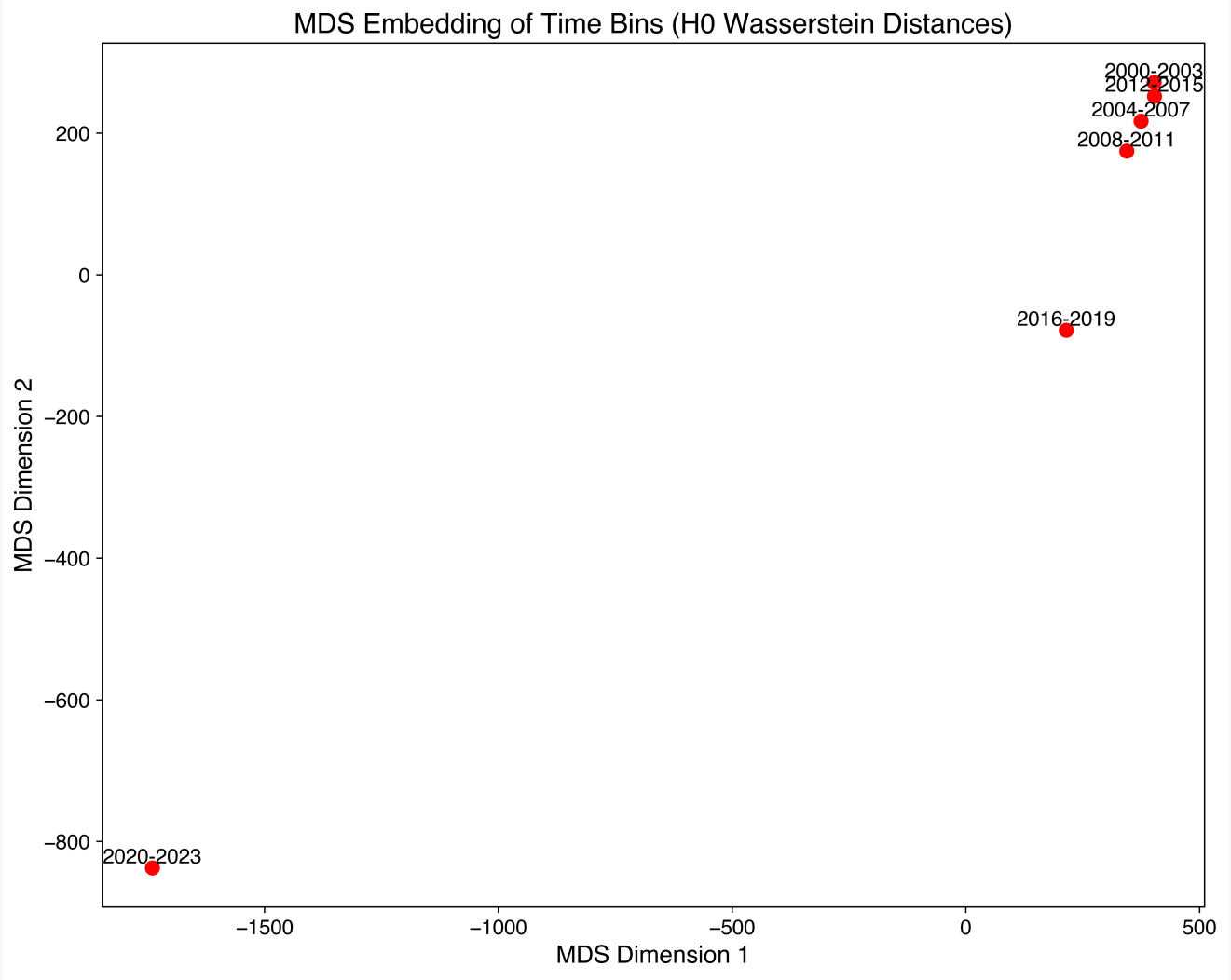
Persistence Statistics

Time Bin	Dimension	Persistence Entropy	Total Persistence	Number of Bars
2000–2003	H0	7.271806	4214.796202	1445
2000–2003	H1	3.755381	10.081300	47
2004–2007	H0	7.268863	4153.611334	1444
2004–2007	H1	4.567066	23.149875	103
2008–2011	H0	7.260125	4095.293433	1436
2008–2011	H1	4.655132	27.129023	112
2012–2015	H0	7.274373	4206.591410	1451
2012–2015	H1	4.357307	19.915693	85
2016–2019	H0	7.221857	3776.876870	1396
2016–2019	H1	5.282973	55.482896	216
2020–2023	H0	5.764442	817.141295	327
2020–2023	H1	4.092898	17.011624	68

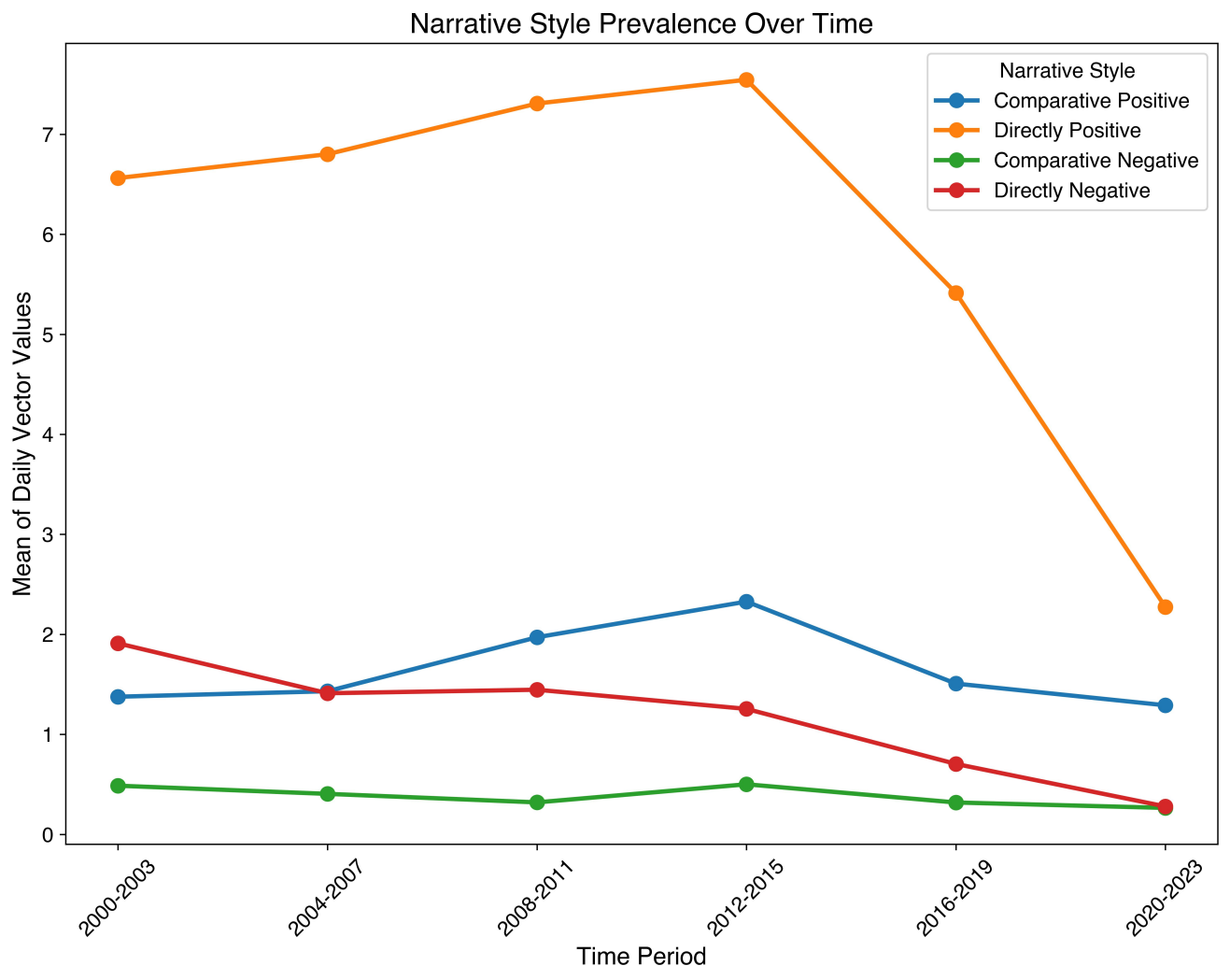
Interpretation: Persistence Statistics

- Early periods: high entropy and high total persistence imply diverse, stable patterns.
- 2016–2019: contractions in H_0 and expanded H_1 loops indicate reorganized structure.
- 2020–2023: dramatic simplification across all metrics.

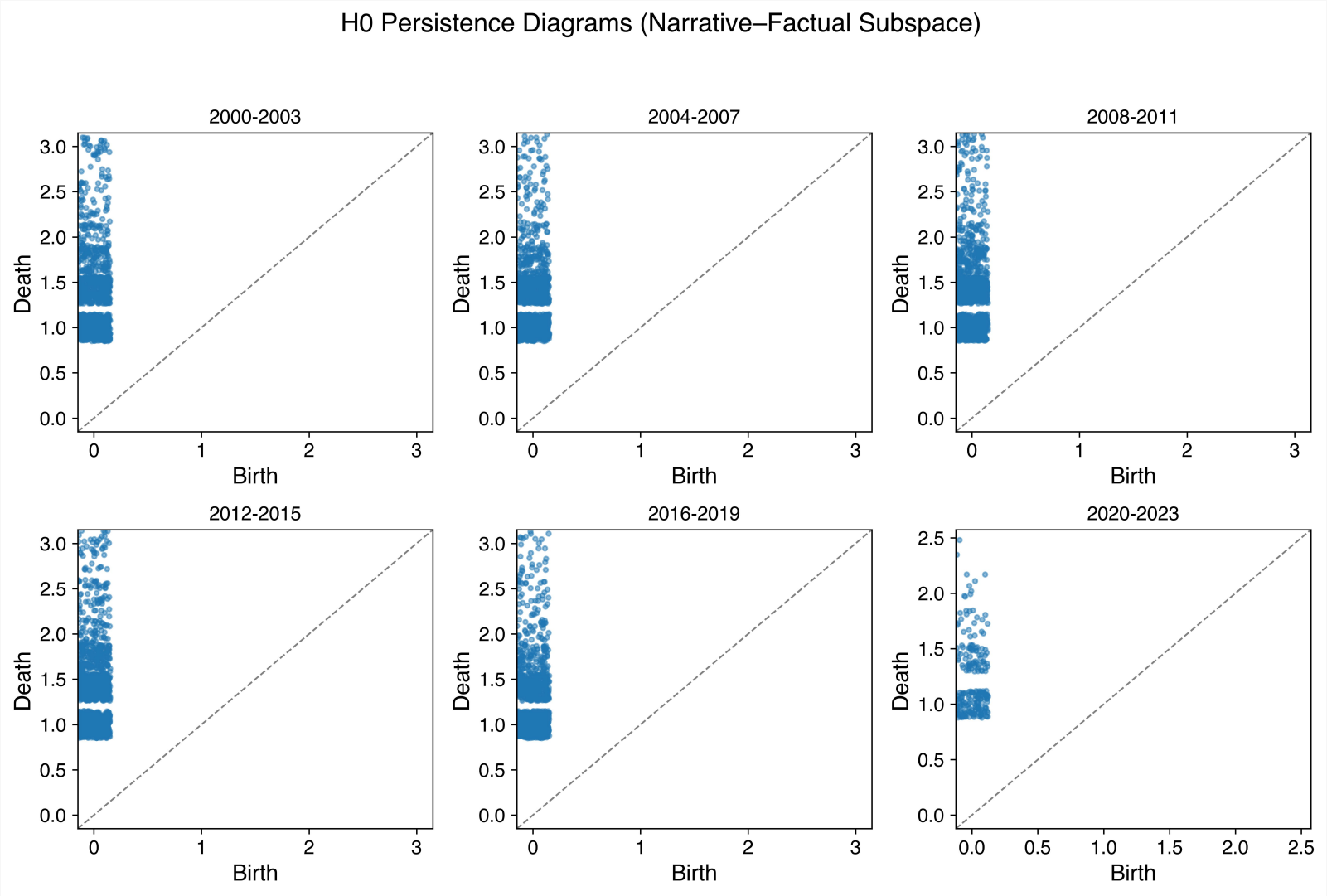
Wasserstein Distance MDS



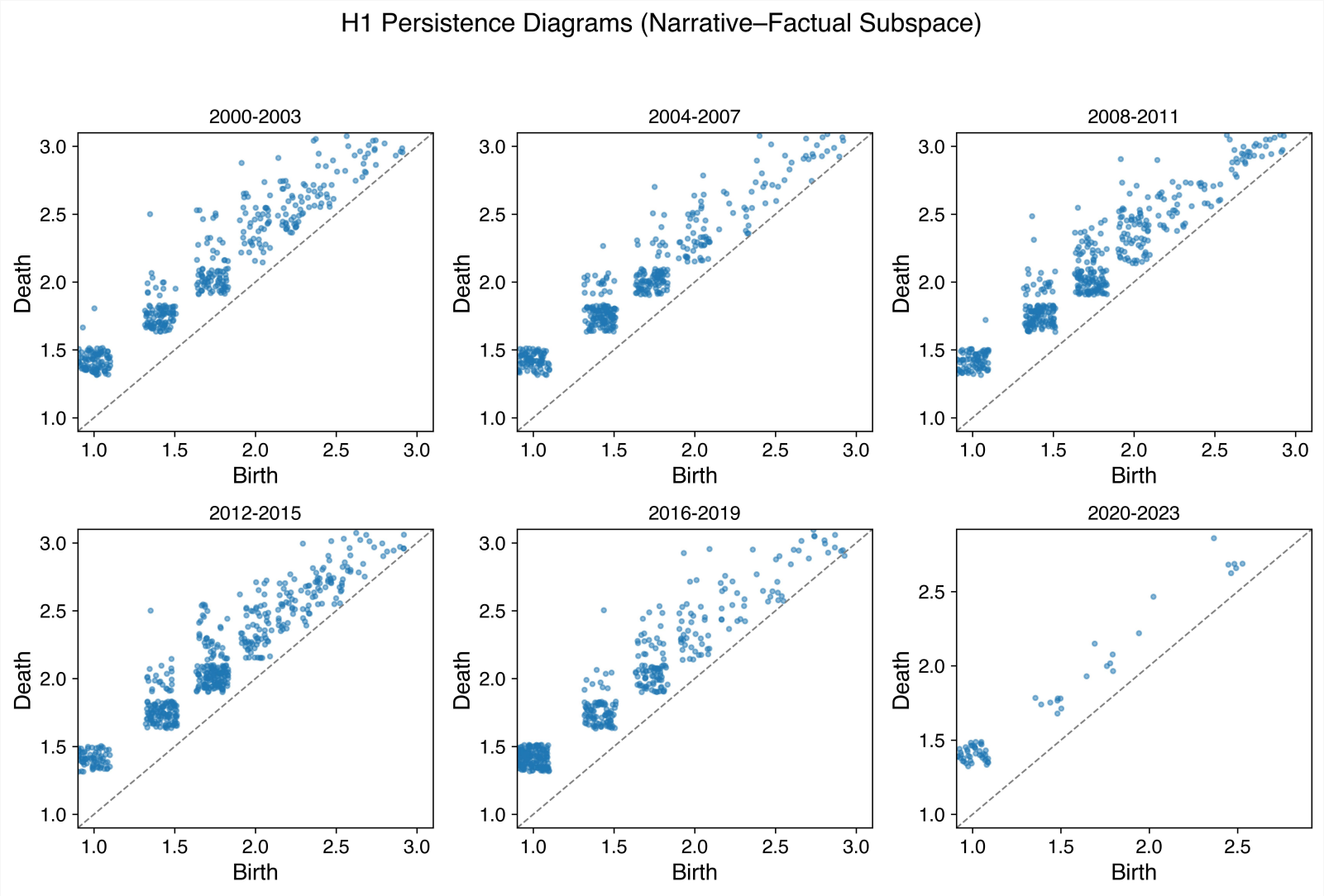
Narrative Time Series



H0 Diagrams (Narrative–Factual Subspace)



H1 Diagrams (Narrative–Factual Subspace)



Conclusion

- TDA reveals three structural regimes in U.S.-related reporting:
  1. **2000–2015:** stable, diverse, consistent structure.
  2. **2016–2019:** major reorganization; more uniform and constrained narratives.
  3. **2020–2023:** sharp homogenization; collapse of persistent structure.
- Persistent homology provides a rigorous, nonparametric way to measure shifts in propaganda strategy.
- Highlights structural—not just tonal—changes in Chinese media narratives.

References I


Chen, Jidong and Yiqing Xu (2017). “Information manipulation and reform in authoritarian regimes”. In: *Political Science Research and Methods* 5.1, pp. 163–178.


Kao, Jay C (2025). “Mistrust and Backfire: Information on Government Responsiveness and Tax Compliance in China”. In: *Social Science Quarterly* 106.4, e70049.