Topological Data Analysis Framework for Knowledge Graph Health and Efficiency: A Technical Assessment

1. Introduction

The Fully Unified Model (FUM) relies critically on its emergent Unified Knowledge Graph (UKG), a structure representing the system's accumulated knowledge and reasoning capabilities derived from underlying neural dynamics. Ensuring the health and operational efficiency of this dynamic graph is paramount for FUM's performance and reliability. However, current methodologies within FUM lack quantitative metrics capable of reliably assessing knowledge graph efficiency (information flow, inference speed, resource use) and detecting structural pathologies (fragmentation, bias amplification vectors, reasoning failure precursors). This deficiency limits FUM's self-monitoring and proactive intervention capabilities, particularly during continuous learning phases where graph structures evolve.

Existing approaches within FUM, primarily based on basic graph metrics such as node degree, edge density, path length statistics, and clustering coefficients, have proven insufficient. These measures fail to capture the higher-order structural organization and topological features—such as cycles, voids, and fragmentation patterns—that are hypothesized to be crucial for understanding knowledge representation and information flow within the complex, emergent UKG.[1] They lack the sensitivity to detect subtle structural changes or pathologies that can significantly impact reasoning performance.

To address this gap, a novel framework based on Topological Data Analysis (TDA) has been proposed. This framework utilizes persistent homology (PH), a core TDA technique, to analyze the "shape" of the UKG across multiple scales.[3] The goal is to derive quantitative topological metrics that correlate with graph efficiency and the presence of structural pathologies. Specifically, the framework introduces two primary metrics: M1 (Total B1 Persistence), quantifying cycle complexity, and M2 (Component Count), measuring fragmentation. The application of TDA, while established in other domains [4], is presented as a novel approach for analyzing the unique structural properties of emergent knowledge graphs within neuromorphic systems like FUM. This report provides a technical assessment of the proposed TDA framework, evaluating its mathematical formalism, empirical validation, integration feasibility, novelty, limitations, and overall potential significance.

2. The Proposed TDA Framework

The proposed framework applies persistent homology to characterize the topological structure of the FUM Knowledge Graph. The core components of the methodology are outlined below.

2.1 Knowledge Graph Representation

The UKG is modeled as an undirected weighted graph $G=(V,E,W)$, where V represents the set of concepts or entities (vertices), E represents the relationships between them (edges), and W contains real-valued weights signifying the strength or salience of these relationships. This graph representation serves as the foundational input for the topological analysis.

2.2 Simplicial Complex Construction and Filtration

Persistent homology requires constructing a sequence of nested topological spaces, known as a filtration, from the input data.[4] The proposed framework utilizes the Vietoris-Rips (VR) complex construction to build this filtration from the graph G. A VR complex is a standard method in TDA for building higher-dimensional structures (simplices) from pairwise relationships.[7]

The specific construction involves:

1. **Distance Metric:** A distance metric is defined based on the shortest path distance within a *thresholded* version of the graph G. An edge is included in this thresholded graph if its weight meets a certain criterion (e.g., exceeds a threshold, or inversely, is below a threshold if weights represent dissimilarity).
2. **Filtration Parameter ϵ:** A filtration parameter ϵ is introduced. For a given ϵ, an edge is included in the graph used for VR construction if the distance between the corresponding nodes (based on the thresholded shortest path metric) is less than ϵ.
3. **VR Complex:** At each value of ϵ, a VR complex is built. This complex includes vertices (0-simplices), edges (1-simplices) between nodes within distance ϵ, triangles (2-simplices) if all three pairwise distances are $\leq\epsilon$, tetrahedra (3-simplices) if all four pairwise distances are $\leq\epsilon$, and so on.[9]
4. **Filtration:** By systematically increasing ϵ from 0 up to the maximum possible distance in the graph, a nested sequence of VR complexes is generated: $VR(S,\epsilon 0)\subseteq VR(S,\epsilon 1)\subseteq\cdots\subseteq VR(S,\epsilon k)$ where $0=\epsilon 0<\epsilon 1<\cdots<\epsilon k$.[6] This sequence is the filtration.

This process transforms the static graph into a dynamic sequence of evolving topological spaces, allowing the analysis of structural features across different scales of connectivity.

2.3 Persistent Homology Computation

Persistent homology is then computed on this filtered VR complex. PH tracks the "birth" and "death" of topological features as the filtration parameter ϵ increases.[7] The dimensions of homology groups quantify specific types of features [4]:

- **H0 (0-dimensional homology):** Tracks connected components. The rank of H0, the 0th Betti number ($\beta 0$), counts the number of connected components.
- **H1 (1-dimensional homology):** Tracks loops or cycles (1-dimensional "holes"). The rank ($\beta 1$) counts the number of independent cycles.
- **H2 (2-dimensional homology):** Tracks voids or cavities (2-dimensional "holes"). The rank ($\beta 2$) counts the number of enclosed voids.

The computation results in persistence diagrams (PD0,PD1,PD2) for each dimension. A persistence diagram is a multiset of points (b,d) in a 2D plane, where b is the filtration value (scale) at which a feature appears (is "born") and d is the value at which it disappears (is "merged" or "filled in", i.e., "dies").[7] The persistence of a feature is defined as $d-b$. Features with high persistence (points far from the diagonal $d=b$) are considered robust topological signals, while low-persistence features (points close to the diagonal) are often interpreted as noise or minor structural details.[10]

2.4 Topological Metrics

From the persistence diagrams, the framework derives two specific metrics intended to quantify UKG health:

- M1: Total B1 Persistence (Cycle Structure):
  $M1=\Sigma_{(b,d)\in PD1}(d-b)$
  This metric sums the persistence values of all 1-dimensional features (cycles). It aims to capture the global complexity and robustness of cyclical pathways within the knowledge graph. A higher M1 suggests a more intricate and potentially more computationally demanding structure for information flow.
- M2: Component Count (Fragmentation):
  M2=Number of connected components in the original graph G
  This metric is equivalent to the 0th Betti number ($\beta 0$) of the initial graph (or the number of points in PD0 with infinite persistence if the filtration starts from isolated nodes). It directly measures the degree of fragmentation or disconnectedness within the knowledge graph. A higher M2 indicates a more fragmented graph, potentially representing conceptual silos or pathological states.

These metrics provide quantitative summaries derived from the topological analysis, intended for monitoring and diagnostics within FUM.

3. Novelty and Relation to Prior Art

The proposal claims novelty in applying TDA, specifically persistent homology, to analyze emergent knowledge graphs within neuromorphic systems like FUM. To assess this claim, it is necessary to compare the proposed approach with both prior methods used within FUM and the broader application of TDA in graph analysis.

3.1 Limitations of Prior FUM Metrics

As stated in the proposal, previous analysis of the FUM UKG relied on basic graph metrics: node degree, edge density, path length statistics, and clustering coefficients. While useful for characterizing local connectivity and general graph properties, these metrics fundamentally fall short in capturing the global, multi-scale structure of complex networks.[1] They are typically insensitive to the presence and arrangement of higher-order features like cycles (H1) and voids (H2), which TDA is designed to detect.[3] Furthermore, these basic metrics often fail to reliably identify fragmentation patterns or subtle structural changes that might correlate with processing inefficiencies or reasoning pathologies.[15] TDA, by focusing on the persistent topological features across scales, offers a potentially more robust way to quantify these aspects.[4]

3.2 TDA in Graph and Network Analysis

TDA and persistent homology are established mathematical frameworks with growing applications across various scientific domains, including the analysis of complex networks and graphs.[1] Researchers have applied PH to diverse network types, including social, biological, citation, and information networks, often to extract features for tasks like network comparison, classification, community detection, and anomaly detection.[1] TDA is valued for its ability to provide multi-scale summaries of network topology, capturing features invariant under continuous deformation and robust to noise.[3] Methods exist for applying PH to both undirected and directed graphs, incorporating edge weights, vertex attributes, and even temporal dynamics.[2]

3.3 Assessment of Novelty

Given the existing body of work applying TDA to networks, the fundamental mathematical techniques employed in the proposal (VR complex, PH computation) are not novel in themselves. The novelty, as claimed, resides in the specific *application domain*: analyzing the emergent, dynamically evolving knowledge graph (UKG) within a large-scale neuromorphic system (FUM) for the purpose of assessing *efficiency* and *pathology*. This context presents unique challenges and opportunities: the UKG's structure emerges from neural processes, potentially exhibiting unique topological characteristics not found in typical human-engineered KGs or simpler network models. Applying TDA to understand and monitor the health of such an emergent cognitive architecture represents a specific and potentially valuable contribution, distinct from prior FUM-internal methods and extending the application scope of TDA in AI systems. The focus on linking specific topological metrics (M1,M2) to operational concepts like efficiency and pathology within this specific context is also a key aspect of the proposed contribution.

4. Empirical Validation

The proposal includes results from empirical validation studies designed to assess the feasibility and potential utility of the derived topological metrics (M1,M2).

4.1 Experimental Setup

The validation was performed using 10 synthetically generated knowledge graph snapshots, each with 100 nodes. These graphs were designed to exhibit varying structural properties, including different network topologies (random, small-world, scale-free), varying levels of fragmentation (1 to 17 components), and different densities of cycles. Crucially, each synthetic graph was associated with pre-defined "efficiency" and "pathology" scores, intended to serve as ground truth proxies for the target properties the TDA metrics aim to capture. The exact method for generating these proxy scores is not detailed in the provided document.

4.2 Validation Results

- **Unit Tests:** Basic functionality tests confirmed that the implementation could correctly count connected components (M2) and quantify cycle presence (M1) in the test graphs. Computation times were reported as efficient for these small (100-node) graphs, with all metrics calculated in under 0.1 seconds.
- **System Tests (Correlation Analysis):** The core validation involved correlating the proposed TDA metrics with the predefined proxy scores across the 10 synthetic graphs. The results showed:
  - A strong negative correlation between Total B1 Persistence (M1) and the Efficiency Score (Pearson $r = -0.8676$, $p = 0.001143$). This suggests that graphs with more complex and persistent cycle structures (higher M1) tend to have lower assigned efficiency scores.
  - An extremely strong positive correlation between Component Count (M2) and

the Pathology Score (Pearson r = 0.9967, p = 5.289e-10). This indicates that higher fragmentation (more components, higher M2) strongly correlates with higher assigned pathology scores in the synthetic data.
- **Performance:** Average computation times for the different stages of the analysis pipeline on 100-node graphs were reported: graph construction (~0.0009s), distance matrix calculation (~0.0014s), persistence calculation (~0.046s), and metric extraction (~0.0007s), leading to a total analysis time of approximately 0.05 seconds per snapshot. The persistence calculation dominates the runtime, as expected.

4.3 Critique of Validation Approach

While the reported correlations on synthetic data are statistically significant and align with the intended interpretations of M1 and M2, the validation methodology has significant limitations:

1. **Reliance on Synthetic Data:** The experiments were conducted exclusively on small, synthetic graphs. The structural properties of these graphs may not accurately reflect the complexity, scale, or emergent characteristics of the actual FUM UKG. Validation on real UKG snapshots is essential.
2. **Use of Proxy Scores:** The correlation analysis relies on predefined "efficiency" and "pathology" scores associated with the synthetic graphs. The document does not specify how these scores were generated or whether they reliably represent actual FUM performance or pathological states. The strong correlations observed might reflect the construction of the synthetic data and proxy scores rather than a genuine link between the TDA metrics and real-world FUM behavior. Establishing correlation with *actual* FUM performance metrics (e.g., inference latency, task accuracy, resource consumption, documented failure modes) is crucial.
3. **Non-Standard Filtration:** The proposal describes constructing the VR complex based on shortest path distances in a *thresholded* graph. This is a non-standard approach for VR filtrations, which are typically built directly on a metric space (e.g., using edge weights as distances or embedding nodes in a metric space). The rationale for this specific choice, the thresholding method used, and its potential impact on the resulting homology and metrics are not discussed. Standard practice often involves filtering directly by edge weights or using geometric embeddings.[2] This methodological choice requires justification and sensitivity analysis.

In summary, the initial validation shows promise but suffers from a significant gap between the synthetic test environment and the target application within FUM.

Further validation on representative FUM data against real performance indicators is necessary.

5. FUM Integration Assessment

Integrating the proposed TDA framework into the FUM system involves considerations regarding software components, resource requirements, and scalability.

5.1 Component Additions

Implementation requires adding specific components to FUM's monitoring subsystem:

- A dedicated module for KG topology analysis.
- Integration of a library capable of computing persistence diagrams from filtered complexes, such as Ripser.[18]
- Mechanisms for tracking the derived metrics (M1,M2) across different phases of FUM's operation, particularly during training and continuous learning.

5.2 Resource Impact

The computational resources required by the framework are a significant concern:

- **Memory:** Calculating the all-pairs shortest path distance matrix requires O(n2) memory, where n is the number of nodes (concepts) in the UKG.
- **Computation:** The most computationally intensive step is typically the persistent homology calculation. Standard algorithms for PH can have worst-case complexities ranging from O(m3) to higher powers, where m is the number of simplices in the largest complex in the filtration.[12] The size of the VR complex can grow rapidly, potentially exponentially in n in the worst case, although it is often smaller for real-world data.[20] The proposal cites a worst-case complexity of O(n3) for persistence calculation, likely assuming algorithms efficient for sparse inputs or specific complex types. Even O(n3) represents a significant computational burden for large graphs. Optimizations for sparse graphs are mentioned as potentially applicable to FUM KGs.

5.3 Scaling Considerations

The computational complexity poses a major challenge for applying this framework to potentially large FUM KGs:

- **Large Graphs:** For graphs exceeding approximately 104 nodes, the proposal acknowledges that the direct computation becomes infeasible and suggests resorting to approximation techniques like graph subsampling or landmark-based persistence calculations. The impact of such approximations on the accuracy and reliability of the M1 and M2 metrics needs careful evaluation.

- **Distributed Computation:** The possibility of distributing the computation for large-scale analysis is mentioned, which could leverage FUM's infrastructure but requires specific algorithmic adaptations.
- **Asynchronous Execution:** Performing the TDA analysis asynchronously to the main FUM learning processes is suggested to mitigate performance impact on core operations.

The practical feasibility of this framework hinges critically on addressing these scaling challenges. The reported performance of ~0.05 seconds for 100-node graphs does not extrapolate favorably to graphs with potentially millions or billions of nodes/edges, where O(n3) or even O(n2) computations become prohibitive. Research into scalable TDA algorithms, including approximation methods and potentially leveraging insights from related work like Knowledge Persistence which evaluates KGs using only a fraction of data [21], is essential for practical deployment within FUM. The effectiveness and trade-offs of proposed scaling solutions like subsampling require thorough empirical validation on realistic UKG data.

6. Limitations and Future Work

The proposal acknowledges several limitations and outlines directions for future research.

6.1 Acknowledged and Identified Limitations

- **Computational Complexity:** As discussed above, the poor scaling (O(n3) or worse) with graph size is a primary limitation, potentially rendering the approach impractical for large FUM KGs without effective scaling strategies.[3]
- **Undirected Graph Assumption:** The framework currently operates on an undirected representation of the UKG. Real-world KGs often have directed relationships, and ignoring directionality might discard crucial information about information flow and dependencies.[1]
- **Limited Testing on Real Data:** Validation has been confined to small, synthetic graphs, lacking confirmation on actual FUM UKGs against real performance metrics.
- **Lack of Formal Verification:** The framework's development and validation are empirical. While correlations are observed, the causal links and the framework's behavior under varying conditions are not formally proven.
- **Non-Standard Filtration:** The use of a VR filtration based on shortest paths in a *thresholded* graph is unconventional and lacks clear justification or analysis of its impact compared to more standard filtrations (e.g., based directly on edge

weights).[2]

- **Information Loss in Metrics:** The proposed metrics, M1 (total H1 persistence) and M2 (initial H0 count), are highly simplified summaries of the rich information contained in the full persistence diagrams (PD0,PD1,PD2). They discard details about the number, distribution, and specific birth/death times of individual topological features.[4] This information loss might limit the framework's ability to distinguish between different types of structural issues or its sensitivity to subtle changes. More sophisticated summaries like persistence landscapes, images, or distances between diagrams (e.g., Wasserstein, bottleneck) capture more geometric information from the PDs.[4]
- **Interpretability and Parameter Sensitivity:** While M1 and M2 are presented as interpretable, TDA results, in general, can be sensitive to parameter choices (e.g., filtration construction, distance metrics) and their interpretation can require expertise.[3] The sensitivity of M1 and M2 to the choices made in this framework (especially the thresholding for the distance metric) needs investigation.

6.2 Proposed Future Work

The proposal outlines several directions for future development aimed at addressing these limitations:

- **Scalability Enhancements:** Develop and evaluate spectral graph theory approximations or other methods for faster computation on large graphs.
- **Incorporate Directionality:** Extend the framework to handle directed graph structures (digraphs), potentially using alternative filtrations or homology theories designed for directed graphs.[2]
- **Temporal Analysis:** Track the topological metrics (M1,M2) over time during FUM's learning phases to detect pathological transitions or monitor structural evolution. This aligns with the need to analyze dynamic systems.[23]
- **Explore Higher Dimensions:** Investigate the utility of higher-dimensional homology features (H2 for voids, and potentially beyond) for characterizing more complex knowledge structures.[3]

These future directions are pertinent, particularly the focus on scalability, handling directedness, and temporal analysis, which are critical for practical application to the dynamic FUM UKG. Exploring alternative, potentially more informative, TDA-derived metrics beyond M1 and M2 should also be considered.

7. Broader Context and Significance

The proposed framework exists within the broader context of applying TDA to understand complex data structures, particularly networks and graphs.

7.1 TDA in Network and Knowledge Graph Analysis

TDA is increasingly recognized as a valuable tool for analyzing complex networks, offering insights complementary to traditional graph theory metrics.[1] Its strength lies in quantifying multi-scale topological features like connectivity, cycles, and voids, which often relate to the function and robustness of the network.[4] Applications include identifying community structures [25], comparing network architectures [16], detecting anomalies [4], analyzing biological networks [24], and even evaluating the quality of knowledge graph embeddings.[21] The core idea is that the "shape" of the network connectivity, as captured by persistent homology, reveals fundamental properties missed by purely local or statistical measures.[28]

7.2 Framework Design Choices and Alternatives

The specific choices made in this proposal—using an undirected graph, a particular VR filtration based on thresholded shortest paths, and summarizing results with M1 and M2—represent only one point in a wide design space for applying TDA to graphs. Numerous alternative filtration methods exist, designed for weighted graphs, directed graphs, vertex attributes, or dynamic networks.[2] Similarly, the information from persistence diagrams can be summarized or utilized in various ways beyond simple summation (M1) or counting (M2).

Table 1: Comparison of Selected TDA-Derived Metrics for Graph Analysis

Metric/Representation	Information Captured	Interpretability	Computational Cost (Relative)	Potential Use Cases	References
M2 (Component Count)	Initial Fragmentation (H0)	High	Low (Graph traversal)	Basic connectivity check, detecting gross fragmentation	
M1 (Total B1 Persist.)	Global H1 cycle complexity (sum of	Medium	High (PH calculation)	Overall measure of cyclic structure	

	persistence)			complexity, potential correlate of efficiency	
Betti Curves	Number of features (βk) vs. filtration parameter	Medium	High (PH calculation)	Visualizing feature counts evolution across scales	4
PD Distances (Wass., Bottleneck)	Geometric distance between full PDs	Medium-Low	High (PH + Distance calc.)	Comparing overall topological similarity of graphs, KG completion evaluation	4
Persistence Landscapes /Images	Vector space embeddings of PDs	Low	High (PH + Vectorization)	Input features for ML models, statistical analysis, graph classification	4
Persistent Entropy	Information-theoretic summary of PD point distribution	Medium-Low	High (PH + Entropy calc.)	Quantifying complexity/uncertainty of topological features	5

This table highlights that the proposed M1 and M2 prioritize simplicity and direct interpretability (especially M2) at the cost of potentially significant information loss compared to methods that utilize the full geometry of the persistence diagrams (e.g., PD distances, landscapes, images). The choice of metrics should depend on whether these simple summaries are sufficiently sensitive and predictive for the specific task of monitoring FUM UKG health, or if more complex representations are needed.

7.3 Potential for TDA in FUM Beyond Monitoring

While the primary goal of the proposal is monitoring efficiency and pathology, the capabilities of TDA suggest broader potential applications within FUM. TDA excels at exploratory data analysis and discovering hidden structures in complex, high-dimensional data.[3] Applied to the UKG, TDA could potentially:

- **Characterize Learning Dynamics:** Identify distinct topological signatures associated with different stages of learning or adaptation in FUM.
- **Discover Emergent Structures:** Uncover previously unknown structural patterns (e.g., specific types of voids or cycles) that correlate with higher cognitive functions or specific failure modes.
- **Compare UKG States:** Quantify structural differences between UKGs resulting from different training protocols, architectures, or datasets using metrics like PD distances.[16]
- **Guide Learning:** Incorporate topological features or constraints directly into the FUM learning process, perhaps through topological regularization techniques aimed at promoting desirable structural properties.[6]

These exploratory avenues leverage TDA's strength in understanding the intrinsic "shape" of data [24], potentially offering deeper insights into the FUM's internal knowledge organization beyond simple health checks.

8. Overall Assessment and Recommendations

The proposed TDA framework represents a principled approach to address the critical need for more sophisticated monitoring tools for FUM's Unified Knowledge Graph. By leveraging persistent homology, it offers the potential to quantify higher-order structural features related to efficiency and pathology that are missed by conventional graph metrics. The initial validation on synthetic data demonstrates the framework's feasibility and shows promising correlations between the proposed metrics (M1,M2) and proxy scores for efficiency and pathology.

However, the framework faces significant challenges that must be overcome for practical deployment. The most critical issues are:

1. **Validation Gap:** The reliance on small, synthetic datasets and undefined proxy scores leaves a substantial gap in demonstrating efficacy on real, large-scale FUM UKGs against actual system performance metrics.
2. **Scalability:** The inherent computational complexity of persistent homology (O(n3) or worse) poses a major barrier to analyzing potentially massive UKGs. Proposed scaling solutions (subsampling, approximation, parallelization) require rigorous development and validation.
3. **Methodological Concerns:** The non-standard VR filtration definition requires

justification, and the simple summary metrics (M1,M2) discard considerable information from the persistence diagrams, potentially limiting sensitivity. The restriction to undirected graphs is also a limitation for many KGs.

The potential significance of a validated, scalable TDA framework for FUM is high. It could provide unprecedented insights into the UKG's structural health, enabling proactive maintenance, improved diagnostics, and potentially even guiding FUM's development. However, failure to address the validation and scaling challenges would render the current proposal impractical.

Based on this assessment, the following recommendations are made:

1. **Prioritize Real-World Validation:** Shift focus from synthetic data to rigorous testing on representative FUM UKG snapshots of varying sizes and from different operational phases. Evaluate the correlation of M1, M2, and potentially other TDA metrics against *actual* FUM performance indicators (e.g., inference time, task success rates, resource utilization, known pathological states). Benchmark against baseline graph metrics on this real data.
2. **Conduct Rigorous Scalability Research & Development:** Actively pursue and benchmark scalable TDA solutions tailored to the UKG context:
   - Quantify the trade-offs (accuracy vs. speed) of subsampling and landmark-based PH methods for M1/M2 calculation.
   - Investigate and implement state-of-the-art approximate PH algorithms or related techniques (e.g., spectral methods), assessing their suitability.
   - Develop and evaluate parallel/distributed PH computation strategies leveraging FUM's infrastructure.
   - Explore if techniques inspired by efficient KG evaluation methods like Knowledge Persistence [21] can be adapted.
3. **Investigate Alternative Filtrations and Metrics:** Broaden the methodological exploration beyond the current proposal:
   - Implement and evaluate standard graph filtrations (e.g., based directly on edge weights using intrinsic distances) and compare results with the current thresholded shortest-path VR approach.
   - Explore filtrations designed for weighted and directed graphs.[2]
   - Compute and evaluate alternative TDA summaries (e.g., persistence landscapes/images, PD distances like Wasserstein [4]) alongside M1,M2. Assess if these richer representations offer better predictive power or robustness, potentially as features for a simple predictive model.
   - Evaluate the utility of H0 persistence beyond the initial count and the potential information contained in H2 features.

4. **Perform Sensitivity Analysis:** Systematically analyze the impact of parameter choices (e.g., graph thresholding parameters, filtration parameters) on the stability and predictive power of the TDA metrics. Justify the choice of the VR filtration method.
5. **Develop Temporal TDA Capabilities:** Progress the future work on temporal analysis by implementing and testing methods to track UKG topological evolution during FUM's continuous learning, identifying characteristic changes or early warnings of degradation.

Addressing these recommendations is crucial to determine the true viability and value of this TDA-based approach for monitoring the health and efficiency of the FUM Knowledge Graph.

9. References

- User Query Document: *Topological Data Analysis Framework for Knowledge Graph Health and Efficiency*, Justin Lietz, 4/2/2025.
- [3] arXiv:2401.04250v1
- [7] arXiv:2411.10298
- [6] arXiv:2302.03836
- [11] arXiv:2406.04102v1
- [30] arXiv:2405.04796
- [4] ResearchGate Publication 378214130
- [21] arXiv:2301.12929
- [31] Wasserman, L. (2016). Topological Data Analysis. *Annual Review of Statistics and Its Application*. (Accessed via math.uri.edu)
- [32] Reddit r/datascience discussion (2019)
- [18] MathOverflow Question 141157
- [12] Kerber, M. (2016). Persistent Homology. *Internationale Mathematische Nachrichten*. (Accessed via geometrie.tugraz.at)
- [1] Aktas et al. (2019). Persistent homology for analysing network data. *Applied Network Science*. (Accessed via d-nb.info)
- [14] Horak, D., Maletić, S., & Rajković, M. (2009). Persistent homology of complex networks. *Journal of Statistical Mechanics: Theory and Experiment*. (Accessed via ResearchGate)
- [10] Otter, N., Porter, M. A., Tillmann, U., Grindrod, P., & Harrington, H. A. (2017). A roadmap for the computation of persistent homology. *EPJ Data Science*. (Accessed via math.ucla.edu)
- [2] Aktas, E., Akbas, E., & Aktas, M. E. (2019). Persistent Homology for Network Comparison. *arXiv:1907.08708*.

- [16] Huang, W., & Ribeiro, A. (2015). Persistent Homology Lower Bounds on High Order Network Distances. *IEEE Global Conference on Signal and Information Processing*. (Accessed via seas.upenn.edu)
- [13] Reininghaus, J., et al. (2015). A Stable Multi-Scale Kernel for Topological Data Analysis. *CVPR*. (Accessed via NSF PAR)
- [15] Aksoy, S. G., et al. (2020). Graph Theoretical and Topological Data Analysis for Structure Identification in Social Networks for Bot Detection. *Mathematics*. (Accessed via MDPI)
- [33] MathStackExchange Question 3958944
- [8] Chazal, F. (2022). Introduction to Topological Data Analysis - Lecture 1. (Accessed via spatstat.org)
- [20] MathOverflow Question 288115
- [22] El Gazzar, A., et al. (2023). Topological Data Analysis and Persistent Homology for Brain Network Analysis: A Narrative Review. *Brain Sciences*. (Accessed via PMC)
- [34] Dey, T. K. Simplicial Complexes. Course Notes. (Accessed via cs.purdue.edu)
- [9] CompTΔG Tutorial (2019). Rips Filtration Tutorial for the R-TDA Package. (Accessed via comptag.github.io)
- [35] MathStackExchange Question 3627924
- [36] Dow, E., et al. (2017). Weighted Persistent Homology. *arXiv:1709.00097*.
- [19] Cade, C., et al. (2023). Complexity of Quantum Algorithms for Topological Data Analysis. *PRX Quantum*. (Accessed via APS)
- [37] MathStackExchange Question 4701286
- [23] Amézquita, E. J., et al. (2020). Topological data analysis for discovery in preclinical spinal cord injury studies. *Scientific Reports*. (Accessed via PMC)
- [24] Amézquita, E. J., et al. (2020). What is topological data analysis?. *PLoS Computational Biology*. (Accessed via PMC)
- [5] Huang, W., et al. (2023). A Survey on Topological Data Analysis in Machine Learning. *Entropy*. (Accessed via MDPI)
- [38] ResearchGate Figure 340439430 (Various citations within figure description)
- [28] El Gazzar, A., et al. (2023). Topological Data Analysis for Multivariate Time Series Data with Applications. *Entropy*. (Accessed via MDPI)
- [17] Georgia Tech TDA Projects Website (tda.gatech.edu)
- [29] Quantmetry Blog (2020). Topological Data Analysis with Mapper.
- [26] Yoo, J., et al. (2024). Altered topological structure of the brain white matter structural covariance network in maltreated children: A topological data analysis. *Network Neuroscience*. (Accessed via MIT Press)
- [25] Chen, C., & Paffenroth, R. (2022). Network Community Detection and Characterization using Persistent Homology. *arXiv:2204.03191*.

- [27] PathologyOutlines.com Topic: Informatics - Graph Neural Networks
- [21] Browsing Summary: Knowledge Persistence (KP) Method for KG Completion Evaluation.
- [2] Browsing Summary: Alternative Filtration Methods for Persistent Homology on Networks.
- [14] Browsing Summary: Persistent Homology for Network Type Distinction and Robustness Analysis.
- [10] Browsing Summary: Computational and Statistical Challenges in Persistent Homology.

Works cited

1. Persistence Homology of Networks: Methods and Applications, accessed April 11, 2025, https://d-nb.info/1203059833/34
2. arxiv.org, accessed April 11, 2025, https://arxiv.org/pdf/1907.08708
3. Explaining the Power of Topological Data Analysis in Graph Machine Learning - arXiv, accessed April 11, 2025, https://arxiv.org/html/2401.04250v1
4. CHAPTER 3 Topological Data Analysis for Intelligent Systems and Applications - ResearchGate, accessed April 11, 2025, https://www.researchgate.net/profile/Alperen-Eroglu/publication/378214130_Topological_Data_Analysis_for_Intelligent_Systems_and_Applications/links/65cddb5d34bbff5ba7175936/Topological-Data-Analysis-for-Intelligent-Systems-and-Applications.pdf
5. TREPH: A Plug-In Topological Layer for Graph Neural Networks - MDPI, accessed April 11, 2025, https://www.mdpi.com/1099-4300/25/2/331
6. arXiv:2302.03836v1 [cs.LG] 8 Feb 2023, accessed April 11, 2025, https://arxiv.org/pdf/2302.03836
7. arXiv:2411.10298v2 [cs.CL] 14 Dec 2024, accessed April 11, 2025, https://arxiv.org/pdf/2411.10298?
8. Topological Data Analysis I Applications in Spatial statistics - spatstat, accessed April 11, 2025, https://spatstat.org/Aalborg2022/notes/TDA_Lecture1_part1.pdf
9. Rips Filtration Tutorial for the R-TDA Package, accessed April 11, 2025, https://comptag.github.io/rpackage_tutorials//2019/07/tda-rips-tutorial.html
10. www.math.ucla.edu, accessed April 11, 2025, https://www.math.ucla.edu/~mason/papers/roadmap-final.pdf
11. Chromatic Topological Data Analysis - arXiv, accessed April 11, 2025, https://arxiv.org/html/2406.04102v1
12. Persistent Homology – State of the art and challenges - TU Graz, accessed April 11, 2025, https://www.geometrie.tugraz.at/kerber/kerber_papers/imn_article_2016.pdf
13. Learning metrics for persistence-based summaries and applications for graph classification, accessed April 11, 2025, https://par.nsf.gov/servlets/purl/10180702
14. (PDF) Persistent Homology of Complex Networks - ResearchGate, accessed April 11, 2025,

https://www.researchgate.net/publication/23418949_Persistent_Homology_of_Complex_Networks
15. Bot Detection on Social Networks Using Persistent Homology - MDPI, accessed April 11, 2025, https://www.mdpi.com/2297-8747/25/3/58
16. Persistent Homology Lower Bounds on High Order Network Distances - Penn Engineering, accessed April 11, 2025, https://www.seas.upenn.edu/~aribeiro/preprints/2015_huang_ribeiro_a.pdf
17. Projects - TDAlab @ GA Tech., accessed April 11, 2025, http://tda.gatech.edu/projects
18. Inference using Topological Data Analysis: Is it worth it for a regular statistician to learn TDA? - MathOverflow, accessed April 11, 2025, https://mathoverflow.net/questions/141157/inference-using-topological-data-analysis-is-it-worth-it-for-a-regular-statisti
19. Complexity-Theoretic Limitations on Quantum Algorithms for Topological Data Analysis, accessed April 11, 2025, https://link.aps.org/doi/10.1103/PRXQuantum.4.040349
20. Complexity of computing the Vietoris-Rips complex - MathOverflow, accessed April 11, 2025, https://mathoverflow.net/questions/288115/complexity-of-computing-the-vietoris-rips-complex
21. arxiv.org, accessed April 11, 2025, https://arxiv.org/abs/2301.12929
22. Topological Data Analysis for Multivariate Time Series Data - PMC - PubMed Central, accessed April 11, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC10669999/
23. Using topological data analysis and pseudo time series to infer temporal phenotypes from electronic health records - PubMed Central, accessed April 11, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC7536308/
24. The shape of things to come: Topological data analysis and biology, from molecules to organisms - PMC, accessed April 11, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC7383827/
25. [2204.03191] Efficient Community Detection in Large-Scale Dynamic Networks Using Topological Data Analysis - arXiv, accessed April 11, 2025, https://arxiv.org/abs/2204.03191
26. Altered topological structure of the brain white matter in maltreated children through topological data analysis | Network Neuroscience - MIT Press Direct, accessed April 11, 2025, https://direct.mit.edu/netn/article/8/1/355/118706/Altered-topological-structure-of-the-brain-white
27. Application of graph neural networks to whole slide images - Pathology Outlines, accessed April 11, 2025, https://www.pathologyoutlines.com/topic/informaticsgraphneuralnetworkstowholeslideimages.html
28. Topological Data Analysis for Multivariate Time Series Data - MDPI, accessed April 11, 2025, https://www.mdpi.com/1099-4300/25/11/1509
29. Topological data analysis with Mapper - Quantmetry, accessed April 11, 2025,

https://www.quantmetry.com/blog/topological-data-analysis-with-mapper/
30. [2405.04796] Persistent homology of featured time series data and its applications - arXiv, accessed April 11, 2025, https://arxiv.org/abs/2405.04796
31. arXiv:1609.08227v1 [stat.ME] 27 Sep 2016, accessed April 11, 2025, https://www.math.uri.edu/~thoma/comp_top__2018/Wasserman_2016_TopologicalDataAnalysis.pdf
32. Are concepts of topology relevant to data scientists? : r/datascience - Reddit, accessed April 11, 2025, https://www.reddit.com/r/datascience/comments/dl473r/are_concepts_of_topology_relevant_to_data/
33. Distances between two complexes when using Persistence Homology, accessed April 11, 2025, https://math.stackexchange.com/questions/3958944/distances-between-two-complexes-when-using-persistence-homology
34. Computational Topology for Data Analysis: Notes from Book by, accessed April 11, 2025, https://www.cs.purdue.edu/homes/tamaldey/course/531/Simplicial-complex2.pdf
35. Good Stopping Criteria for Persistent Homology - Mathematics Stack Exchange, accessed April 11, 2025, https://math.stackexchange.com/questions/3627924/good-stopping-criteria-for-persistent-homology
36. arXiv:1709.00097v2 [math.AT] 6 Dec 2018, accessed April 11, 2025, https://arxiv.org/pdf/1709.00097
37. Using topological data analysis to analyze a graph - Mathematics Stack Exchange, accessed April 11, 2025, https://math.stackexchange.com/questions/4701286/using-topological-data-analysis-to-analyze-a-graph
38. Applications of topological data analysis (TDA) to biology. A - ResearchGate, accessed April 11, 2025, https://www.researchgate.net/figure/Applications-of-topological-data-analysis-TDA-to-biology-A-Structural-biology-A_fig6_340439430