# Technical Assessment of the Topological Data Analysis Framework for Knowledge Graph Health and Efficiency by Justin Lietz

Gemini 2.5 Pro: Deep Research

April 12, 2025

## 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 [2]. Ensuring the health and operational efficiency of this dynamic graph is paramount for FUM's performance and reliability [3]. 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) [4]. This deficiency limits FUM's self-monitoring and proactive intervention capabilities, particularly during continuous learning phases where graph structures evolve [5].

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 [6]. 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 [7]. 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 [9]. This framework utilizes persistent homology (PH), a core TDA technique, to analyze the "shape" of the UKG across multiple scales

## 2 The Proposed TDA Framework

[15]

The proposed framework applies persistent homology to characterize the topological structure of the FUM Knowledge Graph

### 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

### 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

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)

   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

- $H_0$ (0-dimensional homology): Tracks connected components. The rank of $H_0$, the 0th Betti number ($\beta_0$), counts the number of connected components.
- $H_1$ (1-dimensional homology): Tracks loops or cycles (1-dimensional "holes"). The rank ($\beta_1$) counts the number of independent cycles.
- $H_2$ (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 ($PD_0$, $PD_1$, $PD_2$) for each dimension

### 2.4 Topological Metrics

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

- $M_1$: Total B1 Persistence (Cycle Structure):

  $$M_1 = \sum_{(b,d) \in PD_1} (d - b)$$

  This metric sums the persistence values of all 1-dimensional features (cycles)
  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

### 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

### 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

### 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

## 4 Empirical Validation

[?]
The proposal includes results from empirical validation studies designed to assess the feasibility and potential utility of the derived topological metrics ($M_1, M_2$)

### 4.1 Experimental Setup

The validation was performed using 10 synthetically generated knowledge graph snapshots, each with 100 nodes

### 4.2 Validation Results

– Unit Tests: Basic functionality tests confirmed that the implementation could correctly count connected components ($M_2$) and quantify cycle presence ($M_1$) in the test graphs

### 4.3 Critique of Validation Approach

While the reported correlations on synthetic data are statistically significant and align with the intended interpretations of $M_1$ and $M_2$, the validation methodology has significant limitations:

(a) Reliance on Synthetic Data: The experiments were conducted exclusively on small, synthetic graphs
In summary, the initial validation shows promise but suffers from a significant gap between the synthetic test environment and the target application within FUM

## 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

### 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 $O(n^2)$ memory, where n is the number of nodes (concepts) in the UKG

### 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 $10^4$ 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 practical feasibility of this framework hinges critically on addressing these scaling challenges

## 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 ($O(n^3)$ or worse) with graph size is a primary limitation, potentially rendering the approach impractical for large FUM KGs without effective scaling strategies.
- Undirected Graph Assumption: The framework currently operates on an undirected representation of the UKG

### 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

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

## 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

## 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 $M_1$ and $M_2$—represent only one point in a wide design space for applying TDA to graphs

Table 1: Comparison of Selected TDA-Derived Metrics for Graph Analysis [?, ?]

| Metric/Representation | Information Captured | Interpretability | Computational Cost (Relative) | Po Ca |
|---|---|---|---|---|
| $M_2$ (Component Count) | Initial Fragmentation ($H_0$) | High | Low (Graph traversal) | Ba ity in ta |
| $M_1$ (Total B1 Persist.) | Global $H_1$ cycle complexity (sum of persistence) | Medium | High (PH calculation) | Ov of co te of |
| Betti Curves ($\beta_k$ vs. $\epsilon$) | Number of features ($\beta_k$) vs. filtration parameter | Medium | High (PH calculation) | Vi tu ev sc |
| PD Distances (Wass., Bottleneck) | Geometric distance between full PDs | Medium-Low | High (PH + Distance calc.) | Co all sin gr pl |
| Persistence Landscapes /Images | Vector space embeddings of PDs | Low | High (PH + Vectorization) | In fo els an cla |
| Persistent Entropy | Information-theoretic summary of PD point distribution | Medium-Low | High (PH + Entropy calc.) | Qu co ity of fea |

This table highlights that the proposed $M_1$ and $M_2$ prioritize simplicity and direct interpretability (especially $M_2$) 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)

## 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

- Characterize Learning Dynamics: Identify distinct topological signatures associated with different stages of learning or adaptation in FUM
  These exploratory avenues leverage TDA's strength in understanding the intrinsic "shape" of data

## 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
However, the framework faces significant challenges that must be overcome for practical deployment

i. 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
   The potential significance of a validated, scalable TDA framework for FUM is high
   Based on this assessment, the following recommendations are made:

   A. 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
      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

## References

## References

[1] Gemini 2.5 Pro. (2025, April 12). Technical Assessment based on user-provided documents regarding FUM TDA Framework. (Generated via conversational interaction).

[2] Lietz, J. (2025, April 2). Topological Data Analysis Framework for Knowledge Graph Health and Efficiency. *User Query Document*. Referenced as source 161 in original doc.

[3] Placeholder - Add Ref.

[4] Aktas et al. (2019). Explaining the Power of Topological Data Analysis in Graph Machine Learning. arXiv:2401.04250v1. https://arxiv.org/html/2401.04250v1

[5] Placeholder - ResearchGate Publication 378214130. Topological Data Analysis for Intelligent Systems and Applications.

[6] Huang, W., et al. (2023). A Survey on Topological Data Analysis in Machine Learning. Entropy, 25(2), 331. https://www.mdpi.com/1099-4300/25/2/331

[7] Placeholder - arXiv:2302.03836v1.

[8] Placeholder - arXiv:2411.10298v2.

[9] Placeholder - Add Ref.

[10] Placeholder - Add Ref.
[11] Otter, N., Porter, M. A., Tillmann, U., Grindrod, P., Harrington, H. A. (2017). A roadmap for the computation of persistent homology. EPJ Data Science, 6(17). https://www.math.ucla.edu/~mason/papers/roadmap-final.pdf
[12] Placeholder - arXiv:2406.04102v1. Chromatic Topological Data Analysis.
[13] Placeholder - Add Ref.
[14] Placeholder - Add Ref.
[15] Placeholder - Browse Summary: Persistent Homology for Network Type Distinction and Robustness Analysis.
[16] Placeholder - Add Ref.
[17] Placeholder - Add Ref.
[18] Placeholder - Georgia Tech TDA Projects Website. http://tda.gatech.edu/projects
[19] Placeholder - MathOverflow Question 141157.
[20] Cade, C., et al. (2023). Complexity-Theoretic Limitations on Quantum Algorithms for Topological Data Analysis. PRX Quantum, 4(4), 040349. https://link.aps.org/doi/10.1103/PRXQuantum.4.040349
[21] Placeholder - MathOverflow Question 288115.
[22] Placeholder - Add Ref.
[23] Placeholder - Add Ref.
[24] Placeholder - Add Ref.
[25] Placeholder - Add Ref.
[26] Placeholder - Add Ref.
[27] 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, 8(1), 355-371. https://direct.mit.edu/netn/article/8/1/355/118706/Altered-topological-structure-of-the-brain-white
[28] Placeholder - PathologyOutlines.com Topic: Informatics - Graph Neural Networks. https://www.pathologyoutlines.com/topic/informaticsgraphneuralnetworkstowholeslideimages.html
[29] Placeholder - Add Ref.
[30] Placeholder - Quantmetry Blog (2020). Topological Data Analysis with Mapper. https://www.quantmetry.com/blog/topological-data-analysis-with-mapper/
[31] Placeholder - arXiv:2405.04796. Persistent homology of featured time series data...
[32] Wasserman, L. (2016). Topological Data Analysis. Annual Review of Statistics and Its Application, 3, 501-528. https://www.math.uri.edu/~thoma/comp_top_2018/Wasserman_2016_TopologicalDataAnalysis.pdf
[33] Placeholder - Reddit r/datascience discussion (2019).
[34] Placeholder - MathStackExchange Question 3958944.
[35] Dey, T. K. Simplicial Complexes. Course Notes. https://www.cs.purdue.edu/homes/tamaldey/course/531/Simplicial-complex2.pdf

[36] MathStackExchange Question 3627924. Good Stopping Criteria for Persistent Homology. https://math.stackexchange.com/questions/3627924/good-stopping-criteria-for-persistent-homology

[37] Dow, E., et al. (2018). Weighted Persistent Homology. arXiv:1709.00097v2 [math.AT]. https://arxiv.org/pdf/1709.00097

[38] MathStackExchange Question 4701286. Using topological data analysis to analyze a graph. https://math.stackexchange.com/questions/4701286/using-topological-data-analysis-to-analyze-a-graph

[39] ResearchGate Figure 340439430. Applications of topological data analysis (TDA) to biology. https://www.researchgate.net/figure/Applications-of-topological-data-analysis-TDA-to-biology-A-Structural-biolog fig6_340439430

[40] Reference to source1.

[41] Wasserman, L. (2016). Topological Data Analysis. Annual Review of Statistics and Its Application, 3, 501-528. *(See Ref 31 in PDF).* https://www.math.uri.edu/~thoma/comp_top_2018/Wasserman_2016_TopologicalDataAnalysis.pdf

[42] Kerber, M. (2016). Persistent Homology: State of the art and challenges. Internationale Mathematische Nachrichten, 232, 1-18. *(See Ref 12 in PDF).* https://www.geometrie.tugraz.at/kerber/kerber_papers/imn_article_2016.pdf

[43] Aktas, E., Akbas, E., Aktas, M. E. (2019). Persistence Homology of Networks: Methods and Applications. Applied Network Science, 4(1), 1-26. *(See Ref - Unnumbered in PDF).* https://d-nb.info/1203059833/34

[44] Horak, D., Maletić, S., Rajković, M. (2009). Persistent homology of complex networks. Journal of Statistical Mechanics: Theory and Experiment, 2009(03), P03034. *(See Ref 14 in PDF).* https://www.researchgate.net/publication/23418949_Persistent_Homology_of_Complex_Networks

[45] Otter, N., Porter, M. A., Tillmann, U., Grindrod, P., Harrington, H. A. (2017). A roadmap for the computation of persistent homology. EPJ Data Science, 6(17). *(See Ref 10 in PDF).* https://www.math.ucla.edu/~mason/papers/roadmap-final.pdf

[46] Aktas, E., Akbas, E., Aktas, M. E. (2019). Persistent Homology for Network Comparison. arXiv:1907.08708 [cs.SI]. *(See Ref 2 in PDF).* https://arxiv.org/pdf/1907.08708

[47] Huang, W., Ribeiro, A. (2015). Persistent Homology Lower Bounds on High Order Network Distances. In 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP) (pp. 157-161). *(See Ref 16 in PDF).* https://www.seas.upenn.edu/~aribeiro/preprints/2015_huang_ribeiro_a.pdf

[48] Reininghaus, J., et al. (2015). A Stable Multi-Scale Kernel for Topological Data Analysis. In Proceedings of the IEEE Con-

ference on Computer Vision and Pattern Recognition (CVPR) (pp. 4741-4748). *(See Ref 13 in PDF)*. https://par.nsf.gov/servlets/purl/10180702

[49] Aksoy, S. G., et al. (2020). Graph Theoretical and Topological Data Analysis for Structure Identification in Social Networks for Bot Detection. Mathematics, 8(3), 58. *(See Ref 15 in PDF)*. https://www.mdpi.com/2297-8747/25/3/58

[50] El Gazzar, A., et al. (2023). Topological Data Analysis and Persistent Homology for Brain Network Analysis: A Narrative Review. Brain Sciences, 13(11), 1581. *(See Ref 22 in PDF)*. https://pmc.ncbi.nlm.nih.gov/articles/PMC10669999/

[51] CompTAG Tutorial. (2019). Rips Filtration Tutorial for the R-TDA Package. *(See Ref - Unnumbered in PDF)*. https://comptag.github.io/rpackage_tutorials//2019/07/tda-rips-tutorial.html

[52] Dow, E., et al. (2018). Weighted Persistent Homology. arXiv:1709.00097v2 [math.AT]. *(See Ref 36 in PDF)*. https://arxiv.org/pdf/1709.00097

[53] Cade, C., et al. (2023). Complexity-Theoretic Limitations on Quantum Algorithms for Topological Data Analysis. PRX Quantum, 4(4), 040349. *(See Ref 19 in PDF)*. https://link.aps.org/doi/10.1103/PRXQuantum.4.040349

[54] Amézquita, E. J., et al. (2020). Topological data analysis for discovery in preclinical spinal cord injury studies. Scientific Reports, 10(1), 16211. *(See Ref 23 in PDF)*. https://pmc.ncbi.nlm.nih.gov/articles/PMC7536308/

[55] Amézquita, E. J., et al. (2020). What is topological data analysis?. PLoS Computational Biology, 16(7), e1007859. *(See Ref 24 in PDF)*. https://pmc.ncbi.nlm.nih.gov/articles/PMC7383827/

[56] Huang, W., et al. (2023). A Survey on Topological Data Analysis in Machine Learning. Entropy, 25(2), 331. *(See Ref 5 in PDF)*. https://www.mdpi.com/1099-4300/25/2/331

[57] El Gazzar, A., et al. (2023). Topological Data Analysis for Multivariate Time Series Data with Applications. Entropy, 25(11), 1509. *(See Ref 28 in PDF)*. https://www.mdpi.com/1099-4300/25/11/1509

[58] Quantmetry Blog. (2020). Topological data analysis with Mapper. *(See Ref 29 in PDF)*. https://www.quantmetry.com/blog/topological-data-analysis-with-mapper/

[59] 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, 8(1), 355-371. *(See Ref 26 in PDF)*. https://direct.mit.edu/netn/article/8/1/355/118706/Altered-topological-structure-of-the-brain-white

[60] Chen, C., Paffenroth, R. (2022). Efficient Community Detection in Large-Scale Dynamic Networks Using Topological Data Analysis. arXiv:2204.03191 [cs.SI]. *(See Ref 25 in PDF)*. https://arxiv.org/abs/2204.03191

[61] Browse Summary: Knowledge Persistence (KP) Method for KG Completion Evaluation. *(See Ref 21 in PDF)*. (Refers to e.g., arXiv:2301.12929).

[62] Browse Summary: Alternative Filtration Methods for Persistent Homology on Networks. *(See Ref 2 in PDF)*. (Refers to various methods, e.g., weight rank homology).

[63] Browse Summary: Computational and Statistical Challenges in Persistent Homology. *(See Ref 10 in PDF)*. (Refers to e.g., Otter et al. 2017).

[64] Chazal, F. (2022). Introduction to Topological Data Analysis - Lecture 1. *(See Ref 8 in PDF). Accessed via spatstat.org.* https://spatstat.org/Aalborg2022/notes/TDA_Lecture1_part1.pdf

[65] ResearchGate Figure 340439430. *(See Ref 38 in PDF)*. (Various citations within figure description). https://www.researchgate.net/figure/Applications-of-topological-data-analysis-TDA-to-biology-A-Structural-biolog fig6_340439430

[66] MathOverflow Question 141157. *(See Ref 18 in PDF)*. Inference using Topological Data Analysis: Is it worth it for a regular statistician to learn TDA? https://mathoverflow.net/questions/141157/inference-using-topological-data-analysis-is-it-worth-it-for-a-regular-statisti

[67] Chen, C., Paffenroth, R. (2022). Efficient Community Detection in Large-Scale Dynamic Networks Using Topological Data Analysis. arXiv:2204.03191 [cs.SI]. *(See Ref 25 in PDF)*. (Note: Link in PDF seems slightly different title than text). https://arxiv.org/abs/2204.03191

[68] Yoo, J., et al. (2024). Altered topological structure of the brain white matter in maltreated children through topological data analysis. Network Neuroscience, 8(1), 355–371. *(See Ref 26 in PDF). Accessed via MIT Press Direct.* https://direct.mit.edu/netn/article/8/1/355/118706/Altered-topological-structure-of-the-brain-white (Note: Title slightly different in PDF link text).

[69] Quantmetry Blog (2020). Topological data analysis with Mapper. *(See Ref 29 in PDF)*. https://www.quantmetry.com/blog/topological-data-analysis-with-mapper/

[70] MathStackExchange Question 4701286. *(See Ref 37 in PDF)*. Using topological data analysis to analyze a graph. https://math.stackexchange.com/questions/4701286/using-topological-data-analysis-to-analyze-a-graph