TOPOQUERY PIPELINE

Explainable Query Recommendations via Persistent Homology

1

Runtime: **~2 minutes**

2

Code: **~1,400 lines**

3

Visualizations: **14 outputs**

4

Components: **7 modules**

5

INPUT
Palmer Penguins
Database

STEP 1

QUERY
GENERATION
31 SQL Queries

STEP 2

FEATURE
EXTRACTION
41 Dimensions

STEP 3

DISTANCE
COMPUTATION
Multi-Metric

STEP 4

PERSISTENT
HOMOLOGY
HOLE Library

333 observations
3 species, 7 features
3 islands

SELECT, WHERE
GROUP BY, JOINs
Aggregations

Structural patterns
Semantic features
Context metadata

Euclidean
Cosine
Mahalanobis

GUDHI backend
Dimensions 0 & 1
Stability analysis

STEP 5

TOPOLOGICAL DATA ANALYSIS
Core Computations

- Rips filtration across distance scales
- Connected component tracking
- Birth-death persistence pairing
- Feature stability & lifetime analysis
- Hierarchical cluster evolution
- Critical threshold identification
- Multi-scale pattern discovery

KEY FINDINGS & RESULTS

✓ 4 stable query clusters identified

✓ Feature lifetime: 2-5 distance units

✓ Maximum persistence: 4-10 (highly stable)

✓ Hierarchical structure (no loops detected)

✓ Multi-scale patterns confirmed

✓ Robust to noise and variations

✓ Explainable topological relationships

RECOMMENDATIONS
& VISUAL
EXPLANATIONS

PERSISTENCE
DIAGRAMS

Birth vs Death plots
Feature significance

PERSISTENCE
BARCODES

Topological proximity
Explainable results
Multi-scale insights

Feature lifetimes
Stability bars

PCA/MDS
PROJECTIONS

2D embeddings
Cluster visualization

DISTANCE
HEATMAPS

Proximity matrices
Relationship patterns

MULTI-SCALE ANALYSIS

Analyzes all distance thresholds
simultaneously, not just one

EXPLAINABLE RESULTS

Visual explanations via persistence
diagrams and projections

UNSUPERVISED METHOD

HOLE Library

No training data required
works on any SQL workload

TECHNOLOGY STACK

GUDHI • Python 3.9+ • NumPy • SciPy • scikit-learn
Matplotlib • Seaborn • NetworkX • SQLite