Persistent Homology Guided Force-Directed Graph Layouts

Ashley Suh, Mustafa Hajij, Bei Wang, Carlos Scheidegger & Paul Rosen
Presented by Javier Castillo-Arnemann
Background

- Graphs/networks are commonly used to encode relationships among entities, used for data exploration.
- Used to model social networks, digital networks, biological interactions, etc.
- Their abstractness makes them difficult to analyze.
- A good graph vis should present structure quickly and clearly, and support further investigation of the data.
Background

- Force-directed/spring-mass layouts are the most popular graph layout for interactive applications.
- Good at showing topological structures in sparse graphs, but as graphs get bigger unrelated structures overlap.
Approach

- **What?** Improve force-directed graph layout visualization.
- **Why?** Identify, compare and separate important topological structures.
- **How?** Use persistence homology features to enable interactive manipulation of the graph layout, visually separating PH features.
Approach

1. Embed a **weighted undirected graph in metric space** by inducing a distance between all nodes.
2. Extract **Persistence Homology (PH) features** of the metric space structure and sort them based on persistence.
3. PH features can be used either to **contract the nodes that created the feature**, or to **repulse the graph into two subsets**, depending on user input.
Approach

1. Embed an undirected graph in metric space by inducing a distance between all nodes.
   - The inverse of the edge weight is the distance, and the shortest path between two nodes is the distance between them.
   - The rest of the algorithm operates in this metric space.

If the graph is unweighted, the Jaccard index, edge centrality or other measures can be computed and used as weights.
Approach

2. Extract **Persistence Homology (PH) features** of the metric space structure and sort them based on persistence.

- The persistence homology of the graph can be calculated by finding its Minimum Spanning Tree (MST).
- Subset of edges that connects every node without cycles and minimizing the edge weights.
Approach

2. Extract **Persistence Homology (PH) features** of the metric space structure and sort them based on persistence.

- Every node corresponds to a point in the metric space.
- Consider the set of balls centered at every point with a diameter $t$ (metric space resolution).
- Every time two balls (components) merge as $t$ grows, one of the components disappear and its “time of death” $t$ is recorded.
- The lifetime of a component is its **persistence** $(0, \ 1 / w)$.

![Fig. 2](image-url)
2. Extract **Persistence Homology (PH) features** of the metric space structure and sort them based on persistence.

Every PH feature is associated with:

- **Persistence** measure $1/w$
- **Cause of death** $u$ and $v$
- **Subsets of nodes** $V_u$ and $V_v$, the sets of connected nodes after removing the feature from the MST.
- **Subset ratio** $|V_u| : |V_v|$, measures the centrality of the feature within the MST.
Approach

3. PH features can be used either to **contract the nodes that created the feature**, or to **repulse the graph into two subsets**, depending on user input.

- Video Demo

![Diagram](Fig. 6)
Comparison

- Comparing the PH of the traditional force-directed source layout and the user-selected target layout.
- PH calculated on Euclidean distance between the nodes, not topological distance.
Discussion

Strengths:

- Real-time selection of the PH features instead of choosing number of clusters beforehand.
- Combination of repulsive and attractive forces allows better separation than either force alone.
- Can be extended to other force-directed layouts.
- Performance comparable to traditional force-directed layouts.

Weaknesses:

- Shallow explanation of the modifications to the force-directed algorithm.
- Performance with unweighted graphs. Extra computations for weights?
- Can this approach be used with other community detection/network clustering methods instead of PH features?
- What happens when the number of PH features is too high to make the barcode impractical?