Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Topological Analysis of Graphs

CHALLENGE

Cyber graphs are constantly evolving while computers and other components come on and offline, and communications occur sporadically. We believe that tracking evolution from a topological perspective, including higher dimensional structures, rather than just a one-dimensional graph point of view, will allow us to learn behaviors within the network. We hypothesize that different states and behaviors of dynamic cyber networks can be identified using methods from topological data analysis, coupled with novel forms of graph statistics. From this approach, cyber system administrators will be able to tlearn more about how their system is evolving.

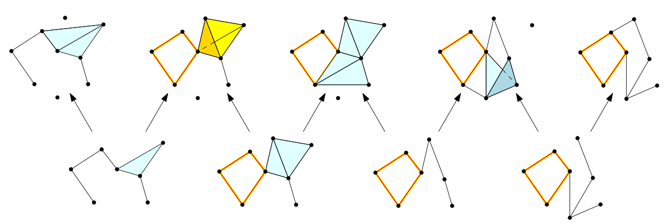
CURRENT PRACTICE

Current techniques in graph analysis of cyber networks are limited to the exploration of typical graph metrics (e.g., centrality, degree distribution); topological information in the graphs is not being utilized, though is being applied elsewhere e.g. biological systems, mobile sensor networks, signal processing, and medical image analysis. We seek to discover the topological properties of cyber graphs and use them to gain information about the state of the cyber system being modeled.

APPROACH

There are three major techniques in our approach: persistent homology, manifold learning, and Hodge theory.

Using proven tools from algebraic topology to study the evolution of cyber network structure



Evolution of a graph showing a persistent cycle

1. In persistent homology, topological features in data are identified which "persist" across different scalar resolutions, and these features are deemed more representative of the data, while those which occur only at a few levels of resolution are more likely to be noise. We are applying this tool to cyber networks to discover different types of evolution in the graph structure.
2. The topological technique we will apply is manifold learning. A graph on n vertices and m edges has $n+m$ pieces of information describing it. In the case of cyber graphs this can easily be in the millions or

U.S. DEPARTMENT OF ENERGY

billions, which can be difficult for any analyst to explore. Our conjecture is that these complex cyber graphs actually have fewer degrees of freedom, and thus they can be represented in a lower dimensional space, through manifold learning. We are experimenting with graph spectra and other novel graph metrics to embed a graph into a lower dimensional space.

3. While topological persistence is focused on structural questions, Hodge theory was developed to understand the structure of vector fields and flow data. We propose to merge the notions of persistent homology and Hodge theory so that we can better understand what, if anything is structural in network flow. This should assist in the detection of network state change, as a persistent Hodge theory would be sensitive to small changes in the source/sink structure of network flow.

Our approach brings established mathematical tools to bear in a new domain. Persistent homology has had widespread success in the area of topological data analysis of point cloud data, e.g., materials modeling. Additionally, researchers in other fields, e.g., medical imaging, have had success using manifold representations of data in order to find average behavior in static objects. In this project, we will bring these techniques to dynamic objects. We believe that extending these types of analysis to evolving graph data can have a similar level of success in learning core behavior of cyber networks.

IMPACT

The impact of this work will be seen through a deeper understanding of the evolution of cyber systems through a variety of topological techniques. System administrators are in need of additional tools that identify anomalies in their systems and are able to get local information about the problem. In this project, we give a different viewpoint from the theory of topological spaces which allows us to get a more robust picture of the system evolution and identify anomalies. This work is relevant to any sponsor working in the cyber domain, e.g., next-generation smart grid technologies in the DOE.

Contact

Emilie Purvine
Principal Investigator
(509) 375-3819
emilie.purvine@pnnl.gov

Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by **Battelle** *Since 1965*