Euler Characteristic Transform (ECT) of Embedded Graphs

I. Collaborators

- Dr. Liz Munch, CMSE & Math, MSU
- Elena Wang, CMSE Graduate Student, MSU

II. Basic idea

At this point, it is cliche to point out that data science intersects all aspects of our lives, online or not. While often this is interpreted as a need to build analysis methods for *big* data, we also need methods for analyzing *complex* data. Where the former is concerned with cases where we have terabytes upon terabytes of input, the latter is focused on finding and quantifying interesting structure in data. The field of topological data analysis (TDA) [8] takes the second vantage point by providing quantifiable, comparable, robust, and concise summaries of the shape of data. This may mean measuring shape in an obvious shape-focused context such as quantifying shape and structure in plant leaves, or could mean understanding high-dimensional point clouds by providing lower dimensional skeletons of the underlying embedded structures.

In this research project, we will focus on graphical signatures of data, specifically embedded graphs. A graph is a combinatorial object consisting of vertices and edges, however we can also view it as a continuous topological space, particularly when we are concerned with drawings of the graph. That is, a drawing of a graph G is a continuous map $\Phi : G \to \mathbb{R}^2$.

Our goal is to find methods to compare two graph embeddings [2]. This is particularly useful in applications such as comparison of geographic networks such as rivers or roads. We are particularly interested in creating studying map reconstructions where, for example, we have a reconstructed road network given GPS trajectory data. The main idea is to provide quality guarantees for computed approximations of a ground truth map.

We will utilize the idea of the Euler Characteristic Transform (ECT) [10, 9] to provide a summary of one aspect of the shape of the data. The Euler characteristic for a simplicial complex (a generalization of a graph) is defined to be an alternating sum of the number of simplicies in each dimension,

$$\chi(K) = \#(\text{vertices}) - \#(\text{edges}) + \#(\text{faces}) - \cdots$$

By picking an orientation in the plane, $\omega \in S^1$, any drawing of a graph $\Phi : G \to \mathbb{R}^2$ can provide a single valued function for the graph $\Phi_\omega : G \to \mathbb{R}$ by looking at the height of the graph in that particular direction. The Euler Characteristic Curve (ECC) in that direction is a function giving the Euler characteristic for each sublevel set. Stacking this information together gives an ECC for each direction; data which is collectively known as the Euler Characteristic Transform (ECT) of the input shape. This construction has already proven to be useful in many settings, including for analyzing barley seed morphology [3], predicting clinical outcomes in glioblastoma [5], and shape classification [7]. We will investigate how the ECT can be used for comparison of two graph drawings, $\Phi : G \to \mathbb{R}^2$ and $\Psi : H \to \mathbb{R}^2$.

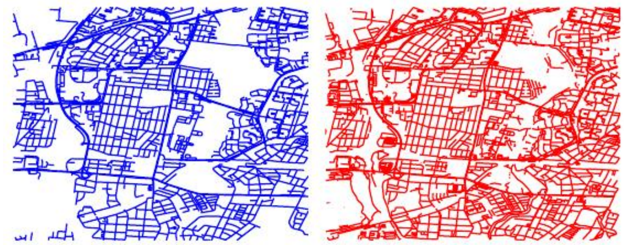

Fig. 1. Two reconstructed maps of Berlin. How do we compare the two graphs? From [1].

III. Semester research goals

- Understand the ECT construction.
- Compute ECTs for example graphs, by hand and/or by computer.
- Develop a comparison measure for two embedded graphs using this technique.
- Apply the results to toy data sets.

IV. Useful background knowledge

Note that while experience with any of these directions will be useful, they are not required as we can pick up what is needed as we go along.

- Graph theory/discrete math.
- Point-set topology.
- Some experience with python for computation.

V. Project plan

We will start by working with embedded graphs, which for now we will loosely define as 2-dimensional graphs with vertices and edges. Our goal is to analyse data in this form using Euler characteristic transform (ECT).

1) First, we need to get an understanding of basic graph theory. For what we need, likely Bondy Murty Ch 1-4 will suffice [4].
2) We will do a literature review and learn how ECT is used to represent geometric shapes in the context of TDA.
3) We will take a look at data sets such as the IAM Handwriting database, and binary graphs and their skeletons [6]. We aim to analyse the data using ECT and get output data by developing a comparison measure for two embedded graphs.
4) After having sufficient data, we will then use ML techniques such as SVMs to see if ECT is a good representation of embedded graphs.

Of course, this is research, so we have a lot of questions of how this procedure works. An example of some is below.

- How many or which direction(s) makes this comparison work well?
- How representative is the distance obtained from the comparison measure via ECT?
- How well can ML capture this information?

References

[1] Mahmuda Ahmed, Brittany Terese Fasy, and Carola Wenk. "Local persistent homology based distance between maps". In: *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM. 2014, pp. 43–52.

[2] Hugo A. Akitaya, Maike Buchin, Bernhard Kilgus, Stef Sijben, and Carola Wenk. "Distance Measures for Embedded Graphs". In: (Dec. 21, 2018). arXiv: `1812.09095 [cs.CG]`.

[3] Erik J Amézquita, Michelle Y Quigley, Tim Ophelders, Jacob B. Landis, Daniel Koenig, Elizabeth Munch, and Daniel H Chitwood. "Measuring hidden phenotype: Quantifying the shape of barley seeds using the Euler Characteristic Transform". In: *in silico Plants* (2021). DOI: `10.1093/insilicoplants/diab033`.

[4] John A Bondy and Uppaluri SR Murty. *Graph theory, volume 244 of Graduate Texts in Mathematics*. 2008.

[5] Andrew X. Chen Sayan Mukherjee Lorin Crawford Anthea Monod and Raúl Rabadán. "Predicting Clinical Outcomes in Glioblastoma: An Application of Topological and Functional Data Analysis". In: *Journal of the American Statistical Association* 115.531 (2020), pp. 1139–1150. DOI: `10.1080/01621459.2019.1671198`.

[6] Diego Alejandro Macrini. *Indexing and matching for view-based 3-d object recognition using shock graphs*. Citeseer, 2003.

[7] Lewis Marsh, Felix Y. Zhou, Xiao Qin, Xin Lu, Helen M. Byrne, and Heather A. Harrington. *Detecting Temporal shape changes with the Euler Characteristic Transform*. 2022.

[8] Elizabeth Munch. "A User's Guide to Topological Data Analysis". In: *Journal of Learning Analytics* 4.2 (2017). DOI: `10.18608/jla.2017.42.6`.

[9] Elizabeth Munch. *An Invitation to the Euler Characteristic Transform*. 2023. eprint: `arXiv:2310.10395`.

[10] K. Turner, S. Mukherjee, and D. M. Boyer. "Persistent homology transform for modeling shapes and surfaces". In: *Information and Inference* 3.4 (2014), pp. 310–344. DOI: `10.1093/imaiai/iau011`.