The topological shape of cellular patterns

Erik Amézquita

09/02/2024

- **Mentor:** Erik Amézquita
- **Affiliation:** Plant Science & Technology and Mathematics @ MU
- **Area:** Applied Mathematics (Topological Data Analysis [TDA])
- **Start time:** Fall 2024
- **Commitment:** Attend one-on-one regular weekly meetings, plus a weekly training meeting with BIPS fellows.
- **Duration:** The whole semester with possibilities to extend further depending on progress
- **Compensation:** Compensation through BIPS Broader Impacts Grant. Circa $12/hr — expected 10 hrs a week

Overview

Shape plays a fundamental role across all organisms at all observable levels. Molecules and proteins constantly fold and wrap into intricate designs inside cells. Cells arrange into elaborate motifs to form sophisticated tissues. Layers of different tissues come together to form delicate vascular systems that sustain leaves. Each of these tissues evolved as part of a distinct branch of the ever-growing tree of life. From micro-biology to macro-evolutionary scales, shape and its patterns are foundational to biology. Measuring and understanding the shape is key to extracting valuable information from data, and push further our insights.

We will study the potential of Topological Data Analysis (TDA) for plant shape quantification. TDA is a combination of different mathematical and computational results based on algebraic topology that seeks to describe concisely and comprehensively the shape of data in a general setting. In very succinct terms, TDA consists of two basic ingredients. First we think the data as a collection of points. Second, we define a notion of distance between every pair of points. The points could be atoms, biomolecultes, cells' nuclei, image pixels, or an organism itself. Distances between points could be the Euclidean, geodesic, genetic, or correlation-based. Once we have data points and distances, we merge systematically the points, starting with those that are closer to each other. The key idea is to keep track of distinct blobs, loops, and voids that form and disappear as we merge several points. This versatile idea is not constrained to a particular dimension or set of landmarks, which makes it ideal to compare a vast array of possible different shapes.

In particular, I am interested in understanding applications of TDA to measure spatial patterns to the apply them to plant cell data. This will require reading relevant work from the Porter Lab, mainly Feng and Porter (2021). In that paper, they provide ways to analyze the intrinsic shape of geospatial patterns of voting districts despite the heterogeneity of district sizes and shapes. Being able to parse through such size and shape heterogeneity will be key to study the shape of cellular patterns.

This would be a highly interdisciplinary research project, combining mathematics, programming, and biology. Once we have a basic grasp of TDA side of things, we will be involved in active discussion with plant biologists to truly understand the significance of our results and fuel future directions.

BIPS Funding

The research will be funded by the BioInformatics in Plant Sciences (BIPS) program. BIPS is an interdisciplinary undergraduate program where students from engineering and math work in pairs with plant biology students to work on a project that requires the expertise of both fields. From the start it was clear that students working in wet and dry labs needed to learn how to communicate with each other and this requires language fluency opportunities to understand the challenges and opportunities of both fields.

BIPS students work on their projects ~10hrs/week and are required to attend a weekly meeting coordinated by an MU graduate student (BIPS coordinator) who meets with students to discuss their projects, find solutions and alternatives to bottlenecks within their projects, hear about faculty conducting inter-disciplinary research, career opportunities, demos of new of software for data mining, phenotyping, genomics and to prepare posters and talks for presentations.

BIPS students will be compensated above $12/hr. US citizenship is **not** required. Funding can extend beyond spring semester depending on progress. BIPS is currently administered by Dr. David Mendoza-Cozatl from the Division of Plant Sciences & Technology.

Profile of student

The student should feel comfortable with some linear algebra and calculus at the bare minimum. Some knowledge on the basics of group theory will be useful to get a quick grasp of the TDA fundamentals. Basic coding skills (e.g. feel comfortable using `numpy` in python) would be very helpful but not required.

The exact path of the research project will vary depending on the student's strengths.

The student should be open to discuss their knowledge with biologists and be receptive of biology input when it comes to data analysis and interpretation.

References

- Feng, M., Porter, M.A. (2020) Spatial applications of topological data analysis: Cities, snowflakes, random structures, and spiders spinning under the influence. *Phys. Rev. Res.* 2: 033426. DOI: 10.1103/PhysRevResearch.2.033426
- Feng, M., Porter, M.A. (2021) Persistent Homology of Geospatial Data: A Case Study with Voting. *SIAM Review* 63(1): 67–99. DOI: 10.1137/19M1241519
- Feng, M., Hickok, A., Porter, M.A. (2022) Topological Data Analysis of Spatial Systems. In *Higher-Order Systems*. Springer: Cham. p. 389-399. DOI: 10.1007/978-3-030-91374-8_16
- Hickok, A., Needell, D. Porter, M.A. (2022) Analysis of Spatial and Spatiotemporal Anomalies Using Persistent Homology: Case Studies with COVID-19 Data. *SIAM Journal on Mathematics of Data Science* 4(3): 1116–1144. DOI: 10.1137/21M1435033