Persistent Homology for Data of Leaf Morphology

Yuqing Xing

Adviser: Zhen Huan

September 12, 2021

Outline

Outline

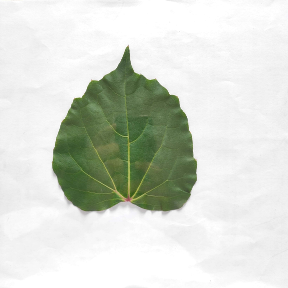
Motivation

In order to classify plants, the study of plant leaves is a direct and effective method. Characteristics such as leaf venation, leaf contour and leaf color can be used to identify different leaves. **How do we classify leaves of plants according to their topological characteristics?**

Leaf venation is the distribution form of leaf veins on leaves. **Leaf contour** characterizes the shape of the leaf. Both of them have fine topology and geometry structures, providing important bases for plant classification.

Leaf morphology






Background

In 2017, Li M, Frank MH, Coneva V, Mio W, Chitwood DH, Topp CN presented a morphometric technique based on topology, using a persistent homology framework, to measure the morphology of leaves and classify them by plant family. For a given topological feature, PH measures the persistence of components of the feature across the scales of a filtration set, then creates persistence barcodes that can be used to compare the overall topological similarity between objects.

Combined with analysis methods (e.g., canonical variant analysis (CVA), principal component analysis(PCA)), persistent homology can be used to classify the morphology of plants. Topological features provide a framework to quantify leaves of plants comprehensively.

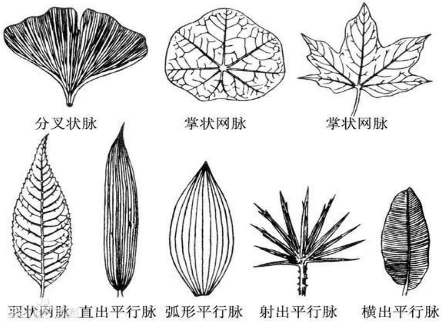
Leaf venation

Study on the topological structure of leaf venation is not ample so far. In our research, we mainly use topological data analysis method to mine the topological characteristics of leaf venation. We expect to combine the topological features of leaf venation with the topological features of leaf contours to distinguish different leaves topologically.

Three main types of leaf venation:
(1) reticular venation (2) parallel venation (3) bifurcated venation

Reticular venation

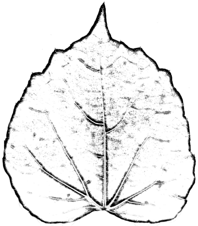
Reticular venation can be divided into pinnate venation and palmate venation.

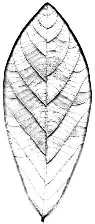


Two examples: ivy and Bridelia monoica

The venation of ivy have palmate vein structure and the venation of Bridelia monoica have pinnate vein structure. We use the PH method to analyze the venation of ivy leaves and Bridelia monoica leaves. Here, persistent homology method for leaf venation can be based on **the height function** / **the geodesic distance function**, which captures the height / the length of the shortest path from veination points to the base of leaves.

Two examples: ivy and Bridelia monoica





Outline

Introduction of persistent homology

Let S be a finite set of points, r be a positive real number, define the **Rips complex** $R(r) = \{\sigma \subset S \mid diam\, \sigma \leq r\}$, where $diam\, \sigma$ is the diameter of σ, i.e. the maxiamal distance between any two points of σ.

A filtration is a finite increasing sequence of sets $K = \{K_i \mid K_i \subset K_j \quad for\, i < j,\; i, j \in R\}$.

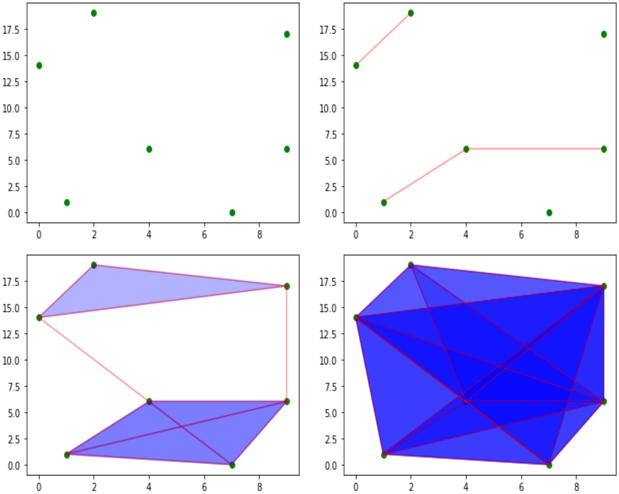
Introduction of persistent homology

Ways of forming filtrations:

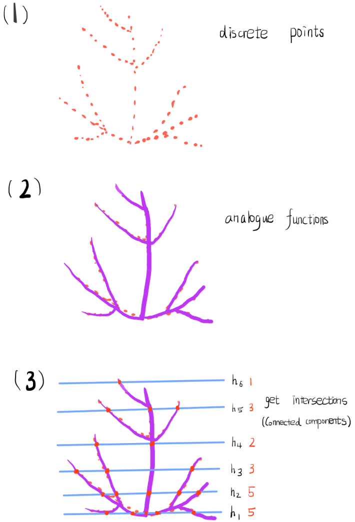
(1) The input is a finite set of points (a point cloud) embedded in Euclidean space or other metric space. For any real number $r > 0$, consider the union of all spheres of radius r centered at fixed points in the point cloud. As the radius r increases, the union of balls provides a filtration.

(2) The input is a real valued function on space X. For instance, X can be a Euclidean space. Considering the superlevel set $\{x \in X \mid f(x) \geq r\}$, the filtration is obtained as the threshold r decreases.

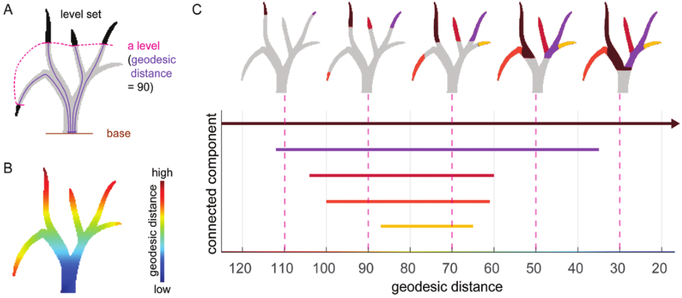
Forming filtrations



Forming filtrations: height function



Forming filtrations: geodesic distance function

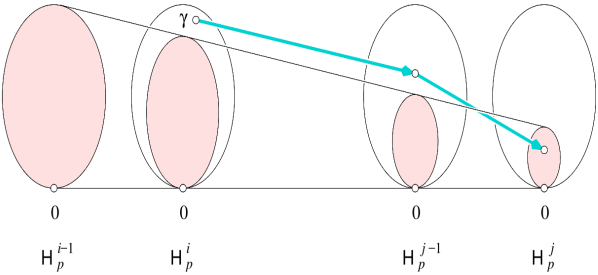


Persistent homology group

The inclusion $f^{i,j} : K_i \to K_j$ induces a homomorphism $f_p^{i,j} : H_p(K_i) \to H_p(K_j)$. The p-th persistent homology group $H_p^{i,j}$ is defined as the image of $f_p^{i,j}$. The p-th persistent betti number is defined as the rank of $H_p^{i,j}$.

Let be a class in $H_p(K_i)$, γ is born at K_i if $\gamma \notin H_p^{i-1,i}$; furthermore, γ dies entering at $K_j (j > i)$ if $f_p^{i,j-1}(\gamma) \notin H_p^{i-1,j-1}$ and $f_p^{i,j}(\gamma) \in H_p^{i-1,j}$. If homology class γ is born at K_i and dies entering K_j , define the persistence of γ as $j - i$.

Persistent homology group

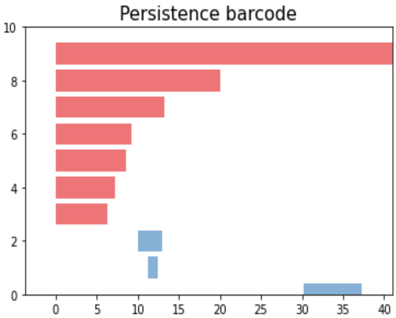


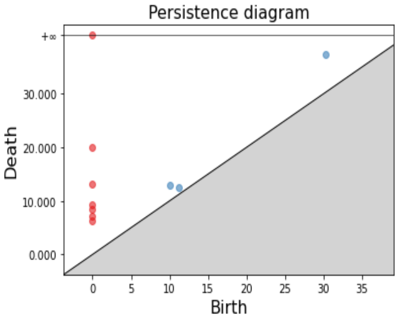
Persistence barcode/Persistence diagram

Persistent homology can be represented in two equivalent forms: either as a **persistence barcode** or as a **persistence diagram**;

A persistence barcode is a collection of bars whose left end points and right end points are birth and death respectively. Each interval in the persistence barcode is represented in the persistence diagram by a point in the plane, with its **birth coordinate** on the horizontal axis and with its **death coordinate** on the vertical axis. Points of a persistence diagram all lie above the diagonal line $y = x$.

Persistence barcode/Persistence diagram





Bottleneck distance

A persistence diagram is made up of 2-dim points. **Bottleneck distance** can measure the similarity of two persistent diagrams. The bottleneck distance between digram A and digram B is defined as follows:

$$d(A, B) = \inf_{\alpha} \sup_{x} \mid x - \alpha(x) \mid$$

where $x \in A, u = (x, y), |u| = max\{|x|, |y|\}, \alpha$ traverses all the bijections from A to B.

The bottleneck distance is a robust metric of similarity between two leaf structures. The more similar the diagrams are, the smaller it is.

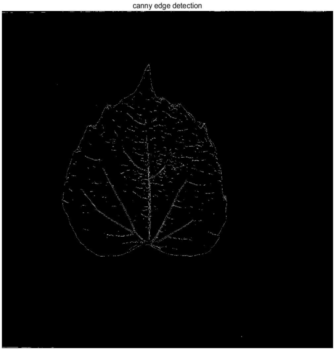
Outline

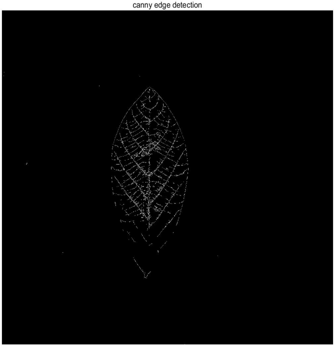
Steps of the PH method

Steps of the PH method

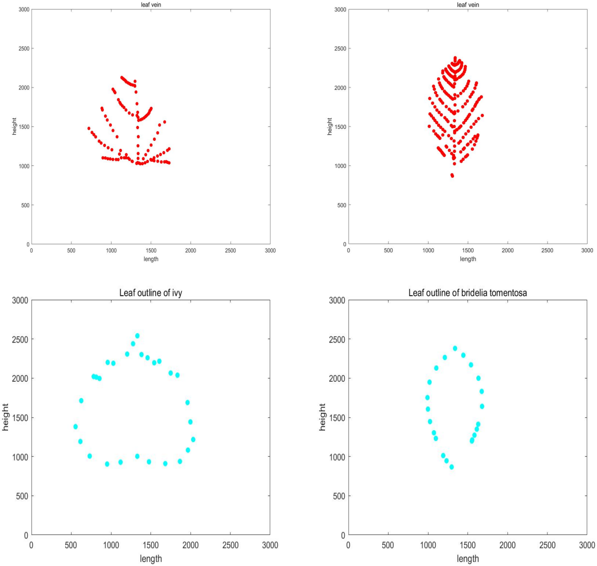
1. Get two-dimensional leave images
2. Extracting point cloud datasets from leave images
3. Constructing filtrations
4. Computing persistent homology(e.g., persistent barcode, bottleneck distance, Betti number, Euler characteristic curve)
5. Anayzling persisting homology results(e.g.,PCA, Factor Analysis, Cluster Analysis)

Image extraction of leaf venations and contours

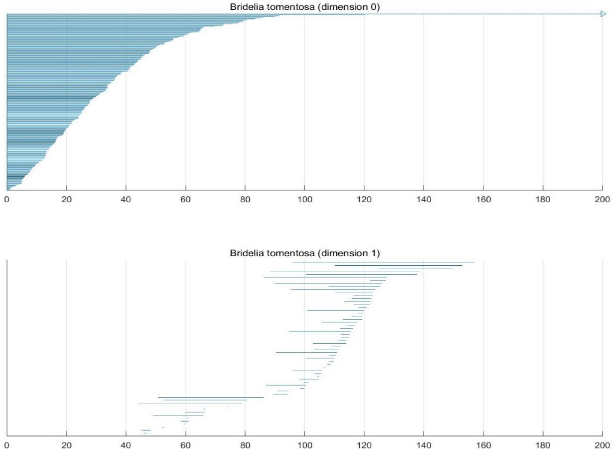




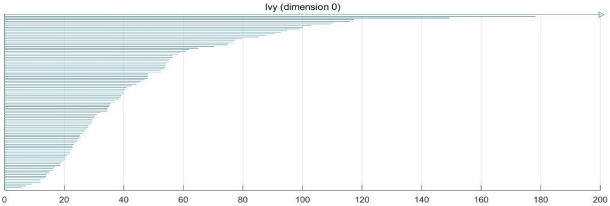
Point clouds of venations and contours

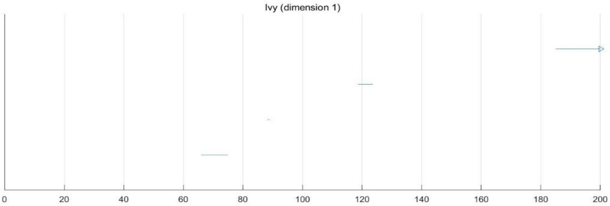


Barcodes: leaf venation of Bridelia monoica



Barcodes: leaf venation of ivy





More analysis are needed

At present, the analysis of the topological characteristics of plant leaves is preliminary, and more analysis needs to be done. And I have some problems:
1. How to program the filter constructed by geodesic distance?
2. All points in the leaves were treated equally, and persistent spectral graph or weighted persistent homology was not used.
3. How to add the geometric information or non-geometric information in the blade data into the topological invariants?
4. Ways to select leaf data is various.

References

M Li, LL Klein, KE Duncan, N Jiang, CN Topp. Characterizing 3D inflorescence architecture in grapevine using X-ray imaging and advanced morphometrics: implications for understanding cluster density, Journal of Experimental Botany, Volume 70, Issue 21, 1 November 2019, Pages 6261–6276;

Mao L, Duncan K, Topp C N, et al. Persistent homology and the branching topologies of plants[J]. American Journal of Botany, 2017, 104(3).

Adams H, Moy M. Topology Applied to Machine Learning: From Global to Local[J]. 2021.
Rote G , Vegter G . Computational Topology: An Introduction[J]. Springer Berlin Heidelberg, 2006.

References

Li M, Frank MH, Coneva V, Mio W, Chitwood DH, Topp CN. The Persistent Homology Mathematical Framework Provides Enhanced Genotype-to-Phenotype Associations for Plant Morphology, Plant Physiology Aug 2018, 177 (4) 1382-1395;

Edelsbrunner H, Harer J. Persistent homology—a survey[M]. 2008.