The connectivity threshold on random graphs

Daniel Etrata
Mentors: Benjamin Antieau and Samuel Cole
Mathematical Computing Laboratory
UIC

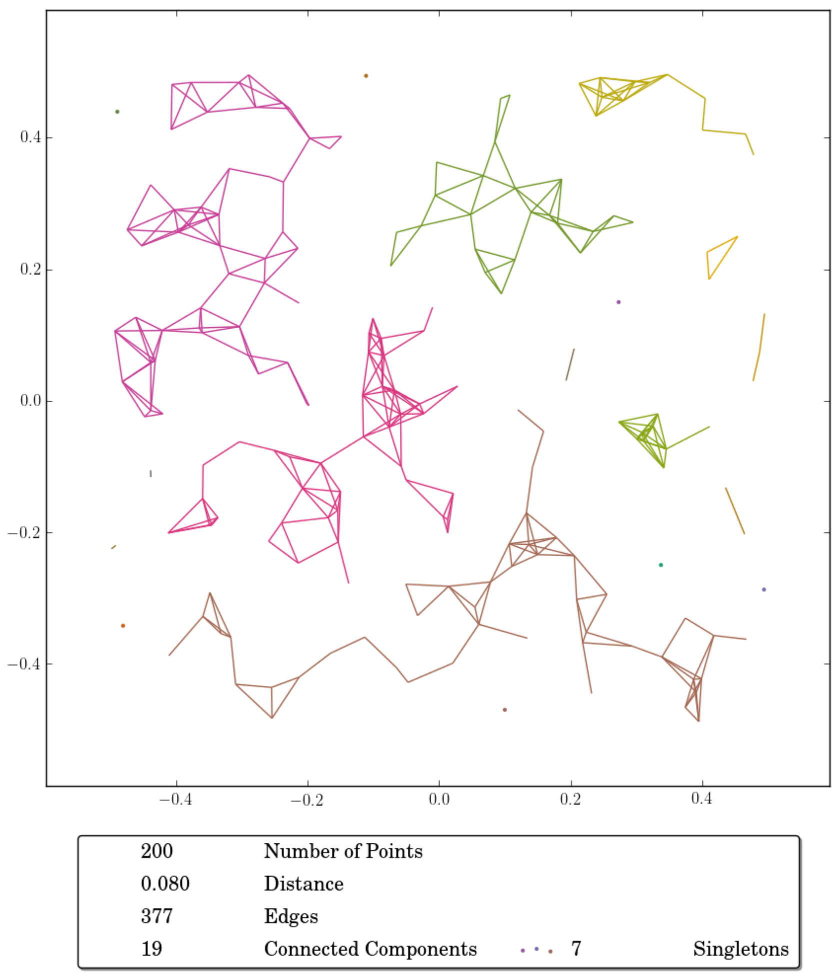
Summary

This semester's research focused on determining to what extent we can use properties of random graphs to study points on algebraic varieties. Along the way, we developed a series of tools in Python called `persispy` to understand the subject material.

Motivation

In higher dimension and with multiple equations, it becomes very difficult to understand basic properties of their set of solutions. We are developing `persispy` to study the approximate solution sets of algebraic varieties. We should say that this poster shows solutions in the reals for the sake of plotting points. Additionally, our tool can also return the set of solutions in the complex, which is nice when an algebraic variety only has complex solutions.

Neighborhood graphs in the plane



Here is an example of a neighborhood graph in the plane whose distance is in the critical regime. We construct each component by choosing a random vertex, running a depth first search, and visiting all the neighbors.

The connectivity property and its threshold

Let n be the number of points, ε be the maximum edge between points, $G(n)$ be a random graph, and P be a general property of the graph. Then $G(n)$ has P a. s. as $n \to \infty$. It turns out when we define $\varepsilon(n)$, we can find a threshold function for that property, here denoted as ε_P. Then, $G(n)$ has P a. s. if $\varepsilon \gg \varepsilon_P$. The complement is that $G(n)$ does *not* have P a. s. if $\varepsilon \ll \varepsilon_P$.

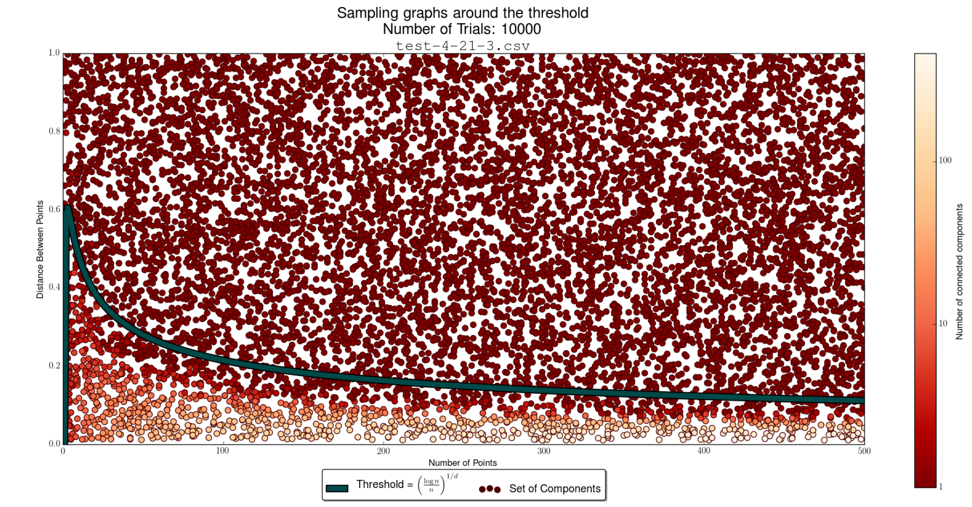
In particular, we are interested in the connectivity property, where given a point, there is a path to another point in the graph. We want to choose a *good* n and ε such that all vertices are connected and topological features are not obscured. For points in the plane, then the threshold is given when

$$\varepsilon(n) = \left(\frac{\log(n)}{n}\right)^{\frac{1}{2}}, \tag{1}$$

then a.s. $G(n,\varepsilon)$ is connected [1]. We can also generalize the threshold to the d-dimensional box. Let d be the dimension and $c \in \mathbb{R}$. Then the threshold is

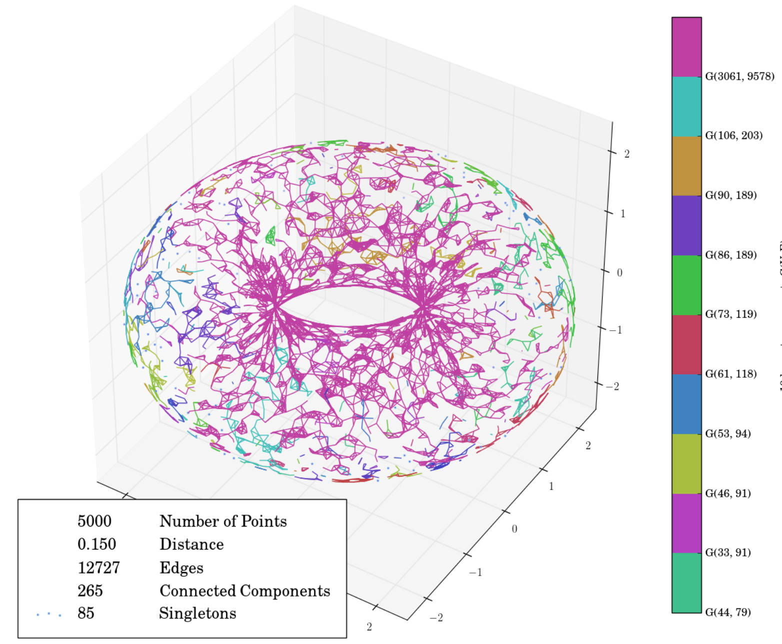
$$\varepsilon(n) = \left(\frac{\log(n)\cdot c}{n}\right)^{\frac{1}{d}}.$$

Threshold graphs



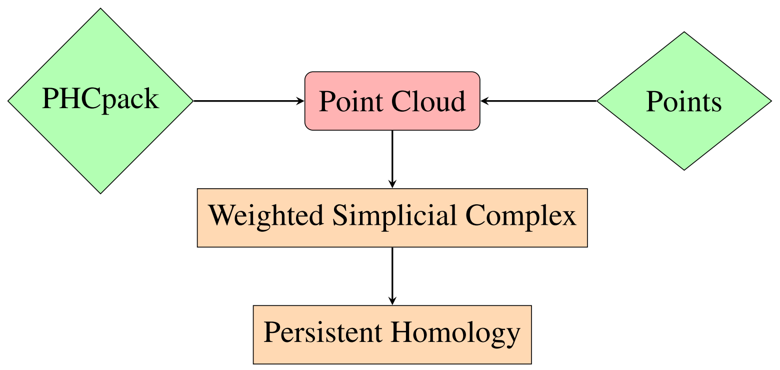
We use a Monte Carlo method to sample graphs around the threshold. We take a random n and ε and return the neighborhood graph. Then, we count the number of connected components. We also plot the super critical threshold from Equation 1 for comparison. Notice points around and above the threshold are dark red and therefore either have very few components or are completely connected.

Large connected components



Here is a neighborhood graph of 5000 random points with $\varepsilon = 0.15$ on the torus. In the picture, there are 250 components, 86 of which are singletons. The largest component has 3061 vertices and 9578 edges.

`persispy` Tool Chain



There are a few tools to `persispy`. In short detail, we use the `phc` or `points` point cloud process to return a point cloud from `point_cloud`. From here, we can get a neighborhood graph from `weighted_simplicial_complex`. We can then filter through the resulting simplicial complexes in `persistent_homology`.

This poster focuses on the results of `point_cloud` and `weighted_simplicial_complex` modules.

Citation

Bobrowski, O., and Kahle, M., *Topology of Random Geometric Complexes: A Survey.* September 17, 2014.