AMAT 584 Homework 2

Mathew Goberdhan (mgoberdhan@albany.edu)

May 10, 2021

1. Problem 1

What is the expected number of vertices of $S(n,p)$?

$S(n,p)$ is the set of vertices that originate from L, which is defined as

$$L = \left\{ i \begin{bmatrix} 1 \\ 0 \end{bmatrix} + j \begin{bmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix} : i, j \in \{0, ..., n-1\} \right\}$$

Then, $S(n,p)$ can have at most n^2 vertices. In fact, $S(n,p)$ has n^2 vertices when $p = 1$ and 0 vertices when $p = 0$. For $0 \leq p \leq 1$, our hypothesis is $S(n,p)$ is expected to have n^2p vertices, and we test with $p = 0.9$, while n tending to large numbers. Using our function $S(n,p) =$ *sitePercolationGraph(n,p)*, with $p = 0.9$, we test n as it approaches 2000, and compare n to ratio $\frac{card\left(S(n,p)\right)}{n^2p}$ via scatter plot. We want to observe the ratio converging to 1 as n becomes larger. We obtain the following results in Figure 1:

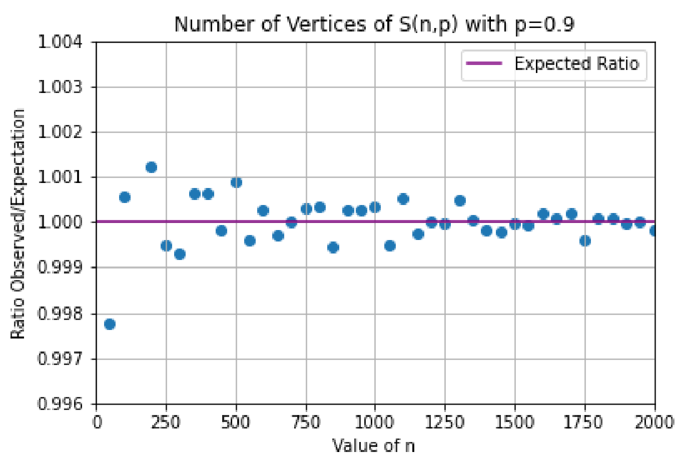

Figure 1: Test on 'Number Of Vertices' with parameter $p = 0.9$.

We observe as n becomes larger, the cardinality of $S(n, 0.9)$ tends to $0.9n^2$. Repeated for any $0 \leq p \leq 1$, we observe the same trend. Therefore, we expect the cardinality of $S(n,p)$ to be n^2p.

2. Problem 2

Let $S(n,p)$ denote a (random) sit percolation graph with parameters n and p and let $C(n,p)$ be the largest connected component of $S(n,p)$. We call a random polymer $C(n,p)$ a site percolation polymer. Using *SitePercolationGraph.ipynb* and an appropriate function from the NetworkX library, write a function to sample a random polymer $C(n,p)$.

a) Create a figure showing how the proportion of the vertices of $S(n,p)$ in its largest connected component changes with p. Sample polymers $C(400,p)$ for $p \in \{0.40, 0.42, ..., 0.60\}$ and plot p against $|C(400,p)|/|S(400,p)|$.

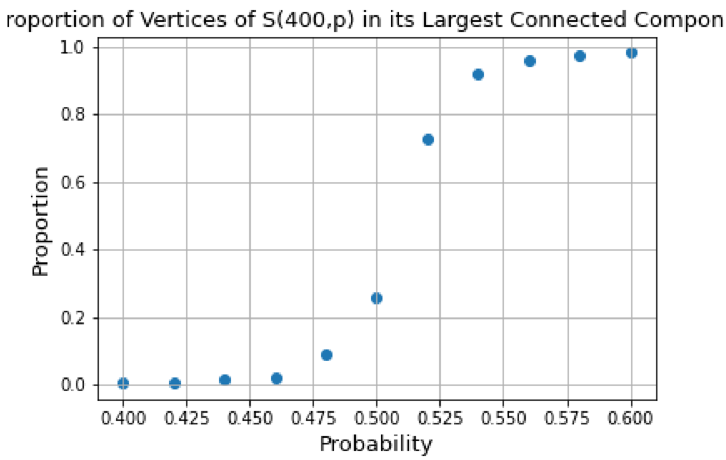

Figure 2: Proportion of vertices with varying p-value.

b) Interpret the figure you generated in the previous plot. What is the significance of $p = \frac{1}{2}$?

The behavior of this curve is that of logistical regression. In this plot, $p = \frac{1}{2}$ represents the point of inflection for this curve. The significance is that when p finally exceeds $\frac{1}{2}$, the proportion of vertices will automatically jump closer to 1. Values of $p > \frac{1}{2}$ statistically will give you denser polymers which lack significant structure (holes/gulfs), whereas when $p < \frac{1}{2}$, polymer size is sometimes too small where significant structure disappears for the overall polymer.

3. Problem 3

Write code to generate an image of a site percolation polymer by drawing disks of radius $\frac{1}{2}$ centered at a collection of points. Make sure you circles appear as circles rather than as ellipses, and that the aspect ratio is set to 'equal'.

a) Find an example polymer $C(n, 1/2)$ with about 1000 disks and generate a figure showing this polymer.

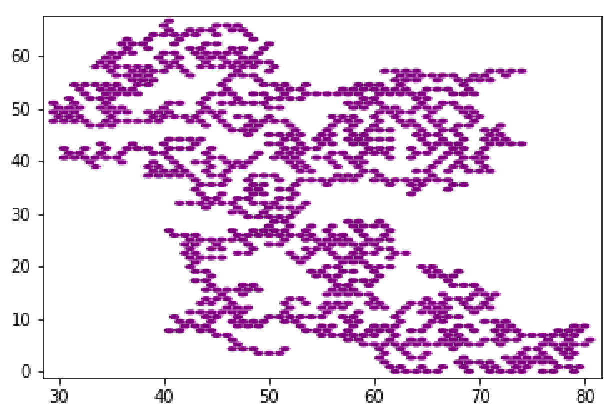

Figure 3: n=80, p=$\frac{1}{2}$, and 1009 disks.

b) Create figures showing example polymers $P5$ from $C(n, 1/2)$ and $P6$ from $C(n, 3/5)$ that have around 100,000 disks each (±10,000 say).

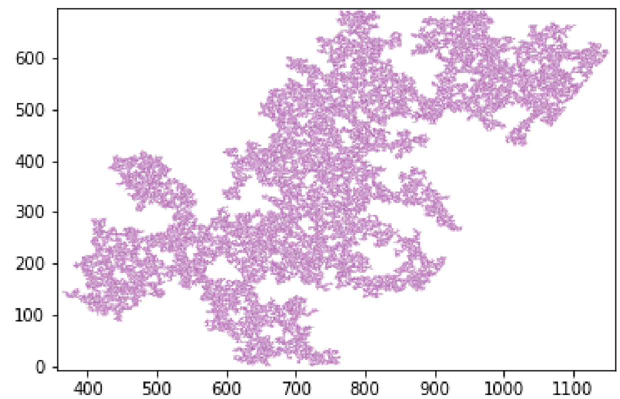

Figure 4: P_5 | n=800 and 101,935 disks |

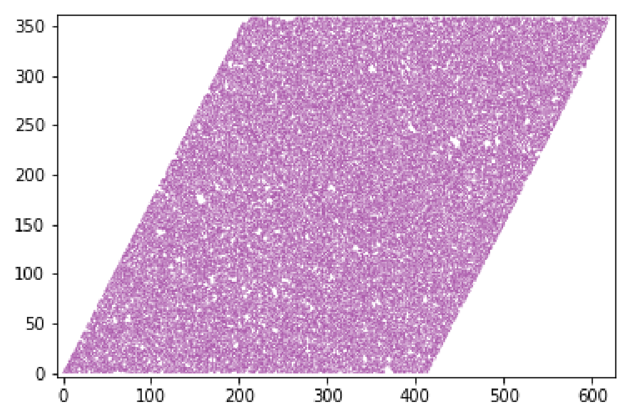

Figure 5: P_6 | with n=415 and 101,537 disks |

c) Comment on the visual differences between P_5 and P_6.

When observing P_5, we can see significant gulfs and holes in its structure. Specifically, we see some particular gulfs having potential to form into holes if you step up p minimally. If we observe P_6, we don't see any significance. Visually, all gulfs and holes are small enough where any minimal increase in p will have most of the structures disappearing. This dramatic change from $p = \frac{1}{2}$ to $p = \frac{3}{5}$ can be observed from the scatter plot in Problem 2A.

Additionally, it is obvious P_6 has self-similarity. There are infinite parts of the polymer that is similar to the whole polymer. When you analyze the structure of P_5, however, it appears the is no self-similarity, strictly speaking. If we are loose about 'similar,' then it is possible there is one piece of P_5 that could make it self-similar, which is outlined below.

Lastly, the diameter of P_5 exceeds that of the diameter of P_6. This observation makes sense due to the density of the points in P_6. You would also expect a larger diameter of P_5 due to the multiple gulfs and holes of significant size.

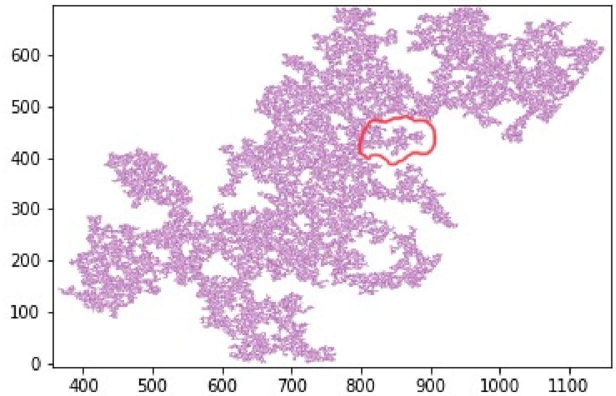

Figure 6: P_5. Potential self-similarity due to outlined piece of the whole polymer.

4. Problem 4

Use the function *polymerPH1(points)* in *SitePercolationGraph.ipynb* to compute the one-dimensional persistent homology of the polymers P_5 and P_6.

a) Create scatter plot of the data (birth, death) for the one-dimensional persistent homology of P_5 and P_6.

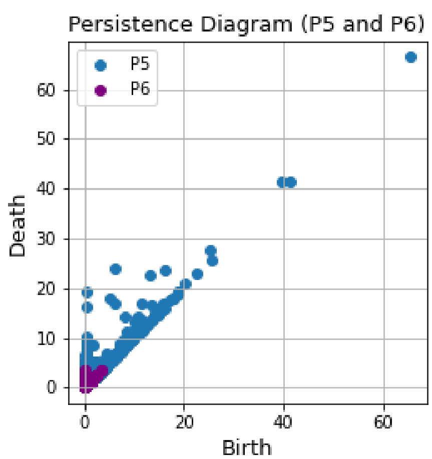

Figure 7: Persistence Diagram for P_5 and P_6 which shows significant differences in polymer structure.

.

b) Compute the five most common intervals (birth, death) for each of P_5 and P_6.

Most Common Intervals (b, d) of P_5 and P_6			
Intervals of P_5	Frequency (P_5)	Intervals of P_6	Frequency (P_6)
(0, 0.07735027)	54,683	(0, 0.07735027)	73,475
(0.3660254, 0.5)	20,687	(0.3660254, 0.5)	22,741
(0, 0.5)	2,900	(0.0, 0.5)	5,668
(0.82287566, 0.84715063)	2,111	(0.3660254, 0.65470054)	1,926
(0.3660254, 0.65470054)	1,884	(0.5, 0.65470054)	1,161

Table 1: Most common birth/death intervals for P_5 and P_6 and their frequency. Some interesting trends are present in the data as p increases amongst the intervals in common with both polymers.

The proportion of total intervals of P_5 and P_6 that fall into the five most common cases are 0.88748 and 0.93888, respectively. You can observe these results in the scatter plot above, as majority of the coordinates are contained closer to the origin. With larger p, we will begin to expect a higher frequency for (0, 0.07735027) and proportion of total intervals to most common cases to tend closer to 1.

Additionally, the following geometric configurations are for the first and third most common intervals of P_6 when radius starts at $\frac{1}{2}$.

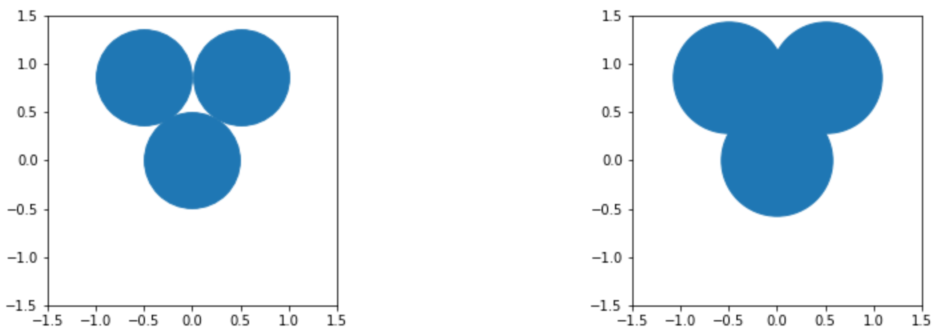

Figure 1: Left is $\epsilon = 0$ and right is $\epsilon = \frac{1}{\sqrt{3}} - \frac{1}{2}$. Configuration for birth/death interval $(0, \frac{1}{\sqrt{3}} - \frac{1}{2})$

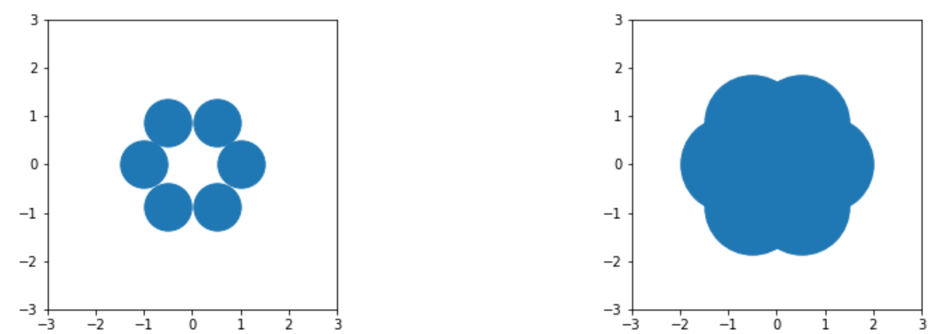

Figure 1: Left is $\epsilon = 0$ and right is $\epsilon = \frac{1}{2}$. Configuration for birth/death interval $(0, \frac{1}{2})$

c) Recall that the size of a gulf can be measured using the quantity $\frac{b+d}{2}$, with b=birth and d=death. Compute the empirical distribution of gulf sizes for P_5 and P_6.

d) Use the function *inverseCumulative(data)* in *SitePercolationGraph.ipynb* to compute the inverse cumulative distribution functions for P_5 and P_6, respectively. Plot the inverse CDFs on a single figure, with log scaling on both axes.

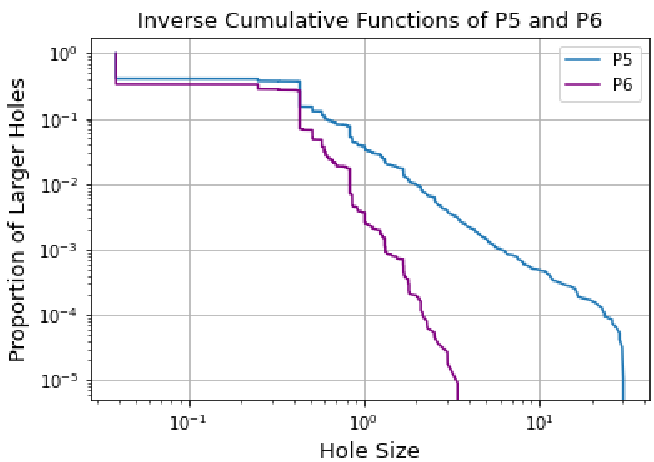

Figure 8: Inverse CDFs of P_5 and P_6.

e) In preparation for your analysis, compute interesting statistics of your choosing based on the persistent homology data of P_5 and P_6.

Statistical Data Of Gulf Sizes Of P_5 And P_6						
	Mean (μ)	Standard Deviation (σ^2)	1st Quartile	Median	3rd Quartile	Interquartile Range
P_5	0.28354	0.64782	0.03868	0.03868	0.43301	0.39433
P_6	0.18031	0.21810	0.03868	0.03868	0.43301	0.39433

Table 2: Statistical data supporting our analysis of P_5 and P_6.

.

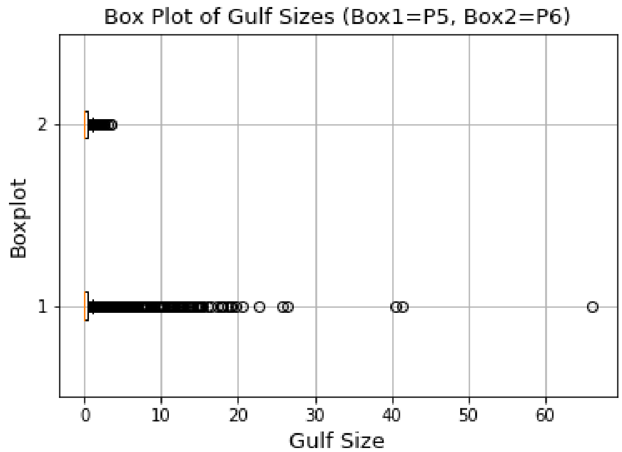

Figure 9: Box and whisker plot of gulf sizes in comparison to gulf sizes of P_5 and P_6. Clear difference in outliers and their size range. Visual representation of our results will be useful for our analysis.

5. Problem 5

Write a paragraph summarizing your work and comparing the geometry of the site percolation polymers P_5 and P_6.

When comparing the site percolation polymers P_5 and P_6, it was intuitive to visually compare the polymers from Figure 4 and Figure 5. With polymers having a difference by 398 disks and p-value difference of 0.1, the one-dimensional persistent homology changes drastically, specifically, gulfs disappearing and holes either dying or shrinking close to death. Our analysis on P_5 and P_6 suggests as p passes the inflection point ($p = \frac{1}{2}$), the structure of the site percolation polymer will rapidly approach the structure of the site percolation itself.

Analysis can support our claim; Figure 2 earlier supported our expectation to the proportion of the site percolation polymers to P_5 and P_6, respectively. We can see there is a rapid spike in proportionality at between $(0.475, 0.525)$, which jumps in value by approximately 0.6. It wishes to converge to a densely packing parallelogram of disks. In Figure 7, we can see the birth/death coordinates of P_5 and P_6. It is clear to see the dense packing of disks in P_6 due to the small birth/death intervals plotted. Results from Table 1 can confirm the frequency of the most common birth/death intervals, majority being of the smallest interval. In comparison, our larger birth/death intervals are seen in the graph for P_5 due to the gulf and hole sizes pictured.

If we look at the statistical data collected (Table 2) and the box plot (Figure 9) together, we can further validate our claim. Table 2 suggests the 'box' of the plot for both polymers are the same for Figure 9. Therefore, the intriguing results from this box plot lies in the outliers. The outliers suggest the gulf sizes are far from its expectation. Specifically, P_5 shows to have a significant amount of outliers, ranging past sizes of 60, suggesting the presence of holes and gulfs. The box plot for P_6 indicates that it is closer to a densely packed polymer like its site percolation, suggesting the gulf sizes are shrinking and wanting to be more of its site percolation.