Principles of Complex Systems, Vols. 1, 2, & 3D
CSYS/MATH 300, 303, & 394
University of Vermont, Fall 2022
Solutions to Assignment 08

What's The Story?

Gods come and go, and still the Turtle Moves.

Name: Krishna Kannan Srinivasan

Conspirators:

---

1. (3 + 3)

   Repeat the first question from the previous assignment changed from $\gamma = 5/2$ to $\gamma = 3/2$.

   Now $1 < \gamma < 2$ so we should see a very different behavior.

   Here's the question reprinted with γ switched to 3/2:

   For $\gamma = 3/2$, generate $n = 1000$ sets each of $N = 10$, 10^2, 10^3, 10^4, 10^5, and 10^6 samples, using $P_k = ck^{-3/2}$ with $k = 1, 2, 3, \ldots$

   How do we computationally sample from a discrete probability distribution?

   Hint: You can use a continuum approximation to speed things up. In fact, taking the exact continuum version from the first two assignments will work.

   (a) For each value of sample size N, sequentially create n sets of N samples. For each set, determine and record the maximum value of the set's N samples. (You can discard each set once you have found the maximum sample.)

   You should have $k_{\mathrm{max},i}$ for $i = 1, 2, \ldots, n$ where i is the set number. For each N, plot the n values of $k_{\mathrm{max},i}$ as a function of i.

   If you think of n as time t, you will be plotting a kind of time series.

   These plots should give a sense of the unevenness of the maximum value of k, a feature of power-law size distributions.

   (b) Now find the average maximum value $\langle\rangle i k_{\mathrm{max},i}$ for each N.

   The steps again here are:

   1. Sample N times from P_k;
   2. Determine the maximum of the sample, k_{max};
   3. Repeat steps 1 and 2 a total n times and take the average of the n values of k_{max} you have obtained.

   Plot $\langle k_{\mathrm{max}} \rangle$ as a function of N on double logarithmic axes, and calculate the scaling using least squares. Report error estimates.

   Does your scaling match up with your theoretical estimate for $\gamma = 3/2$?

How to sample from your power law distribution (and similar kinds of beasts):

We now turn our problem of randomly selecting from this distribution into randomly selecting from the uniform distribution. After playing around a little, $k = 10^6$ seems like a good upper limit for the number of samples we're talking about.

Using Matlab (or some ghastly alternative), we create a cdf for P_k for $k = 1, 2, \ldots, 10^6$ and one final entry $k > 10^6$ (for which the cdf will be 1).

We generate a random number x and find the value of k for which the cdf is the first to meet or exceed x. This gives us our sample k according to P_k and we repeat as needed. We would use the exactly normalized $P_k = \frac{1}{\zeta(3/2)} k^{-3/2}$ where ζ is the Riemann zeta function.

Now, we can use a quick and dirty method by approximating P_k with a continuous function $P(z) = (\gamma - 1)z^{-\gamma}$ for $z \geq 1$ (we have used the normalization coefficient found in assignment 1 for $a = 1$ and $b = \infty$). Writing $F(z)$ as the cdf for $P(z)$, we have $F(z) = 1 - z^{-(\gamma-1)} = 1 - z^{-1/2}$. Inverting, we obtain $z = [1 - F(z)]^{-2}$. We replace $F(z)$ with our random number x and round the value of z to finally get an estimate of k.

Solution:

(a) Plots of k_{max} vs experiment for each N

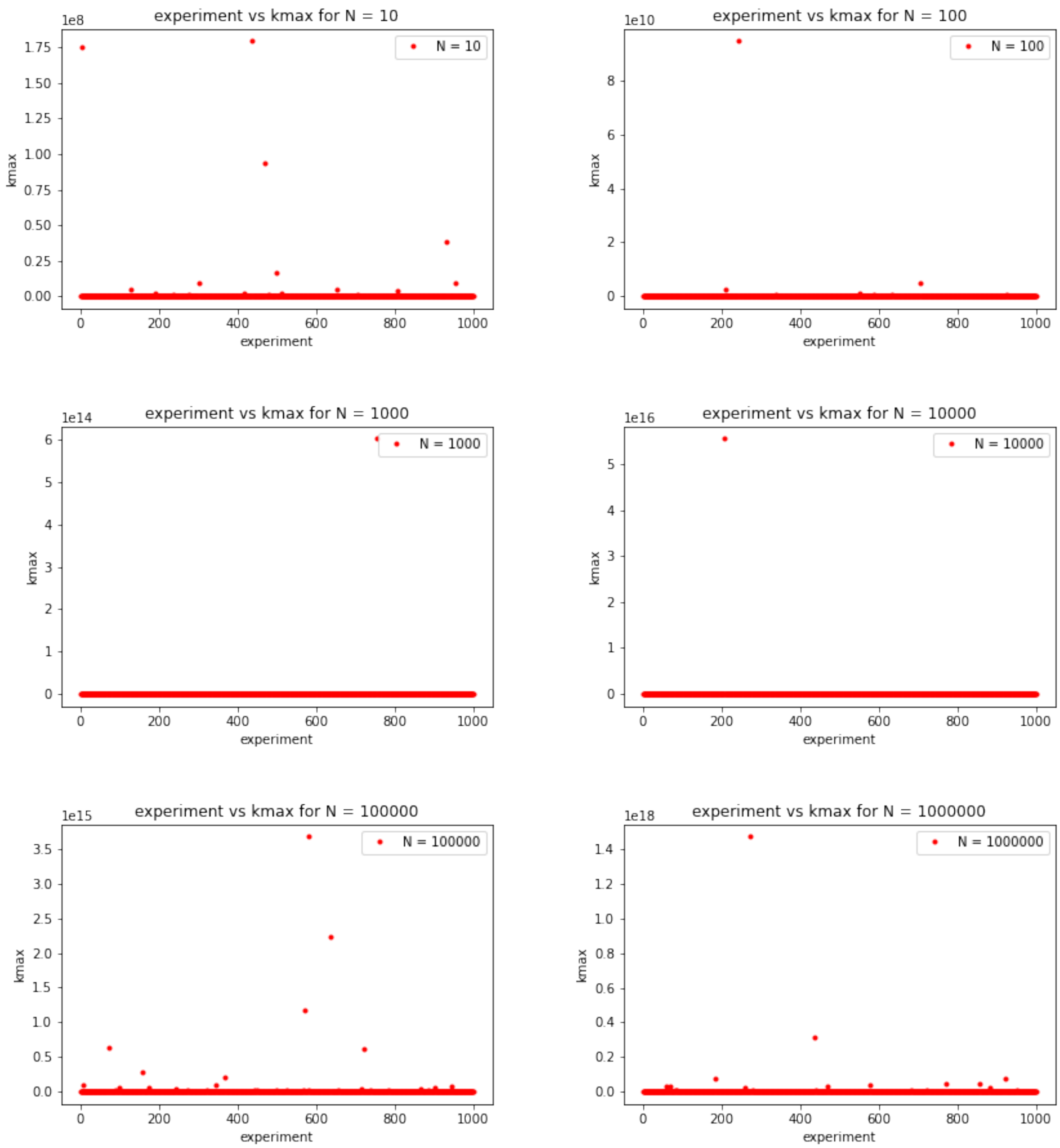


Figure 1: Kmax vs experiments(1 to 1000) for various N

(b) Plots of Average of k_{max} from all experiments for each N

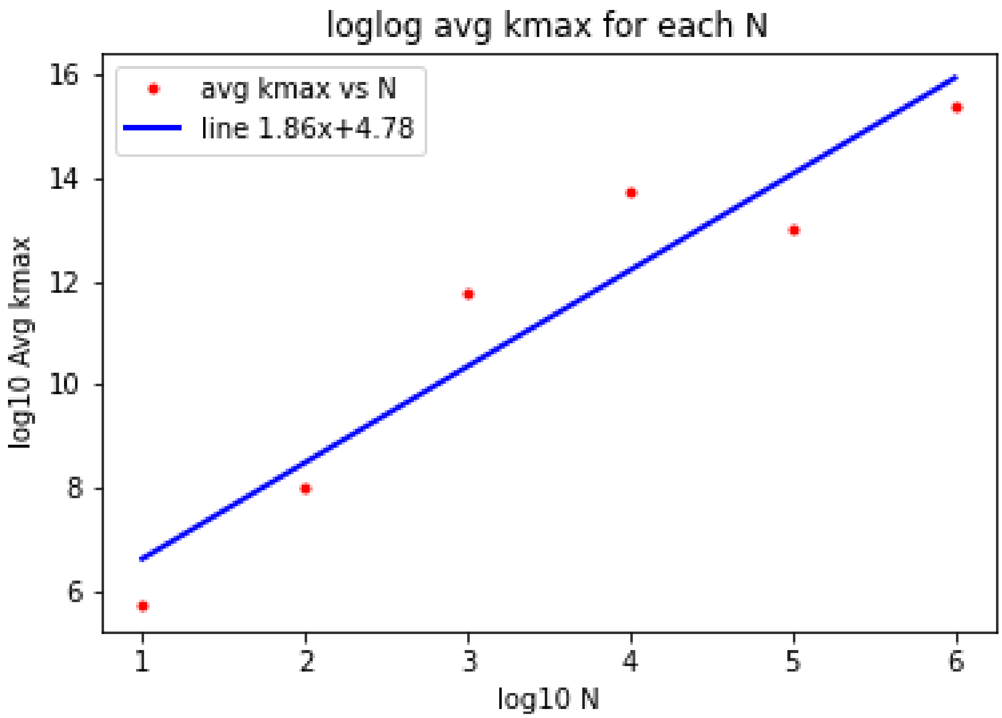


Figure 2: Average k_{max} for each N

Theoretical estimate of Average of $k_{\max,i}$ in terms N:

$$Average k_{\max,i} \propto N^{\frac{-1}{\gamma-1}}$$

For $\gamma = 3/2$, the exponent we find is 2. Here comparing with theoretical estimate of $\approxeq 2$, we get a slope of 1.86 which is close to 2.

Sum of least squares errors of fit: 6.875992 which is large.

$\square$

2. The 1-d theoretical percolation problem:

   Consider an infinite 1-d lattice forest with a tree present at any site with probability p.

   (a) Find the distribution of forest sizes as a function of p. Do this by moving along the 1-d world and figuring out the probability that any forest you enter will extend for a total length ℓ.

(b) Find p_c, the critical probability for which a giant component exists.

Hint: One way to find critical points is to determine when certain average quantities explode. Compute $\langle l \rangle$ and find p such that this expression goes boom (if it does).

Solution:

(a) To find the distribution of forest sizes as a function of p
Let ℓ be the forest size on a 1D lattice.
Let p be the probability a site is a forest.
Let P(X $= l$) be the distribution of forest sizes.
Then,

$$P(X = \ell) = p^{\ell}(1 - p)$$

Where p^{ℓ} represents the sequence of forests.
and $(1 - p)$ represents ending the sequence with "not a forest".

$$P(X = \ell) = (1 - p)e^{\ln(p^{\ell})}$$

$$P(X = \ell) = (1 - p)e^{\ell \ln(p)}$$

$$P(X = \ell) = (1 - p)e^{-\ell \ln(\frac{1}{p})}$$

$$P(X = \ell) = ce^{f(l))}$$

Here the distribution is exponential in ℓ.
(b) Find Critical probability for p when a giant component exists.
Giant component when

$$< \ell > \to \infty$$

$$< \ell > = \sum_{\ell=0}^{\infty} \ell P(X = \ell)$$

$$< \ell > = \sum_{\ell=0}^{\infty} \ell(1 - p)p^{\ell}$$

$$< \ell > = (1 - p) \sum_{\ell=0}^{\infty} \ell p^{\ell}$$

We would want to express the derivative of sum of p^l.

$$< \ell > = p(1 - p) \sum_{\ell=1}^{\infty} \ell p^{\ell-1}$$

Since

$$
\begin{aligned}
\sum_{\ell=0}^{\infty} \ell p^{\ell} &= p \sum_{\ell=1}^{\infty} \ell p^{\ell-1} \\
&= p \sum_{\ell=1}^{\infty} \frac{d}{dp} p^{\ell} \\
&= p \frac{d}{dp} \sum_{\ell=1}^{\infty} p^{\ell}
\end{aligned}
\tag{1}
$$

Then,

$$
< \ell >= p(1-p) \frac{d}{dp} \sum_{\ell=1}^{\infty} p^{\ell}
$$

Using the relation of Geometric sum,

$$
\sum_{\ell=1}^{\infty} p^{\ell} = \frac{p}{1-p}
$$

$$
< \ell >= p(1-p) \frac{d}{dp} \left(\frac{p}{1-p} \right)
$$

Using quotient rule for derivatives,

$$
\frac{d}{dx} \frac{u}{v} = \frac{v.du - u.dv}{v^2}
$$

$$
\begin{aligned}
< \ell > &= p(1-p) \left[\frac{(1-p).1 - p(-1)}{(1-p)^2} \right] \\
&= \frac{p(1-p)}{(1-p)^2} \\
< \ell > &= \frac{p}{1-p}
\end{aligned}
\tag{2}
$$

This quantity explodes when p = 1.
Then $p_c = 1$.

$\square$

3. Show analytically that the critical probability for site percolation on a triangular lattice is $p_c = 1/2$.

   Hint—Real-space renormalization gets it done.:
   http://www.youtube.com/watch?v=JlkbU5U7QqU

Solution:
Using real space renormalization, we can look at percolation between nodes in a triangle. For a triangle, percolation can occur in 4 cases:

(a) 2 nodes forest - 3 combinations possible

(b) all 3 nodes forest - 1 combination possible

Let P(triangle allows percolation) be p^1. If p be the probability that a node in a triangle is a forest,

$$p^{'} = p^3 + 3p^2(1-p)$$

To find the critical probability, we need $p^{'} = p = p_c$.

$$p_c = p_c^3 + 3p_c^2(1-p_c)$$
$$3p_c^2 - 3p_c^3 + p_c^3 - p_c = 0$$
$$2p_c^3 - 3p_c^2 + p_c = 0$$
$$p_c\left(2p_c^2 - 3p_c + 1\right) = 0$$
$$p_c(2p_c - 1)(p_c - 1) = 0$$

This gives us three roots for p_c.

(a) $p_c = 0$, no percolation

(b) $p_c = \frac{1}{2}$, transition point where a giant component comes to exist.

(c) $p_c = 1$. complete percolation since all nodes are "forests".

□

4. (3 + 3)

Coding, it's what's for breakfast:

(a) Percolation in two dimensions (2-d) on a simple square lattice provides a classic, nutritious example of a phase transition.

Your mission, whether or not you choose to accept it, is to code up and analyse the L by L square lattice percolation model for varying L.

Take $L = 20$, 50, 100, 200, 500, and 1000.

(Go higher if you feel $L = 1000$ is for mere mortals.)

(Go lower if your code explodes.)

Let's continue with the tree obsession. A site has a tree with probability p, and a sheep grazing on what's left of a tree with probability $1 - p$.

Forests are defined as any connected component of trees bordered by sheep, where connections are possible with a site's four nearest neighbors on a lattice.

Each square lattice is to be considered as a landscape on which forests and sheep co-exist.

Do not bagelize (or doughnutize) the landscape (no periodic boundary conditions—boundaries are boundaries).

(Note: this set up is called site percolation. Bond percolation is the alternate case when all links between neighboring sites exist with probability p.)

Steps:

i. For each L, run N_{tests}=100 tests for occupation probability p moving from 0 to 1 in increments of 10^{-2}. (As for L, you may use a smaller or larger increment depending on how things go.)

ii. Determine the fractional size of the largest connected forest for each of the N_{tests}, and find the average of these, S_{avg}.

iii. On a single figure, for each L, plot the average S_{avg} as a function of p.

(b) Comment on how $S_{\mathrm{avg}}(p; N)$ changes as a function of L and estimate the critical probability p_c (the percolation threshold).

For the few Matlabbers, a helpful reuse of code (intended for black and white image analysis): You can use Matlab's bwconncomp to find the sizes of components. Very nice.

Solution:

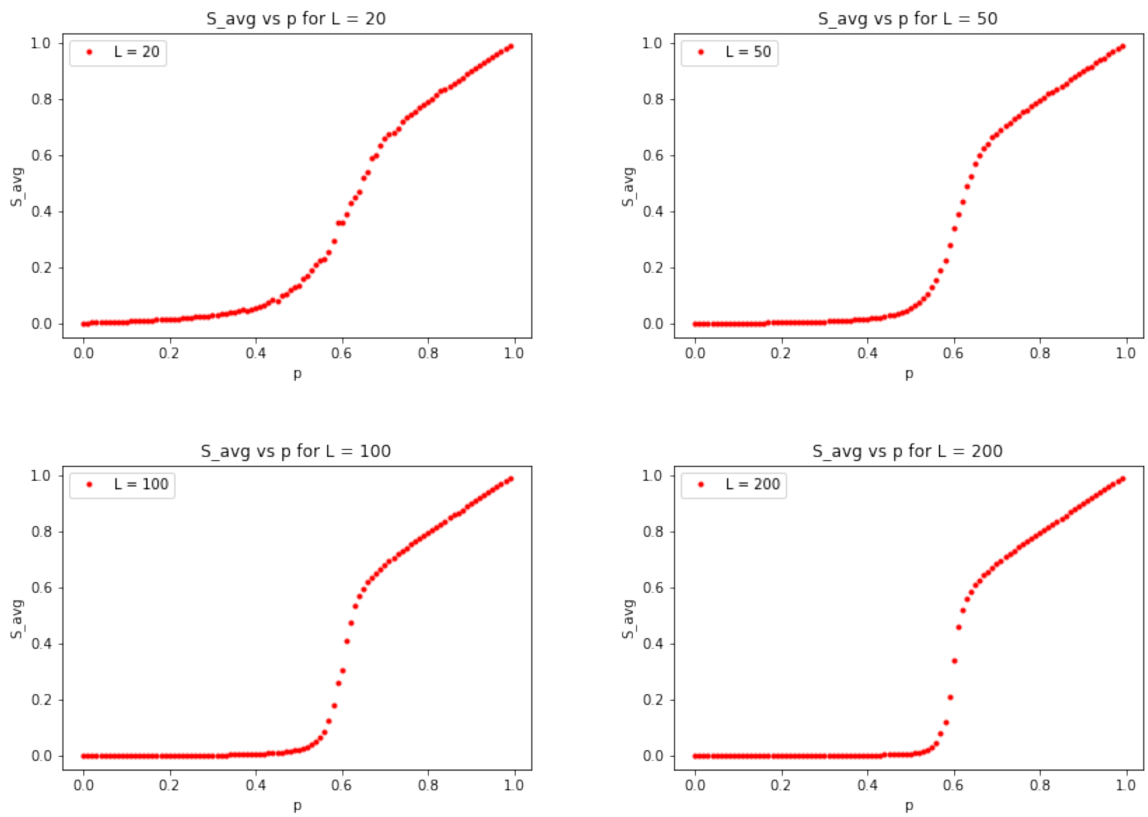


Figure 3: S_{avg} vs p for increasing L

Here we can see that as we increase L the grid size, we can identify the sharp increase the max grid size (the giant component) that results in percolation around $p_c = 0.6$. $\square$

5. (3 + 3)

   (a) Using your model from the previous question and your estimate of p_c, plot the distribution of forest sizes (meaning cluster sizes) for $p \simeq p_c$ for the largest L your code and psychological makeup can withstand. (You can average the distribution over separate simulations.)

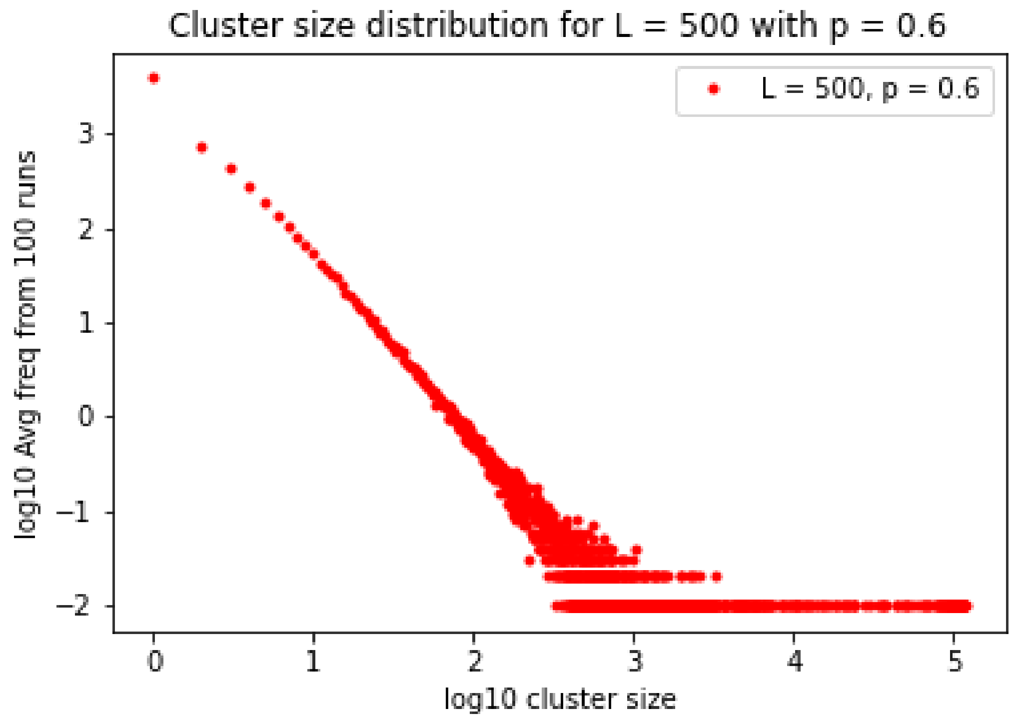
   Comment on what kind of distribution you find.

   (b) Repeat the above for $p = p_c/2$ and $p = p_c + (1 - p_c)/2$, i.e., well below and well above p_c.

   Produce plots for both cases, and again, comment on what you find.

Solution:

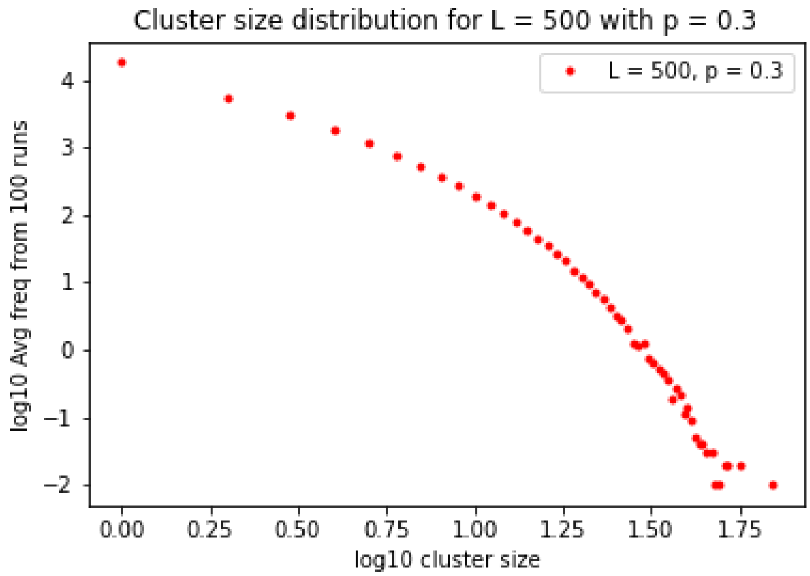
(a) Cluster Size distribution for $p = p_c$,

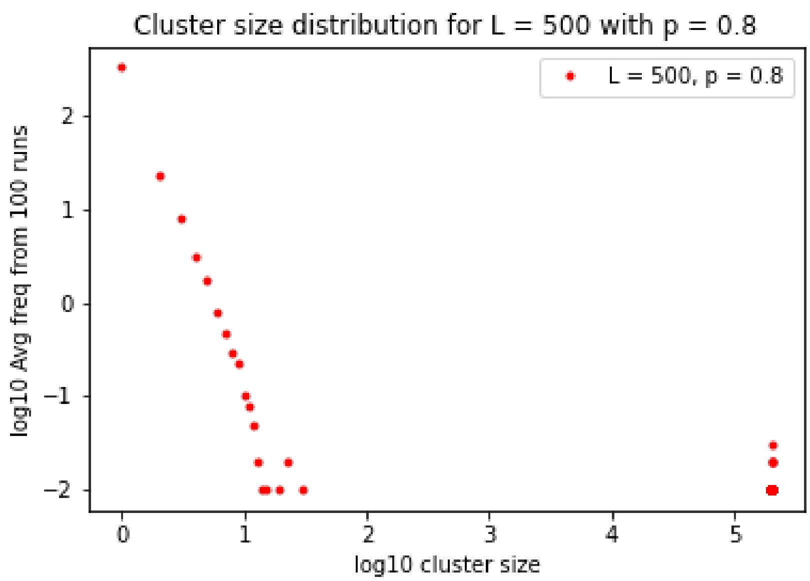


Here I plot the averaged distribution over 100 runs for L $=$ 500 at p $= p_c$.

We can see that this resembles a power law distribution.

(b) Cluster Size distribution for $p < p_c$ and $p > p_c$:





We can see that for $p = 0.3 < p_c$, the average distribution of the cluster size is truncated at the end(size $\approx log10size = 1.6$

And for $p = 0.8 > p_c$, there are two different regions where at the end, the large component stays isolated and the smaller cluster sizes form a truncated distribution. $\square$