Homework 6

Yuancheng Xu

Dec 1, 2020

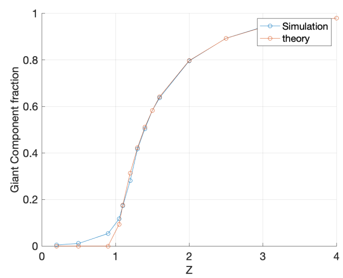
Code link:
`https://github.com/Yuancheng-Xu/AMSC-808N/tree/master/HW6`

Problem 1

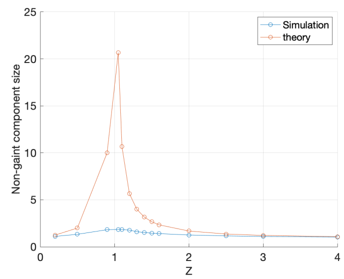
In this problem we are concerned with the fraction S of a giant component (if there is one) and the average size $\langle s \rangle$ of non-giant components. Note that we only consider the fraction of a giant component because the size of it should be infinity if n goes to infinity, while the size of non-giant components is finite. I implement DFS which returns all component sizes. In the code, the size (when n is finite) of the giant component is the largest component size, which will be large if $z > 1$ and will be negligible if $z < 1$. The average size of non-giant components is computed as the average components size except the largest one (when $z < 1$, the average of all components sizes are computed directly).

The simulated S (fraction of a giant component) as well as $\langle s \rangle$ (mean size of non-giant components) are shown in Figure 1, together with their theoretical values. The graph is Poisson and z is varied from 0 to 4, with denser grid around the singularity 1. For each z, 100 random graphs are simulated and the statistics are averaged. I find that simulated S quite match the theoretical values, with 0 when $z < 1$ and suddenly increase when z surpasses 1 and then converge to 1 as z goes to ∞. However, experiments and theories do not agree in the case of $\langle s \rangle$, when z is close to 1. Theories predict much larger non-giant component size. I suspect that this disagreement comes from the fact that n is too small. Indeed, when $n = 10000$, the simulated $\langle s \rangle$ is a little sharper than that in the $n = 1000$ case, but still far from the theory. I expect that larger n will get more accurate results when z is around 1. That is, near singularity point $z = 1$, much larger n should be used in order to approximate the theoretical values. Due to computational restrictions, I do not try higher n.

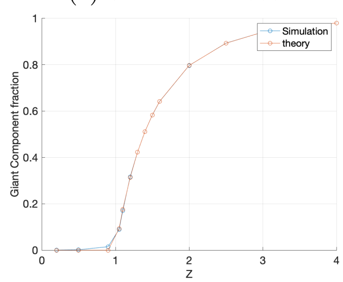
Interesting phenomenon: when z increases (the mean degree increases), it is intuitive that a random vertex will connect to more points. Indeed, the fraction that a giant component occupies increases and thus more fraction of vertices are connected to "infinitely many" other vertices. However, observe that **the average size of non-giant components decreases** (converges to



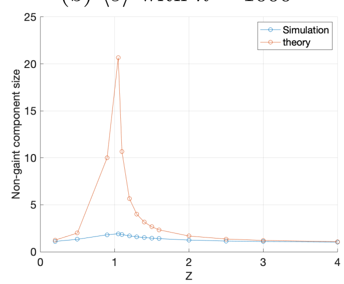
(a) S with $n = 1000$



(b) $\langle s \rangle$ with $n = 1000$



(c) S with $n = 10000$



(d) $\langle s \rangle$ with $n = 10000$

Figure 1: Fraction S of a giant component (if exists) and average size $\langle s \rangle$ of non-giant components.

1), which is somewhat counter-intuitive. That is, when mean degree increases, larger fraction of points tend to stay together, while the isolated points will be more isolated. I interpret this phenomenon as the following: assume that z is very large, a isolated point (those not in the giant component) should have degree 0 (corresponding to a component consisting of itself with size 1) because if it is connected to another point, than with high probability that point is in the giant component and thus the isolated point is not isolated anymore.

Also, when z is slightly larger than 1, I found in the simulations that sometimes there are several "large" components. Indeed, the theoretical S is less than 0.5 so it is possible to have several giant components, although further analysis is needed.

Problem 2

In this problem we want to compute the average length of shortest paths. I choose $z = 8$ instead of $z = 4$ since in simulations when $z = 4$, few of my randomly chosen vertex will be outside of the giant components (and thus are isolated by my argument in Problem 1), so their shortest path will be something useless. By choosing a larger z value, we basically can assume that the whole graph is connected (so the derivation of the theory is valid). The result is shown in Figure 2. We can see that the simulated l is indeed log in n. However, there is a gap between the simulation and theory, possibly because n is too small or the approximations in the theory are not very accurate.

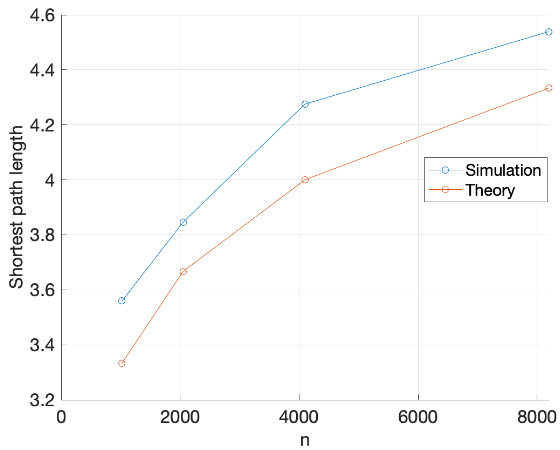


Figure 2: The average length of shortest path with different n, when $z = 8$.

Problem3

In this problem I will derive the formula for the average size $\langle s \rangle$ of a non-giant component (picked by randomly choosing a vertex from non-giant components), when the degree of the random graph is prescribed as p_k.

Some notations

- G_0 and G_1 are generating functions for the vertex degree p_k and excess degree q_k.
- H_0 and H_1 are generating functions for the size of a non-giant component which is selected by a random vertex and a random edge, respectively. Note that $H_0(1) = H_1(1) = 1 - S$ which is the fraction of non-giant components.

Also we need the following equalities. First, by assuming that each non-giant component is a tree and decomposing into one (excess) degree, two (excess) degree, etc, we have the following

$$H_1(x) = xG_1(H_1(x)) \tag{1}$$

$$H_0(x) = xG_0(H_1(x)) \tag{2}$$

Therefore, by taking derivative and substituting $x = 1$ we have

$$H_1'(1) = \frac{G_1(H_1(1))}{1 - G_1'(H_1(1)))} \tag{3}$$

$$H_0'(1) = G_0(H_1(1)) + G_0'(H_1(1))\frac{G_1(H_1(1))}{1 - G_1'(H_1(1)))} \tag{4}$$

The final equality we need is the relationship between excess degree and vertex degree

$$G_1(x) = \frac{1}{z}G_0'(x) \tag{5}$$

Now, by observing Equation 1, we know that $H_1(1)$ is the smallest non-negative real solution of $u = G_1(u)$. Denote $u = H_1(1)$. Also, by substituting $x = 1$ in Equation 2, we have $1 - S = H_0(1) = G_0(H_1(1)) = G_0(u)$ and therefore $G_0(u) = 1 - S$. Finally, substitute $x = u$ in Equation 5 we have $G_0'(u) = zG_1(u) = zu$. Therefore, we can rewrite Equation 4 as

$$H_0'(1) = 1 - S + zu\frac{u}{1 - G_1'(u)} = 1 - S + \frac{zu^2}{1 - G_1'(u)} \tag{6}$$

Since $H_0(1) = 1 - S$, we need to normalize $H_0'(1)$ in order to obtain the average size of a non-giant component:

$$\begin{aligned} \langle s \rangle &= \frac{H_0'(1)}{H_0(1)} \\ &= \frac{1}{1 - S}(1 - S + \frac{zu^2}{1 - G_1'(u)}) \\ &= 1 + \frac{zu^2}{(1 - S)(1 - G_1'(u))} \end{aligned} \tag{7}$$