Solution 9-11

Ex 9.11: Average component size [4]

In this chapter we derived a method to compute c_k, the probability that a randomly picked node is in a component of k nodes. Now, let s_k be the probability that a randomly picked component has k nodes. Formulate an equation that relates c_k and s_k. Then translate your equation into an equation that relates the C and S, the generating functions of c_k and s_k. Use your equations to compute the average component size (the expectation value of s_k, for the networks from Ex. 9.5 and Ex. 9.6) [This one has some pitfalls, proceed with caution.]

Solution

First, let's ask ourselves why are c_k and s_k different at all? Does it make a difference if we pick select a component by picking a random component straight away, instead of picking a random node and then selecting this nodes component? Of course it does. At this stage of this book we realize that this is a similar to the situation that we encountered when discussing excess degree. In Chap. 6 the key insight was that when we follow links we have a higher chance of arriving at nodes that have more links. Here we have the same. If we select a component by picking a random node then we have a higher chance of picking components that have more nodes.

Finding a relationship between s_k and c_k becomes easier if we think in terms of numbers rather than probabilities. So, let's suppose the and the number of components is N_c. This means that the number of components of size k is $s_k N_c$, and the number of nodes in components of size k is $k s_k N_c$. Therefore the probability that a randomly-picked node is in a component of size k is

$$c_k = \frac{k N_c s_k}{N}, \tag{1}$$

where N is the total number of nodes. Now we have to take care because not all of the symbols we have used are independent. If we specify N_c and s_k, we can compute N as

$$N = \sum k N_c s_k. \tag{2}$$

Baking this condition into the equation above gives us

$$c_k = \frac{k N_c s_k}{\sum k N_c s_k}. \tag{3}$$

We can now cancel N_c (you have realized that it was a crutch that would disappear again, haven't you?). The result is

$$c_k = \frac{k s_k}{\sum k s_k}. \tag{4}$$

which is very similar to an equation for that we derived for the relationship between degree distribution and excess degree distribution. (In this case there is no $k + 1$ because we count

the selected node itself when we compute the size of the component, whereas we did not count the selected link when we count the excess degree)

Now that we have found the relationship between c_k and s_k we can translate it into a relationship between the generating functions. At this point we could just multiply x^k to the equation above and then sum both sides over k. However, let's rather follow our translation procedure one more time. We start by defining the generating functions

$$S = \sum s_k x^k \qquad C = \sum c_k x^k. \tag{5}$$

Now we substitute our knowledge about c_k into the generating function for C,

$$C = \sum \frac{k s_k}{\sum k s_k} x^k. \tag{6}$$

We can write this more nicely as

$$C = \frac{\sum k s_k x^k}{\sum k s_k}. \tag{7}$$

The terms that appear are very similar to

$$S' = \sum k s_k x^{k-1}. \tag{8}$$

If we multiply this by x we find

$$x S' = \sum k s_k x^k \tag{9}$$

Hence we can write

$$C = \frac{\sum k s_k x^k}{\sum k s_k} = \frac{x S'}{\sum k s_k} \tag{10}$$

The term in the denominator is just $S'(1)$, and hence

$$C = x \frac{S'}{S'(1)}. \tag{11}$$

This is very similar to the generating function that relates G and Q (degree and excess degree generating functions), $Q = G'/G'(1)$. Again an extra x appears, because we count the node that we picked randomly as part of the component.

Now that we have an equation we can test it. The simplest test is offered by the network from Ex. 9.5, where every node has degree 1. In that exercise we found

$$C = x^2. \tag{12}$$

Our challenge is now to compute S. Solving our generating function equation for S is a bit tricky, but we can write

$$S' = \frac{C S'(1)}{x} \tag{13}$$

At this point it may be confusing that S still appears on the right side in the form of the factor $S'(1)$. The solution is to just ignore this problem for now. If this makes you unhappy, consider that our challenge at the moment is to find the function S, whereas $S'(1)$ is just a number. Once we know the function in principle, finding such a missing factor is usually simple. Substituting the known $C = x^2$ into the equation yields

$$S' = \frac{x^2 S'(1)}{x} = x S'(1) \tag{14}$$

2

To find S we now need to integrate this equation. In the integration we again keep in mind that $S'(1)$ is only a number, so it can be pulled out of the integral

$$S = \int xS'(1)\mathrm{d}x = S'(1)\int x\mathrm{d}x = \frac{1}{2}S'(1)x^2 \tag{15}$$

Normally an integral also leads to a constant of integration. However, in this case we know this constant must be 0. If there was a constant term in S it would imply would mean that there is a non-zero probability that picking a component at random results in a component with zero nodes. Because this is not possible. Therefore we are interested in the solution to the integral where the constant part vanishes, which means $S(0) = 0$. In the network considered here this means that the constant of integration must be zero. (This isn't always the case in this type of calculation, as we may sometimes require a constant of integration to ensure $S(0) = 0$, as we will see in the following.)

To find $S'(1)$ we need some more information. Fortunately we know that S should be a properly normalized generating function so $S(1) = 1$. Evaluating S we find

$$1 = S(1) = \frac{S'(1)}{2} \tag{16}$$

which tells us

$$S'(1) = 2, \tag{17}$$

which is also the average component size(!) Moroever, substituting back gives

$$S = x^2, \tag{18}$$

so this is telling us if we pick a random component we will get a component with two nodes all the time. This makes sense as every component in the network contains two nodes.

Having tested our equation on the simple test example we can now move on to the network from Ex. 9.6. In this case we have

$$C = \frac{x}{(2-x)^2}. \tag{19}$$

Substituting into $C = xS'/S'(1)$ yields

$$\frac{x}{(2-x)^2} = x\frac{S'}{S'(1)} \tag{20}$$

and hence

$$S' = \frac{S'(1)}{(2-x)^2}. \tag{21}$$

So this is a constant with a quadratic function in the denominator. So we might guess that the solution is of the form

$$S = \frac{a}{2-x} + b \tag{22}$$

where a and b are constants. We can check this guess by differentiating, which yields

$$S' = \frac{a}{(2-x)^2} \tag{23}$$

So by comparison with $S' = S'(1)/(2-x)^2$ we see that $a = S'(1)$. So far we have

$$S = \frac{S'(1)}{2-x} + b. \tag{24}$$

3

To find the value of b we can use again the information that there is zero chance to randomly pick a component with 0 nodes, that is $S(0) = 0$. Hence we know,

$$0 = \frac{S'(1)}{2} + b \tag{25}$$

So $b = -S'(1)/2$ and thus,

$$S = \frac{S'(1)}{2-x} - \frac{S'(1)}{2} \tag{26}$$

Finally we can find $S'(1)$ by demanding proper normalization:

$$1 = S(1) = S'(1) - \frac{S'(1)}{2} = \frac{S'(1)}{2} \tag{27}$$

and so

$$S'(1) = 2. \tag{28}$$

So let us think for a moment what this means. In Ex. 9.6 we found that if we pick a random node then the mean number of nodes that we find in the respective component is 3. However, our result $S'(1) = 2$ tells us that the average component size in the network is 2. This difference appears because in the former case the larger components enter the average with larger weight (there is a greater chance to pick larger components when we pick nodes). In the latter case we pick a random component straight away so all components enter with equal weights.

We can finish by writing the final generating function for components size (when components are picked with equal probability)

$$S = \frac{2}{2-x} - 1 = \frac{x}{2-x}. \tag{29}$$