Networks and Complexity

Exercise Sheet 7: The Importance of Being Random

This is an exercise sheet from the forthcoming book Networks and Complexity.
Find more exercises and solutions at https://github.com/NC-Book/NCB

Ex 7.1: Independence [1]

One problem of the G(n,M) is statistical interdependence between links. Suppose we have a simple graph with $N = 3$ nodes and $K = 2$ links.

a) Compute the probability that there is a link between node 1 and node 2.
b) Suppose you know that there is a link between nodes 2 and 3, what is the probability for a link between node 1 and 2 now?

Ex 7.2: Combinatorics [2]

Compute the following:

a) In the German Democratic Republic there were four ice cream flavours. So, how many different ways were there to make an ice cream cone containing two scoops of different flavors? Assume the order of scoops does not matter.
b) As part of their degree program students have to pick three optional courses. In total 20 optional courses are offered. How many possible combinations of three of these courses exist?
c) On a lottery ticket you can pick 6 numbers out of 49. How many possible combinations are there?

Ex 7.3: Degree distribution of ER graphs [2]

We discovered that the degree distribution of ER graphs can be written as $p_k = z^k \mathrm{e}^{-z}/k!$.

a) Show that p_k is a correctly normalized probability distribution. (i.e. $\sum p_k = 1$)
b) Compute the mean degree from p_k to confirm that it is z.
c) Finally, compute the excess degree distribution q_k and the mean excess degree q.

Ex 7.4: Motifs in a large network [2]

In this chapter we derived the formula $n_{\text{motif}} = z^k N^{n-k}/s$ for the approximate number of instances of a motif with n nodes, k links, and symmetry number s in a network with N nodes and mean degree z. Let's use this formula to estimate some motif counts in a network of $N = 10.000$ nodes and $z = 10$.

a) To build up some confidence in the formula, let's use it to estimate the number of nodes in the network (so the motif we are looking for is just a node.)
b) Does the motif formula also yield the right result when we use it to compute the number of links?
c) Now, estimate the number of 4-cycles ($\square$).

Ex 7.5: 3-chains done differently [3]

Let's try to estimate the number of 3-chains in an ER graph in a different way. Revisit the reasoning from Chap. 5 regarding the number of nodes that we find at distance 2 from a typical node. Use this type of reasoning to estimate the number of 3-chains in the network. Show that the resulting estimate is consistent with n_{--} from the present chapter.

Ex 7.6: Loners, pairs and triads [3]

A sailing club has $N = 1000$ members. Friendships in the club form an ER graph with mean degree $z = 4$.

a) Calculate the number of isolated nodes, i.e. the number of members who have no friends in the club. (Hint: $0! = 1$)

b) Calculate the number of isolated pairs of friends in the club, i.e. the number of components of size 2. (Hint: This is not the same as the number of link motifs, because motifs are not necessarily isolated. You will have to find a different way.)

c) Finally compute the expected number of components of size 3.

Ex 7.7: The most common cycle [3]

We know our random graphs are locally treelike, but we also know that if $z > 1$ then there must be some cycles. To explore this, consider an ER graph with $N = 101$ nodes and mean degree $z = 2$. Find out what the most common length l for a cycle is.

Ex 7.8: Apparent competition [4]

Food webs are the networks of who-eats-who in ecology. In these networks the nodes represent species and directed links between them are predator-prey interactions. An important ecological motif is the 'apparent competition' motif, which consists of one species feeding on two other species. How many apparent competition motifs would we expect in a (small) network of N species and K predator-prey links if the links were distributed randomly.

Ex 7.9: 3-chains in yet another way [4]

Let's estimate the number of three-chains in large network in yet another way: This time we argue from the perspective of links. So we pick two links from the network and then ask if these to links happen to be incident on the same node, such that they form a three node chain. Then we check the next pair of links, and so on, until we have checked all possible pairs of links that can be picked from the network. Does this way of thinking also give you a formula for the number of three node chains?

Ex 7.10: Poisson approximation [4]

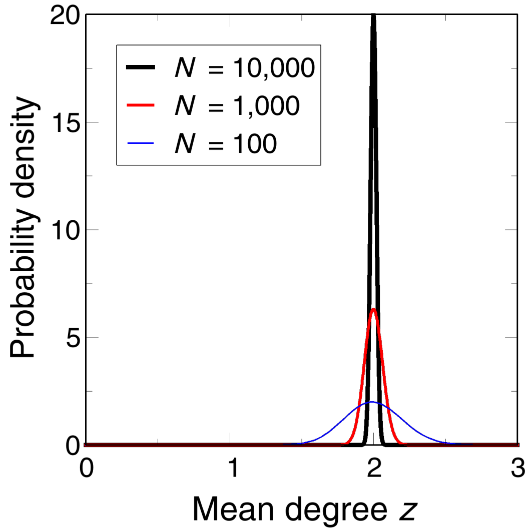
Using approximations from the chapter and

$$(1 + x)^n \approx e^{xn} \tag{55}$$

for small x show that the Binomial distribution can be approximated by the Poisson distribution.

Ex 7.11: Mean degree distribution [4]

The following figure shows the probability that a network drawn from an ER ensemble with mean degree 2 has actually a mean degree of z, where z is not necessarily z,



But how was this figure actually made?

a) Find a way to compute the probability that a network of 100 nodes drawn from an ensemble with mean degree 2 actually has mean degree 2.

b) Bonus: How does your probability relate to the probability density shown in the figure, and why does the figure show probability density, rather than probability directly? (If you are unsure check out the explanation in the solution)

Ex 7.12: Counting lollypops [4]

The lolly motif consists of a 4-cycle with an additional link attached to the nodes in the cycle ($\diamond -$).

a) Use the formula from the lecture to estimate the number of lolly motifs, $n_{\diamond -}$, in a large network with given z.

b) Now suppose that we already know the number of four-cycles, $n_{\square}$ in our network. Think about the probability that a given four-cycle has an additional node attached to it, so that it is part of a lolly. Use this approach to estimate for the number of lolly motifs based on $n_{\square}$ and z.

c) Now, substitute the estimate for the number of 4-cycles from Ex. 7.4 into your solution from (b) and show that the result is consistent with part (a).