Assignment 4: Solutions

TOTAL MARKS: 0

1. All of the nodes in an r-regular graph will have node degree r, and hence the row and columns sums of the adjacency matrix A will all be r.

   Multiplying a matrix by $\mathbf{1} = (1, 1, \ldots, 1)$ is simple the operation of calculating the row sum, *i.e.,* it creates a vector $\mathbf{v}$, whose elements are the sums of the rows. As these will all be r we have $\mathbf{v} = r\mathbf{1}$, *i.e.,*

   $$A\mathbf{1} = r\mathbf{1},$$

   which says that $\mathbf{1}$ is an eigenvector by the definition, and that r is its eigenvalue.

2. For a node to become disconnected, each of the links to it should be rewired and no other links wired back to it.

   The number of links in the network is $nr/2$, so the number of rewired links, on average is $nrp/2$. Hence, the number of *ends* of rewirings is nrp on average. Thus there are rp ends of rewirings for any one node on average.

   Taking the number of rewirings to reach a node as Binomial, we can apply the Poisson limit, and hence the number of rewirings to a node is distributed as

   $$p_n = \frac{e^{-rp}(rp)^n}{n!}.$$

   Thus the probability of no rewirings **to** the node is

   $$p_0 = e^{-rp}.$$

   The probability that all r links to the node are rewired is

   $$\text{Prob}\left\{r \text{ links rewired}\right\} = p^r.$$

   These two events are independent, and so the joint probability that there are no rewirings to the node, and that all of its links are rewired is

   $$p^r e^{-rp} = \left(pe^{-p}\right)^r.$$

3. (a) The probability that none of the new node's k edges (for a degree k node) connect to the largest connected component is approximately (assuming **independence**) is the probability that all of these link join the fraction of nodes outside the largest connected component, *i.e.*,

$$P(\text{ node's CC is outside large CC}|k) = (1-q)^k.$$

(b) The Poisson approximation for degree[1] is

$$P(k) \simeq e^{-(n-1)p}\frac{((n-1)p)^k}{k!}.$$

Hence, the probability of interest can be written approximately (using the Theorem of Total Probability) as

$$\begin{aligned}
&P(\text{ node's CC is outside large CC}) \\
&= \sum_{k=0}^{\infty} P(\text{ node's CC is outside large CC}|k) e^{-(n-1)p}\frac{((n-1)p)^k}{k!} \\
&= e^{-(n-1)p}\sum_{k=0}^{\infty}\frac{((n-1)p(1-q))^k}{k!} \\
&= e^{-(n-1)p}e^{(n-1)p(1-q)} \\
&= e^{-(n-1)pq}
\end{aligned}$$

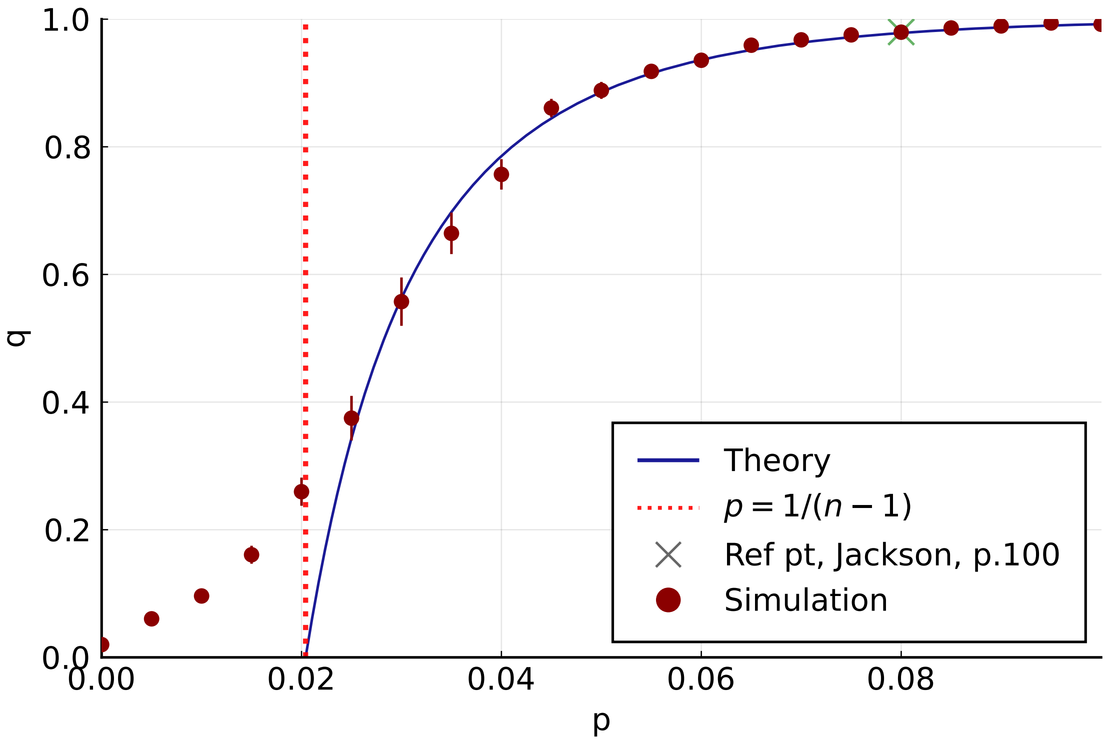
(c) Note that by definition $1-q = P(\text{ node's CC is outside large CC})$, therefore we get the equation

$$1-q = e^{-(n-1)pq},$$

or

$$q = 1 - e^{-(n-1)pq}.$$

(d) This is my exemplar figure:



[1] Note that in lectures we approximated $(n-1)p \simeq np$ but not here.

It shows both the calculated q, and the mean proportion of nodes $\hat{q}$ in the largest connected component in (100) simulated GER random graphs (with 95th percentile confidence intervals for the estimate of $\hat{q}$) as well as a reference point from Jackson.

Note that

- there are only positive solutions to the equation when $(n-1)p > 1$,
- the approximation (which uses the Poisson approximation to node degree distribution) is only valid for sparse networks, with $np = \lambda$ not too large.

and these facts have helped choose the range of interest, and further, this is the range that displays the threshold behaviour that is often interesting for such networks.

The approximation for q is good for larger values of p, but near the threshold the approximation breaks down. A simple explanation is that the proof above presumes that a single large connected component exists and covers the majority of nodes. The probability of such a connected component drops near the threshold (for a finite graph), and hence, the underlying assumption of the approximation is invalid.

Below the threshold, all bets are off. In this regime, we know that there may be a largest connected component of $O(\log n)$, and thus $q = 0$ is an underestimate of the true proportion.

4. Solutions will be evaluated individually depending on the graph model chosen.