Complex Networks
Solutions 7

---

1. (a) K_4 is the square with both its diagonals present.

   (b) Every subset of 4 nodes can support exactly one copy of K_4. There are $\binom{n}{4}$ ways of choosing a subset of 4 nodes; K_4 is present on these nodes if the corresponding 6 edges are present, which has probability p^6. Hence, by the linearity of expectation, the expected number of copies of K_4 in $G(n,p)$ is exactly $\binom{n}{4}p^6$.

   (c) Let $A_1, A_2, \ldots, A_{\binom{n}{4}}$ denote all the possible 4-node subsets of V. Let $\chi(A_i)$ denote the indicator that the subgraph of $G(n,p)$ induced by A_i is K_4, or equivalently, that all six edges between pairs of nodes in A_i are present in the random graph. Thus, $\chi(A_i)$ is a random variable. Let N_4 denote the number of copies of K_4 in $G(n,p)$, which is also a random variable. We then have

$$N_4 = \sum_{i=1}^{\binom{n}{4}} \chi(A_i).$$

   Consequently,

$$\text{Var}(N_4) = \sum_{i,j=1}^{\binom{n}{4}} \text{Cov}(\chi(A_i), \chi(A_j)). \tag{1}$$

   Now, if the node sets A_i and A_j have only zero or one nodes in common, then they have no edges in common, and the random variables $\chi(A_i)$ and $\chi(A_j)$ are independent. Consequently, their covariance is zero. On the other hand, if these node sets have two nodes in common, then they have one edge in common, and

$$\text{Cov}(\chi(A_i), \chi(A_j)) = \mathbb{E}[\chi(A_i)\chi(A_j)] - \mathbb{E}[\chi(A_i)]\mathbb{E}[\chi(A_j)] = p^{11} - p^{12}.$$

   To see the last equality, note that the expectation of the indicator $\chi(A_i)$ is the probability of the event for which it is the indicator, namely that K_4 is present on A_i; for this to happen, six edges need to be present, the probability of which is p^6. On the other hand, the product of the indicators is 1 if and only if both indicators are 1, i.e., if K_4 is present on both A_i and A_j. Hence, the expectation of the product is the probability that both copies of K_4 are present. But for this to happen, we need 11 edges to be present, as one edge is common. Hence, the probability of this event is p^{11}.

   Similarly, we next look at the case when A_i and A_j have 3 nodes in common. In this case, they have $\binom{3}{2} = 3$ edges in common (edge between each pair of these 3 nodes). Hence, reasoning similarly to above, we get

$$\text{Cov}(\chi(A_i), \chi(A_j)) = \mathbb{E}[\chi(A_i)\chi(A_j)] - \mathbb{E}[\chi(A_i)]\mathbb{E}[\chi(A_j)] = p^9 - p^{12}.$$

   The first term is p^9 because 3 edges being in common means that we need 9 distinct edges to be present in order for there to be a copy of K_4 on each of A_i and A_j. Finally, if A_i and A_j have all 4 nodes in common, i.e., they are the same set, then they have all six edges in common, and

$$\text{Cov}(\chi(A_i), \chi(A_j)) = \mathbb{E}[\chi(A_i)\chi(A_j)] - \mathbb{E}[\chi(A_i)]\mathbb{E}[\chi(A_j)] = p^6 - p^{12}.$$

We need to substitute these different estimates for the covariance into (1), counting the number of contributions of each type. The number of ways we can choose two sets of 4 nodes overlapping in two nodes is the number of ways we can choose 6 nodes in total, and then choose which 2 of them are going to belong to both sets, and then how to split the remaining 4 nodes into two sets. This can be done in $\binom{n}{6} \times \binom{6}{2} \times \binom{4}{2}$ ways. Ignoring constants, this term is n^6. Similarly, the number of ways in which we can choose two sets of 4 nodes overlapping in three nodes is $\binom{n}{5} \times \binom{5}{3} \times \binom{2}{1}$ ways. Again ignoring constants, this term is n^5. Finally, the number of ways of choosing two sets of 4 nodes having 4 nodes in common is clearly the number of ways of choosing 4 nodes, which is $\binom{n}{4}$. Ignoring constants, this is n^4. Putting together these estimates, and substituting in (1), we get

$$\text{Var}(N_4) = n^4(p^6 - p^{12}) + n^5(p^9 - p^{12}) + n^6(p^{11} - p^{12}). \tag{2}$$

Which of these terms is dominant depends on the value of p.

(d) We saw in part (a) that $\mathbb{E}[N_4]$, the expected number of copies of K_4 in $G(n.p)$ is n^4p^6. Take $\alpha_c = 4/6 = 2/3$. Then, if $p = n^{-\alpha}$, it is easy to see that $\mathbb{E}[N_4]$ tends to zero as n tends to infinity if $\alpha > \alpha_c$, and to infinity if $\alpha < \alpha_c$.
It follows by Markov's inequality that, if $\alpha > \alpha_c$, then

$$\mathbb{P}(N_4 \geq 1) \leq \frac{\mathbb{E}[N_4]}{1} \to 0$$

as n tends to infinity.
Suppose next that $\alpha < \alpha_c$. We know that $\mathbb{E}[N_4]$ tends to infinity, but is this enough to guarantee that there is at least one copy of K_4 in $G(n,p)$, with high probability? To answer this, we need to use Chebyshev's inequality, which gives us

$$\mathbb{P}(N_4 = 0) = \mathbb{P}(N_4 \leq 0) \leq \mathbb{P}(|N_4 - \mathbb{E}N_4| \geq \mathbb{E}N_4) \leq \frac{\text{Var}(N_4)}{(\mathbb{E}N_4)^2}.$$

Now, substituting for $\text{Var}(N_4)$ in the above from (2), with $p = n^{-\alpha}$, we get

$$\begin{aligned} \mathbb{P}(N_4 = 0) &\leq \frac{n^4(p^6 - p^{12}) + n^5(p^9 - p^{12}) + n^6(p^{11} - p^{12})}{(n^4p^6)^2} \\ &= n^{-4+6\alpha}(1 - n^{-6\alpha}) + n^{-3+3\alpha}(1 - n^{-3\alpha}) + n^{-2+\alpha}(1 - n^{-\alpha}). \end{aligned}$$

Now, if $\alpha < \alpha_c = 2/3$ (and $\alpha > 0$), then it is clear that the expression above tends to zero as n tends to infinity. Hence $\mathbb{P}(N_4 \geq 1)$ tends to 1, as we are asked to show.

2. (a) $K_{2,2}$ consists of two sets X' and Y' of 2 nodes each; there are 4 edges, one between each node in X' and each node in Y'.

(b) There are $\binom{n}{2}$ ways of choosing 2 nodes from X and an equal number of ways of choosing 2 nodes from Y. There is exactly one way of placing a copy of $K_{2,2}$ on these 4 nodes. Moreover, the copy is present with probability p^4, the probability that all 4 possible edges are present. Hence, by linearity of expectation, the expected number of copies of $K_{2,2}$ in $G(n,n,p)$ is exactly $\binom{n}{2}^2 p^4$.

(c) Let H denote the subgraph $K_{2,2}$; we use H_i, H_j etc. to denote possible copies of H that might appear in $G(n,n,p)$. Let N_H denote the random number of copies of H present in $G(n,n,p)$, and χ_i the indicator that the i^{th} copy (in some ordering) is present. Then,

$$N_H = \sum_{i=1}^{\binom{n}{2}^2} \chi_i,$$

and so,

$$\text{Var}(N_H) = \sum_{i,j} \text{Cov}(\chi_i, \chi_j). \tag{3}$$

Now, the covariance depends on how the copies H_i and H_j overlap. If they are non-overlapping, then their indicators are clearly independent random variables, and their covariance is zero. This is still true if the ovelap involves some nodes, but no edges. The only non-zero terms in the sum correspond to overlaps that contain one or more edges.

The graph defined by the overlap is necessarily some sub-graph H' of H. What are the possible subgraphs containing one or more edges? They are $K_{1,1}$ (which is just a single edge), $K_{1,2}$ and $K_{2,2}$. We don't write $K_{1,2}$ and $K_{2,1}$ separately, but we will count both possibilities, that one of the nodes is in X and two are in Y, or the other way round. For each of these subgraphs, we can count the number of possible copies of H_i and H_j intersecting in H' that can appear in $G(n, n, p)$.

For example, for $H' = K_{1,1}$, we are asking for two copies of $K_{2,2}$ that share one edge in common. This means choosing 3 nodes in each of the sets X and Y, choosing how the nodes are paired, and which pairing is common to both graphs. This can be done in

$$\binom{n}{3}^2 \binom{3}{1}^2 2 \approx n^6$$

ways. As usual, the approximate count only depends on the number of vertices chosen.

Similiarly, if the intersection is $K_{1,2}$, the number of ways this could occur is the number of ways of choosing either 3 nodes in X and 2 in Y (or the other way round), then picking one distinguished vertex in X and 2 in Y that will support the shared copy of $K_{1,2}$ (or the other way round). The number of ways of doing this is

$$2\binom{n}{3}\binom{n}{2}\binom{3}{1} \approx n^5.$$

Again, the power of n is just the number of vertices to be chosen.

Likewise, the number of ways of choosing vertices is $\binom{n}{2}^2 \approx n^4$ if the intersection is $K_{2,2}$.

Next, let us work out $\text{Cov}(\chi_i, \chi_j)$ in each of these cases. If H_i and H_j intersect in $K_{1,1}$, then $\mathbb{E}[\chi_i \chi_j] = p^7$ as 7 edges need to be present (twice 4 edges, less 1 in common), and so

$$\text{Cov}(\chi_i, \chi_j) = p^7 - p^8 \approx p^7,$$

where the p^8 term is $\mathbb{E}[\chi_i]\mathbb{E}[\chi_j]$, and is negligible in comparison as p tends to zero.

Similarly $\mathbb{E}[\chi_i \chi_j]$ is equal to p^6 if H_i and H_j intersect in $K_{1,2}$ (as two edges are common to both, and so a total of 6 edges are required), and to p^4 if the intersection is $K_{2,2}$. In each case, $\mathbb{E}[\chi_i]\mathbb{E}[\chi_j] = p^8$, and is negligible in comparison to the $\mathbb{E}[\chi_i \chi_j]$ term.

Putting together these expressions for the covariance, and combining them with approximate counts for the various possibilities, we obtain from eq. (3) that

$$\text{Var}(N_H) \approx n^6 p^7 + n^5 p^6 + n^4 p^4.$$

Substituting $p = n^{-\alpha}$ in the above expression, we get

$$\text{Var}(N_H) \approx n^{6-7\alpha} + n^{5-6\alpha} + n^{4-4\alpha}.$$

(d) We saw in part (b) that $\mathbb{E}[N_H] \approx n^4 p^4 = n^{4-4\alpha}$ for $H = K_{2,2}$. Hence, if $\alpha > 1$, then $\mathbb{E}[N_H]$ tends to zero, and it follows by Markov's inequality that $\mathbb{P}(N_H \geq 1)$ tends to zero, as n tends to infinity. Consequently, for α in this range, $G(n, p)$ does not contain $K_{2,2}$, whp.

Next, suppose $0 \le \alpha < 1$. By Chebyshev's inequality,

$$\begin{aligned}
\mathbb{P}(N_H = 0) &= \mathbb{P}(N_H \le 0) \le \mathbb{P}(|N_H - \mathbb{E}N_H| \ge \mathbb{E}N_H) \le \frac{\text{Var}(N_H)}{(\mathbb{E}N_H)^2} \\
&\approx \frac{n^{6-7\alpha} + n^{5-6\alpha} + n^{4-4\alpha}}{(n^{4-4\alpha})^2} \\
&= n^{\alpha-2} + n^{2\alpha-3} + n^{2\alpha-3} + n^{4\alpha-4},
\end{aligned}$$

which tends to zero as n tends to infinity if $\alpha < 1$. Consequently, $N_H \ge 1$ whp if $0 \le \alpha < 1$, i.e., $G(n,p)$ contains at least one copy of $K_{2,2}$.
Thus, $\alpha_c = 1$.

3. (a) The star graph S_k has k nodes and $k-1$ edges, and so its edge density is $(k-1)/k$. Now consider any subset of m nodes from the vertex set of S_k. If this subset contains the hub, then the induced subgraph has $m-1$ edges, and the edge density is $(m-1)/m$. But this is no more than $(k-1)/k$ because $m \le k$. On the other hand, if the node subset does not contain the hub, then the induced subgraph contains no edges, and its edge density is zero. In either case, the edge density is no more than that of S_k. In other words, S_k contains no subgraph which is denser than S_k. Hence, it is balanced.

(b) By the results from lectures, the threshold for appearance of S_k is at $n^k p^{k-1} = 1$ (or a constant), i.e., when $n^{k-\alpha(k-1)} = 1$, i.e. $\alpha = k/(k-1)$. More precisely, if we take $\alpha_k = k/(k-1)$, then

$$\mathbb{P}(G(n, n^{-\alpha}) \supseteq S_k) \to \begin{cases} 0, & \text{if } \alpha > \alpha_k, \\ 1, & \text{if } \alpha < \alpha_k. \end{cases}$$

(c) Suppose that $d_{\max}+1$ colours are available. Consider the greedy algorithm, as described in the question. The algorithm isn't precise on what to do if several choices of colour are available when it comes to colouring a node. (Some choices will be ruled out by those of its neighbours that have already been assigned colours.) It is not important how such ties are broken. You could, for example, pick one of the permissible colours uniformly at random. Alternatively, you could pick the one that would appear first in a dictionary., or in your list of favourite colours. No matter how these choices are made, when it comes to colouring some node v, at most $deg(v)$ colours will be ruled out, one for each neighbour of v. In fact, some of these neighbours may have been assigned the same colour, so the number of colours ruled out could be smaller than $deg(v)$. But it cannot be bigger. In particular, at most $d_{\max}$ colours could have been ruled out. But $d_{\max}+1$ colours are available, so we can always find a colour that can be used to colour v.

(d) Note that the maximum degree of a graph is $d_{\max}$ if and only if it contains a copy of $S_{d_{\max}+1}$ but does not contain a copy of $S_{d_{\max}+2}$.
We know $\chi(G) \le d_{\max}+1$, so we can guarantee that $\chi(G) \le k$ if $d_{\max} \le k-1$. In other words, $\chi(G) \le k$ if G does not contain a copy of S_{k+1}. Hence, from the above, we can write

$$\mathbb{P}(\chi(G(n, n^{-\alpha}) \le k) \to 1 \text{ if } \alpha > \frac{k}{k-1}.$$

Note that we can only give an upper bound on the chromatic number using this approach, and cannot give a lower bound. To see this, suppose $G = S_n$. Then $d_{\max} = n-1$, but this graph can be coloured with just two colours: assign one of the colours to the hub, and the other to each of the leaves!