Problem 4: Prove Lemma 4.1. from the notes:

Solution (by Ferran Espuña): Let us recall the statement of the lemma: Let $\mathbb{G}_{n,p(n)}$ be the random graph where $p(n) = \frac{c}{n}$ and $c > 1$ is a constant. Let us fix a vertex v and start a BFS from v, exploring the graph in the order $x_1 = v$, $x_2 \in \mathcal{N}(v), \cdots$. Let A_k be the set of vertices in the BFS queue after k steps (that is, those that have appeared as neighbors of some vertex x_t for $t \leq k$ but are not themselves of the form x_t for some $t \leq k$). For example, $A_0 = \{v = x_1\}$, $A_1 = \mathcal{N}(x_1) \setminus \{x_1\} = \mathcal{N}(x_1)$, $A_2 = (\mathcal{N}(x_1) \cup \mathcal{N}(x_2)) \setminus \{x_1, x_2\}, \cdots$

Lemma. *With this notation, a.a.s., either:*

1. *the process stops before k^- steps, or*
2. *$|A_k| \geq \frac{c}{2}k$ for all $k \in [k^-, k^+]$ and thus the process survives until k^+.*

Where $k^- = \log(n)$ and $k^+ = n^{\frac{2}{3}}$.

Proof. We will assume condition 1 does not hold, and we will bound the probability of condition 2 *failing for the first time* at some $k \in [k^-, k^+]$. For this, note that condition 2 holding for all previous steps implies that the process has not stopped before k. Furthermore, there is a previous step $\hat{k}$ for which the set of vertices explored in the BFS between steps $\hat{k}$ and k is already predetermined at time $\hat{k}$, because they are all in $A_{\hat{k}}$. indeed, we know for all $k' \in [k^-, k]$ that $|A_{k'}| \geq \frac{c}{2}k'$. If we set $\frac{c}{2}\hat{k} \geq k - \hat{k}$, what we have just said is true. A valid choice for $\hat{k}$ is $\lceil \frac{2}{3}k \rceil \geq \frac{c+2}{2}k$. A different argument is needed for $\lceil \frac{2}{3}k \rceil < k^-$, but we will worry about that later.

We will assume a model in which the edges of the graph are created at random when the BFS algorithm needs them. That is, at step t, the edges from x_t are determined. The probability $p_{u,k}$ of a given vertex $u \notin \{x_1, \cdots x_k\}$ being in A_k bounded below by so of the probability $p_{u,k;\hat{k}}$ of it entering the BFS queue between steps $\hat{k}$ and k. This probability can be calculated right after step $\hat{k}$, and by symmetry is independent of everything explored previously. Because there are $k - \hat{k} \geq \frac{k+3}{3}$ chances for it to happen between $\hat{k}$ and k, we have that

$$p_{u,k;\hat{k}} \geq 1 - (1-p)^{\frac{k+3}{3}} \tag{1}$$

And because there are $n - k$ vertices not in $\{x_1, \cdots x_k\}$, we have that

$$|A_k| \geq \text{Binom}\left(n-k,\, 1 - (1-p)^{\frac{k+3}{3}}\right) \tag{2}$$

with expectation

$$\mathbb{E}(|A_k|) \geq (n-k)\left(1 - (1-p)^{\frac{k+3}{3}}\right) \tag{3}$$

To apply Chernoff's bound, we need to bound

$$f(k) := \frac{\frac{c-1}{2}k}{(n-k)\left(1 - (1-p)^{\frac{k+3}{3}}\right)} \tag{4}$$

As it turns out, $f(k)$ is increasing in k. Plugging in $k = k^+$, we get, as $n \to \infty$,

$$f(k) \leq \frac{(c-1)n^{\frac{2}{3}}}{2(n - n^{\frac{2}{3}})\left(1 - (1-\frac{c}{n})^{\frac{n^{2/3}}{4}}\right)} \leq \frac{c-1}{n^{\frac{1}{3}}\left(1 - e^{-\frac{cn^{-1/3}}{4}}\right)} \tag{5}$$

$\square$