**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 $t \le k$ but are not themselves of the form $x_t$ for some $t \le k$). 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) $|A_k| \ge \frac{c-1}{2}k$ *for all $k \in [k^-, k^+]$ and thus the process survives until $k^+$.*

*Where $k^- = M(c)\log(n)$ 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 $k$.

Let $Z_k$ be the set of vertices seen up to step $k$. That is, $Z_k = A_k \cup \{x_1, \cdots, x_k\}$ so that $|Z_k| = |A_k| + k$. We will shift our focus to $Z_k$ and bound the probability of $|A_k| < \frac{ck}{2} < \frac{c-1}{2}k + k$. At each step $k'$ up to $k$, because we have only conditioned on the total number of seen vertices being high, the number of vertices in $Z_k$ distributed at least as a binomial, where each vertex has $k$ independent chances of being seen. That is,

$$|Z_k| \ge \overline{Z_k} \sim \text{Binom}\left(n, 1 - (1-p)^k\right) \tag{1}$$

with expectation

$$\mathbb{E}\left(\overline{Z_k}\right) = n\left(1 - (1-p)^k\right) \tag{2}$$

To apply Chernoff's bound, we need to bound above

$$1 - \delta = f(k) := \frac{ck/2}{n\left(1-(1-p)^k\right)} = \frac{ck}{2n\left(1-(1-c/n)^k\right)} \le \frac{ck}{2n\left(1 - e^{-ck/n}\right)} \tag{3}$$

**Claim.** *This bound is increasing in $k$*

Plugging in $k = k^+$, we get, as $n \to \infty$,

$$f(k) \le \frac{cn^{2/3}}{2n\left(1 - e^{-cn^{-1/3}}\right)} \le \frac{cn^{-1/3}}{(2-\epsilon)\left(cn^{-1/3}\right)} = \frac{1}{2-\epsilon} < 1 \tag{4}$$

For some small $\epsilon > 0$. Therefore, the $\delta$ in Chernoff's bound is positive, and we can apply it:

$$\mathbb{P}\left(\overline{Z_k} < \frac{ck}{2}\right) = \mathbb{P}\left(\overline{Z_k} < (1-\delta)\mathbb{E}(\overline{Z_k})\right) \le e^{-\frac{\delta^2}{2}\mathbb{E}(\overline{Z_k})} \tag{5}$$

Finally, we will union bound over all $k \in [k^-, k^+]$. Because $k \ge k^-$, each term is at most

$$e^{-\frac{\delta^2}{2}\mathbb{E}(\overline{Z_{k^-}})} \le e^{-\frac{\delta^2 ck^-}{2}} \le e^{-\frac{c\delta^2 M(c)\log(n)}{2}} = n^{-\frac{cM(c)\delta^2}{2}} \tag{6}$$

So the probability of the condition failing at any point in $[k^-, k^+]$ is at most

$$n^{\frac{2}{3}} n^{-\frac{cM(c)\delta^2}{2}} = n^{\frac{2}{3} - \frac{cM(c)\delta^2}{2}} = n^{\frac{4-3cM(c)\delta^2}{6}} \tag{7}$$

For this to be $o(1)$, we need $4 - 3cM(c)\delta^2$ to be negative. To give a sufficient condition, take $\epsilon = \frac{1}{2}$ Then $\delta = 1 - \frac{2}{3} = \frac{1}{3}$. We need $4 - \frac{cM(c)}{3} < 0$, which is equivalent to $M(c) > \frac{12}{c}$. $\square$