CS294-179, Spring 2025 Homework #2 Due 2/14

1 Exponential growth of an epidemic for $R_0 > 1$: a branching process approximation [8 Points]

On many random graphs, the early phase of an SIR epidemic is well approximated by a branching process T_X , with $X \sim D$ describing the random number of people infected by a single individual, and $c = \mathbb{E}[X]$ taking the role of the basic reproduction number R_0 . It is basic folk knowledge, that in the early stages of an epidemic, $R_0 > 1$ implies exponential growth of the epidemic. In the branching process language, this means that Z_n , the number of off-spring in generation n, grows exponential in n.

As we have seen in the course, for c > 1, the branching process T_X has a positive survival probability $\theta = \Pr(|T_X| = \infty) > 0$, and in expectation, $\mathbb{E}[Z_n] = c^n$. However, this does not imply that T_X grows that fast with high probability. In fact, given that $\mathbb{E}[Z_n] = c^n$, it could quite well be that Z_n stays bounded with very high probability, and grows like, e.g., $(2c)^n$ with probability 2^{-n} . In this exercise, you will show that this does not happen, and that conditioned on survival, Z_n grows exponentially in n, with rate close to log c in the sense that

$$\frac{\log Z_n}{n} \to \log c \quad \text{in probability.}$$

More explicitly, you will show that for all $\epsilon > 0$,

$$\Pr\left(\left(c^{(1-\epsilon)n} \le Z_n \le (c^{(1+\epsilon)n} \mid |T_X| = \infty\right) \to 1, \text{ as } n \to \infty.\right.$$
(1)

To prove this, you will use a concentration inequality for i.i.d. random variables X_1, X_2, \ldots with distribution D_X , namely

$$\Pr\left(\sum_{i=1}^{\ell} X_i \le x\ell\right) \le e^{-\ell I(x)} \quad \text{if} \quad x < E[X] \tag{2}$$

where I(x) > 0 is the rate function for X. You derived this bound in Problem 3 of HW1 (with a formula for I(x) given in terms of a supremum). The strict positivity of I(x) for arbitrary distributions D_X and any $x < \mathbb{E}[X]$ has been established in Lecture 2.

(a) [1 Point]: Use Markov's inequality to prove an upper bound on the unconditional probability $\Pr(Z_n \ge c^{(1+\epsilon)n})$. Combine this bound with the fact that $\theta = \Pr(|T_X| = \infty) > 0$ to get an upper bound on the conditional probability $\Pr(Z_n \ge c^{(1+\epsilon)n} \mid |T_X| = \infty)$ and conclude that

$$\Pr\left(Z_n \ge c^{(1+\epsilon)n} \mid |T_X| = \infty\right) \to 0, \text{ as } n \to \infty.$$

(b) [2 Points]: Recall the definition of the number of active vertices Y_t from Lecture 2. In particular, recall that $Y_t > 0$ for all t if $|T_x| = \infty$ (survival), and that $Y_t = 1 + X_1 + \cdots + X_t - t$ if $Y_t > 0$. Fix $\delta > 0$ and an integer k_0 (which we will eventually chose to be δk) and use the bound (2) to show that

$$\Pr\left(\exists t \ge k_0 \text{ s.th. } Y_t \le (c-1-\delta)t \text{ and } |T_x| = \infty\right) \le \frac{e^{-I(c-\delta)k_0}}{1-e^{-I(c-\delta)}}$$

Conclude that conditioned on survival with high probability, i.e., "with probability tending to 1 as k_0 goes to infinity",

 $Y_t > (c - 1 - \delta)t \quad \text{for all} \quad t \ge k_0. \tag{3}$

(c) [2 Points] Recall the inductive definition of t_k from Lecture 5,

 $t_1 = 1$ and $t_{k+1} = t_k + Y_{t_k}$

illustrate this relationship for k = 1 and 2, reproduce the argument that $Y_{t_k} = Z_k$, and show that conditioned on survival, $t_k \ge k$.

(d) [2 Points]: Use what you have proven so far to conclude that conditioned on survival, with high probability (again, as $k_0 \to \infty$),

$$t_k \ge (c-\delta)^{k-k_0} t_{k_0} \ge (c-\delta)^{k-k_0} k_0$$
 and $Z_k = Y_{t_k} \ge (c-1-\delta)(c-\delta)^{k-k_0} k_0$ for all $k \ge k_0$.

(e) [1 Point]: Complete the proof of the lower bound on the growth of Z_k , by setting $k_0 = \lfloor \delta k \rfloor$ and showing that $\delta > 0$ can be chosen small enough to guarantee that for k large enough, the lower bound proven so far is at least $c^{(1-\epsilon)k}$

2 Cycles in G(n, p) [4 Points]

- (a) Let T be the number of triangles in G(n, p). Calculate $\mathbb{E}(T)$ as a function of n and p.
- (b) Show that there are $\frac{(k-1)!}{2}$ ways to have a cycle on k vertices. Then calculate the expected number of cycles of length k in G(n,p). For k = 3 this should agree with part (a).
- (c) Use Markov's inequality to show that the probability that G(n, p) containing any cycle goes to zero if $pn \to 0$. Hint: Write the expected number of cycles as a sum over the expected number of cycles of length k and show that for pn < 1 this expected number is bounded by $\frac{(pn)^3}{1-pn}$.

3 Threshold for the existence of isolated vertices [6 Points]

In class, we proved that for any random variable X with values in $\{0, 1, 2, \ldots\}$, we have

$$\Pr(X > 0) \leq \mathbb{E}[X] \tag{4}$$

$$\Pr(X > 0) \ge \frac{(\mathbb{E}[X])^2}{\mathbb{E}[X^2]}.$$
(5)

In this problem, we use these bounds to establish the threshold for the existence of isolated vertices in G(n, p). For this exercise, we will take X to be the number of isolated vertices in G(n, p)

- (a) Write X as $\sum_{i \in [n]} I_i$, where I_i is the indicator function that *i* is isolated, and show that $\mathbb{E}[X] = n(1-p)^{n-1}$ (Hint: a vertex *i* is isolated if none of the edges incident with *i* is occupied). Use the first moment bound in (4) that with probability tending to 1, G(n, p) has no isolated vertices if $p \ge c \frac{\log n}{n}$ and c > 1.
- (b) For the other direction, you will need to calculate the expectation of X^2 . Write this expectation as a double sum over $i, j \in [n]$. Treating the case i = j and the case $i \neq j$, separately, show that $\mathbb{E}[X^2] = \mathbb{E}[X] + n(n-1)(1-p)^{2n-3}$. Why is the exponent 2n-3, and not just 2(n-1)?
- (c) Use the second moment bound in Eq. (5) to show that with probability tending to 1, G(n, p) has at least one isolated vertex when $p \le c \frac{\log n}{n}$ and c < 1.