CS294-179 Network Structure and Epidemics
Fall 2020
Homework #4
Due Fr. 10/23

Each of the following exercises is worth 9 points. Choose 2, or try all 3, and the one where you get the least number of points will be a bonus exercise.

1 Convergence to Poisson

For two random variables X, Y defined over the same, discrete space Ω, the total variation distance, can be equivalently be defined as

$$d_{TV}(X,Y) = \sup_{A \subset \Omega} |\Pr(X \in A) - \Pr(Y \in A)|$$

$$d_{TV}(X,Y) = \frac{1}{2} \sum_{x \in \Omega} |\Pr(X = i) - \Pr(Y = i)|$$

and

$$d_{TV}(X,Y) = \inf_{\mathbb{P}} \mathbb{P}(X \neq Y)$$

where the infimum goes over couplings of X and Y.

1. Use the second definition above, to show that if $X \sim Be(p)$ and $Y \sim Poi(p)$, then $d_{TV}(X,Y) \leq p^2$. (If you don't quite get this, but a bound $O(p^2)$ for small p, that is fine as well).

2. Recalling that the sum of n $Poi(p)$ random variables has the distribution $Poi(np)$, use the result of (1) to show that $d_{TV}(Bin(n,p), Poi(np)) \leq np^2$.

3. Use (2) to prove that the degree distribution of $G(n,p)$ with $p = c/n$ for a constant c converges to $Poi(c)$ in the distance d_{TV}.

4. Consider the following version of the stochastic block model with k blocks and symmetric similarity matrix $B = B_{\alpha\beta}$: For each vertex $i = 1, \ldots, n$, chose a color $\alpha_i \in [k]$ i.i.d. uniformly at random, and then connect i and j with probability $p_{ij} = \frac{1}{n} B_{\alpha_i \alpha_j}$, independently for all $\binom{n}{2}$ pairs $\{i, j\}$.

   - Condition on the color of vertex 1 to be α, and write the degree d_1 of vertex 1 as sum of the form $\sum_{j=2}^{n} X_j$, where X_j takes values in $\{0, 1\}$.
   - Show that the X_j are i.i.d., calculate their expectation, and use this to write d_1 as $Bin(n-1, \tilde{p})$ for some $\tilde{p} = \tilde{c}/n$. *Hint: if X takes values 0 and 1, even if it's distribution looks very complicated, we know abstractly that it has some probability p' of being 1, and must then be 0 with probability $1-p'$. Thus we know its complicated distribution can be written as $Be(p')$ for some p'; in addition, p' must be equal to $\mathbb{E}[X]$.*
   - Use (2) to calculate the limiting distribution of d_1 as $n \to \infty$.

2 Graphical Sequences

A sequence $d_1, d_2, \dots, d_n$ of non-negative integers is called graphical if it is a degree sequence of a simple graph of size n. In this question we prove the following theorem by Havel and Hakimi.

Theorem 1 (Havel-Hakimi) *Let D be the sequence $n-1 \geq d_1 \geq d_2 \geq \dots \geq d_n > 0$ and $n \geq 2$. Let D' be the sequence obtained from D by discarding d_1, and, subtracting 1 from each of the next largest d_1 entries of D and then keeping the positive integers. So, D' is the sequence $(d_2 - 1, \dots d_{d_1+1} - 1, d_{d_1+2}, \dots d_n)$ after deleting all the zeros. The sequence D is graphical if and only if D' is graphical.*

1. First prove that if D' is graphical then D is graphical.
2. Next, assume D is graphical and let G be a graph with vertices $v_1, \dots, v_n$ and degree sequence D. Assume first v_1 is connected to all vertices v_i with $2 \leq i \leq d_1 + 1$. Show that under this condition, the sequence D' is graphical, i.e., construct a graph G' with the sequence D'.
3. Now assume that the condition in 2 is not satisfied, i.e., assume that there exists a vertex v_i with $2 \leq i \leq d_1 + 1$ that is not connected to v_1. Show that there is a vertex v_j with $j > d_1 + 1$ that is connected to v_1.
4. With part 3's notations, show that one can remove two edges from G and add two other edges to obtain another **simple** graph $\tilde{G}$, in which v_1 is connected to v_i but not v_j.
5. Show that by using the procedure described in 3 and 4 a finite number of times, you must end up in case 2 to prove the theorem.

3 Your friends have more friends than you do

For a graph G and a vertex i in $V(G)$, let d_i^* be the average degree of the neighbors of i, and let $d^*(G)$ to be the average of d_i^* over all vertices i. We will compare it to $\bar{d}(G)$, the average degree of G.

1. Let S_n be the star with n leaves, i.e., let the graph where one vertex, the center, is connected to n others, the leaves, and no other edges. Find $d_*(S_n)$ and $\bar{d}(S_n)$ as a function of n.
2. Let H be a graph that is the disjoint union of a triangle and an isolated edge. Compute $\bar{d}(H)$ and $d^*(H)$.
3. Prove that for any graph G, $d^*(G) \geq \bar{d}(G)$.
   Hint: as an intermediate step you might want to prove that $\frac{d_i}{d_j} + \frac{d_j}{d_i} \geq 2$ and the equality holds when $d_i = d_j$.
4. Characterize graphs for which the equality happens in part 3 ($\bar{d}(G) = d^*(G)$).
5. Express the statement "All your friends have exactly as many friends as you do" in mathematical terms, and prove it for the case that $\bar{d}(G) = d^*(G)$.