Exercises 'Random graphs and ranking'
Nelly Litvak

If numbers of Exercises or statements are mentioned, then they refer to the book 'Random Graphs and Complex Networks' by Remco van der Hofstad.

http://www.win.tue.nl/~rhofstad/NotesRGCN.pdf

Exercise 1 Consider the graph in Figure 1. Compute the following characteristic:

(a) the degree distribution (the distribution of the degree of a uniformly chosen vertex),

(b) the expected degree (of a uniformly chosen vertex),

(c) the distribution of the graph distance from node 1 to a randomly chosen other node,

(d) compute the expected number of friends of a random vertex in a random friendship, and compare this result to the result in (b).

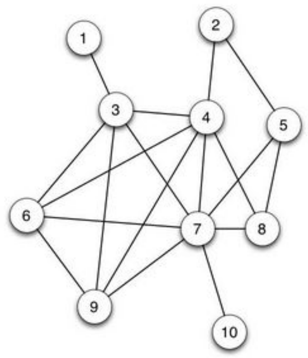

Figure 1: Graph of 10 nodes

Exercise 2 Let a graph G with vertices $\{1, 2, \ldots, n\}$ have connected components $C_1, C_2, \ldots, C_m$. That is, $\{C_i\}_{i=1}^m$ is a partition of $\{1, 2, \ldots, n\}$, every pair of vertices in C_i is connected, and no vertices in C_i are connected to vertices in C_j for $i \neq j$. Recall that H_n denotes the graph distance $\mathrm{dist}_G(U_1, U_2)$ between vertices U_1 and U_2 that are sampled uniformly and independently from $\{1, 2, \ldots, n\}$. Show that

$$\mathbb{P}\{H_n < \infty\} = \frac{1}{n^2} \sum\nolimits_{i=1}^{m} |C_i|^2,$$

where $|C_i|$ denotes the number of vertices in C_i.

Exercise 3 Let G be an undirected graph with vertices $\{1, 2, \ldots, n\}$ and edge set E. We write $\{i, j\} \in E$ if there is an edge between nodes i and j. Let d_i denote the degree of node i. Prove that

$$\sum\nolimits_{\{i,j\}\in E} (d_i^k + d_j^k) = \sum\nolimits_{i=1}^{n} d_i^{k+1}, \quad k \geq 0.$$

Exercise 4 Let a connected graph sequence $\{G_n\}_{n=1}^{\infty}$ have a bounded degree. That is, $d_{\max,n} := \max_{1\leq i\leq n} d_i \leq K$ and G_n is connected for every n. Show that, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left\{ H_n \leq (1-\varepsilon)\frac{\log n}{\log d_{\max,n}} \right\} = 0.$$

Exercise 5 (The friendship paradox) Consider an undirected graph $G = (V, E)$ with $V = \{1, 2, \ldots, n\}$. Let $\mathbf{d} = (d_v)_{v\in[n]}$ be the degree sequence of G. Let D_n be the degree of a randomly sample vertex. Next, let E^* be a corresponding set of directed edges: $(i, j), (j, i) \in E^*$ if and only if $\{i, i\} \in E$. Let (I, J) be a randomly chosen edge from E^*. Let D_n^* be the degree of a random vertex drown from $\{D_I, D_J\}$. Prove that

$$P(D_n^* = k) = \frac{k}{\mathbb{E}(D_n)} \mathbb{P}(D_n = k),$$

where

$$p_k = \frac{|\{i : d_i = k\}|}{n}, \quad \mathbb{E}(D_n) = \frac{1}{n}\sum\nolimits_{i=1}^{n} d_i.$$

Compare this to the result in Exercise 1(d).

(b) Let D_n be the degree of a randomly sampled vertex. Show that Theorem 1.1 follows directly from (a).

Exercise 6 (**Friendship paradox in Erdös-Renyi random graph**) The Erdös-Renyi random graph model creates an undirected random graph, denoted by $G(n,p)$. It is a graph of n vertices, and every possible pair of vertices has an edge between with probability p, independently of other edges.

(a) What is the degree distribution in $G(n,p)$?

(a) Let $\{I, J\}$ be a randomly chosen friendship from graph G generated by $G(n,p)$, and let D_I be the degree of I. Show that D_I is distributed as $X+1$, where $X \sim Binomial(n-2,p)$. Verify that this is consistent with the result in Exercise 5.

(b) Let D_n be the degree of a randomly chosen node in a graph generated by $G(n,p)$. Using (a), verify that $E(D_I) > E(D_n)$.

Exercise 7 (Another mathematical interpretation of "my friends have more friends than I do".) Consider a graph G with vertices $\{1, 2, \ldots, n\}$, edge set E and degree sequence $\{d_i\}_{i=1}^{n}$. Denote by I a uniformly chosen vertex, and by J — a vertex chosen uniformly from the friends of I. Prove that $\mathbb{E}d_J \geq \mathbb{E}d_I$.

Which interpretation, this or the one discussed in the lecture feels more natural to you?

Hint: First, by conditioning $\mathbb{E}[d_J|I = i]$, show that

$$\mathbb{E}d_J = \frac{1}{n}\sum\nolimits_{i=1}^{n}\sum\nolimits_{j=1}^{n} \mathbb{I}\{i \text{ and } j \text{ are connected by an edge}\}\frac{d_i}{d_j}.$$

Then use the inequality

$$\frac{1}{2}\left(\frac{d_j}{d_i} + \frac{d_i}{d_j}\right) \geq 1$$

to show that $\mathbb{E}d_J \geq \mathbb{E}d_I$.

Exercise 8. Consider a discrete power law random variable X, taking values $2, 3, \ldots$, such that $F(x) = P(X \leq x) = 1 - cx^{-\gamma}$, $x \geq x_0$, where $\gamma > 0$.

(a) Find c.

(b) Assume that $\gamma \in (1,2)$. Find $E(X)$. Argue that $Var(X) = \infty$.

(c) Let $X_1, X_2, \ldots, X_n$ be independent observations sampled from F. Provide a heuristic argument that the largest value in the sample is of the order of magnitude $n^{1/\gamma}$. Modify this argument to heuristically derive the order of magnitude for the k-th largest value, and observe how this expression depends on k.

Exercise 9. (a) (Exercise 7.3) Let $n = 2$, $d_1 = 2$ and $d_2 = 4$. Use the direct connection probabilities to show that the probability that $CM_n(\mathbf{d})$ consists of 3 self-loops equals 1/5.

Hint: Note that when $d_1 = 2$ and $d_2 = 4$, the graph $CM_n(\mathbf{d})$ consists only of self-loops precisely when the first half-edge of vertex 1 connects to the second half-edge of vertex 1.

(b) (Exercise 7.4) Let $n = 2$, $d_1 = 2$ and $d_2 = 4$. Use Proposition 7.7 to show that the probability that $CM_n(\mathbf{d})$ consists of 3 selfloops equals 1/5.

Exercise 10. (Exercise 7.9) Fix $CM_n(\mathbf{d})$ with degrees $\mathbf{d}$ given by $(d_i)_{i\in[n]}$, where $(d_i)_{i\in[n-1]}$ is an i.i.d. sequence of integer random variables and $d_n = d_n' + \mathbf{1}_{\{\updownarrow_{n-1}+d_n'\ odd\}}$, where d_n' has the same distribution as d_1 and is independent of $(d_i)_{i\in[n-1]}$. Show that Condition 7.8(a) holds, whereas Condition 7.8(b) and (c) hold when $\mathbb{E}[D] < \infty$ and $\mathbb{E}[D^2] < \infty$, respectively. Here the convergence is replaced with convergence in probability as explained in Remark 7.9.

Exercise 11. (a) (Exercise 7.19) Assume that the degree sequence $(d_i)_{i\in[n]}$ satisfies Conditions 7.8(a) – (c). Let T_n denote the number of triangles in $CM_n(\mathbf{d})$, i.e., the number of (i, s_i, t_i), (j, s_j, t_j), (k, s_k, t_k) such that $i < j < k$ and such that s_i is paired to t_j, s_j is paired to t_k and s_k is paired to t_i. Show that $(S_n, \tilde{M}_n, T_n)$ converges to three independent Poisson random variables and compute their asymptotic parameters.

(b) (Exercise 7.21) Assume that the fixed degree sequence $(d_i)_{i\in[n]}$ satisfies Conditions 7.8(a) – (c). Compute the number of simple graphs with degree sequence $(d_i)_{i\in[n]}$ not containing any triangle.

Hint: use Exercise 11(a) (Exercise 7.19).

Exercise 12. Consider a directed graph $G = (V, E)$. Let d_i^- and d_i^+ be, respectively, in- and out-degree of vertex $i \in [n]$. Denote by $\mathbf{r} = (r_1, r_2 \dots, r_n)$ the vector of PageRank values as in the original formula

$$r_i = \alpha \sum\nolimits_{j:(j,i)\in E} \frac{1}{d_j^+} r_j + (1-\alpha) q_i, \quad i \in [n] \tag{1}$$

where $\alpha \in (0, 1)$ and $(q_i)_{i\in[n]}$ is a probablity distribution: $\sum_{i=1}^n q_i = 1$, and $q_i \geq 0$, $i \in [n]$. Next, consider another version of the PageRank $\pi = (\pi_1, \pi_2 \dots, \pi_n)$, computed as follows:

$$\pi_i = \alpha \sum\nolimits_{j:(j,i)\in E} \frac{1}{d_j^+} r_j + \alpha \sum\nolimits_{j:d_j^+=0} q_i r_j + (1-\alpha) q_i \quad i \in [n]. \tag{2}$$

Note that π_i is a stationary distribution of a Markov chain.

(a) Describe the Markov chain, of which the stationary distribution is given by the system of linear equations (2).

(b) Show that $||\mathbf{r}|| \leq 1$, with the equality iff $|\{j : d_j^+ = 0\}| = 0$, i.e. there is no vertex with no outgoing edges (dangling nodes). Argue that in this case $\mathbf{r} = \pi$ is a stationary distribution of a Markov chain.

Hint: Use the matrix-vector representation of equation (1).

(c) Prove that π and $\mathbf{r}$ *always* give the same ranking (that is, even when dangling nodes are present). Moreover, $\pi = c\mathbf{r}$, where $c = 1 + \sum_{j:d_j^+=0} r_j$. Conclude that π and $\mathbf{r}$ are two equivalent definitions for PageRank.

Hint: Compare the solutions of the linear systems (1) and (1) in the matrix-vector form.

(d) Let $P_k = \{(i_1, i_2, \dots, i_k) : (i_l, i_{l+1}) \in E,\ l \in [k-1]\}$ be a set of paths of length k in G. Prove that

$$r_i = (1-\alpha) q_i \sum\nolimits_{k=1}^{\infty} \alpha^k \sum\nolimits_{p\in P_k : i_k = i} d_{i_1}^{-1} d_{i_2}^{-1} \cdots d_{i_{k-1}}^{-1}.$$

What can you say about the effect of verticies on distance k from i on r_i?