Homework 6

Name:

Student ID:

- Reasoning and work must be shown to gain partial/full credit

- Please include the cover-page on your homework PDF with your name and student ID. Failure of doing so is considered bad citizenship.

1. (1–4 points) **Lecture 11 Question**: You want to simulate the following network dynamics for the Wikispeedia graph. This dataset contains human navigation paths on Wikipedia, collected through the human-computation game Wikispeedia. In Wikispeedia, users are asked to navigate from a given source to a given target article, by only clicking Wikipedia links. A condensed version of Wikipedia (4,589 articles) is used.

   (a) (50% points) Ingest the graph as a **directed graph**. Evaluate the HITS centrality for this graph and comment on its meaning, and what differentiates this metric from other centrality measures. Identify the node with the highest hub value and the node with the highest authority value.

   (b) (50% points) Since the given graph is not strongly connected, the uniform random walk will not have a stationary distribution (i.e, the probability of visiting a node asymptotically depends on the initial state probabilities $\boldsymbol{\pi}(0)$).

   The algorithm used to calculate the **Page Rank** of webpages is a modified uniform random walk with an initial state probability for all the webpages equal to $1/N$. The transition probability is as follows:

   $$w_{ij} = \text{Probability of going from node } i \text{ to node } j$$

   $$= \begin{cases} \frac{1}{N}, & \mathcal{N}_i^{\text{out}} = \emptyset \\ \frac{\alpha}{N} + \frac{1-\alpha}{|\mathcal{N}_i^{\text{out}}|}, & \mathcal{N}_i^{\text{out}} \neq \emptyset, (i \to j) \in \mathcal{E} \\ \frac{\alpha}{N}, & \mathcal{N}_i^{\text{out}} \neq \emptyset, (i \to j) \notin \mathcal{E}, \end{cases}$$

   where α is some small value less than 1 (typically about 0.1), N is the number of nodes in the graph, and $\mathcal{N}_i^{\text{out}}$ is the set of out-neighbors of node i.

   Thus, the state probabilities evolve as follows:

   $$\boldsymbol{\pi}(t+1) = \boldsymbol{W}\boldsymbol{\pi}(t) = \frac{1}{N}\boldsymbol{W}^t\mathbf{1}$$

   The webpage rank produced by PageRank is the limit of the state probabilities, $\boldsymbol{\pi}(t)$, for a large $t \to \infty$. The limiting distribution is the left eigenvector of the probability transition matrix ($\boldsymbol{W}$) that corresponds to an eigenvalue of 1, i.e., $\boldsymbol{\pi}^{\star T} = \boldsymbol{\pi}^{\star T}\boldsymbol{W}$. Calculate the PageRank for the Wikispeedia network using this fact, and compare it to the pagerank obtained by using the networkx method `nx.pagerank`, by plotting them on a scatter plot.

2. (2 points) **Lecture 12 Questions (Randomized Gossiping)**:

   (a) (10% points) Generate a **connected** Random geometric graph with $N = 100$ nodes[1].

   (b) (90% points) Each node in the graph observes an instantiation of Bernoulli(p), where p is unknown to the nodes. The Randomized Gossiping algorithm allows the nodes to cooperatively estimate p. Do the following:

   - (10% points) Select $p = 0.2$, draw i.i.d samples from the distribution Bernoulli($p = 0.2$) and set that to $\boldsymbol{x}(0)$ for all $i \in [1, 2, \ldots, N]$.

---

[1] *Hint: Use `G = nx.random_geometric_graph(N, radius, seed)`. You may use any combination of radius and random seed to produce the graph. One such combination is when the radius is set to 0.225 and the random seed to 96845.*

- (50% points) Now, simulate the randomized "Gossiping" protocol over this graph for $T = 10,000$ steps to estimate p. The specific instance of randomized "Gossiping" is the one with random pairwise exchanges explained in Lecture 12, page 12. Basically, you pick at time t a random edge (i_t, j_t) and you average the two state variable as follows:

$$[\boldsymbol{x}(t+1)]_v = \begin{cases} \frac{x_{i_t}(t) + x_{j_t}(t)}{2} & i_t = v \text{ or } j_t = v \\ x_v(t) & \text{else} \end{cases}$$

- (10% points) Show the evolution of the states of all nodes over time (similar to the plots shown in Lecture 12, slide 16, at the bottom).

- (20% points) Explain why this method is able to estimate p.

(c) (5% bonus points) Calculate analytically $\overline{\boldsymbol{W}} = \mathbb{E}[\boldsymbol{W}(t)]$ and $\lambda_2(\overline{\boldsymbol{W}})$ based on the description of the statistics of the exchange.