Problem Set 3

March 6, 2020

Due date: Fri March 20, 2020 at midnight.

Remember: For any question you answer “I do not know” you get 20% of the grade associated with this question. A totally wrong answer gets 0.

You must prove anything you state unless it is obvious. If a question asks for an example of object X with property P, show that your example has property P.

- All solutions should be typed in LaTeX.
- Plots generated in python should be incorporated in the LaTeX file you submit.
- Your python code should also be submitted.

Exercise 1 (15 points)

1. (**5 points**) Show that for an ergodic Markov chain on n states, with transition matrix M that is symmetric the distribution $\pi = (1/n, \ldots, 1/n)$ is a stationary distribution.
2. (**5 points**) Show that for a simple random walk on a connected undirected graph $G = (V, E)$ a stationary probability distribution is π with $\pi(i) = \frac{d_i}{2|E|}$, where d_i is the degree of node i in G.
3. (**5 points**) Show that for asimple random walk on a weighted connected undirected graph $G = (V, E, w)$ the stationary probability distribution is π with $\pi(i) = \frac{\sum_{j \in N(i)} w(i,j)}{2\sum_{e(i,j)\in E} w(i,j)}$, where $N(i)$ are the neighbors of node i in G and $w(i, j)$ is the weight of edge $e(i, j) \in E$.

Exercise 2 (25 points) Assume the Web graph G and let it consist of n nodes, none of which is a sink node. Let P be the transition matrix of a simple random walk on this graph, and P' the transition matrix of the random walk that is enhanced with a uniform jump, from any node to any other node in G, with probability α. That is, $P' = \alpha P + (1 - \alpha)\mathbf{u}\mathbf{v}^T$, where $\mathbf{u}$ is a vector of all 1's and $\mathbf{v}$ is a uniform vector (all entries have value $1/n$).

1. (**10 points**) Show that since P is a stochastic matrix, then P' is also a stochastic matrix. That is, show that the sum of the entries in every row of P' is equal to 1.
2. (**15 points**) The stationary distribution $\mathbf{q}$ of the random walk described by P' can be computed by a vanilla power-method computation. More specifically, it can be done using the following iterative procedure, known as the power method:

- $\mathbf{q}^0 = \mathbf{v}$ // $\mathbf{v}$ is the uniform vector.
- $t = 1$
- **repeat**
  - $\mathbf{q}^t = \mathbf{q}^{t-1}(P')$
  - $\delta = ||\mathbf{q}^t - \mathbf{q}^{t-1}||$
  - $t = t + 1$
- **until** $\delta < \epsilon$

Note now that the computation of $\mathbf{y} = \mathbf{x}(P')$ can be done faster using the following procedure:

1. $\mathbf{y} = \alpha \mathbf{x} P^T$
2. $\beta = ||\mathbf{x}||_1 - ||\mathbf{y}||_1$
3. $\mathbf{y} = \mathbf{y} + \beta \mathbf{v}$

Prove that this procedure indeed computes $\mathbf{y} = \mathbf{x}P'$. Also discuss the computational gains of using this procedure instead of the vanilla power method algorithm.

Exercise 3 (20 points) This problem shows some technical problems with a stationary probability distribution that may arise if one is not careful.

- (**10 points**) Give an example of an aperiodic Markov chain with more than one stationary distribution. Show they are both stationary distributions.
- (**10 points**) Give an example of an irreducible Markov chain with transition matrix M and an initial probability distribution $\mathbf{x}$, such that $\lim_{t\to\infty} xM^t$ does not exist.

Exercise 4 (20 points) Using the dataset available here: `https://www.cise.ufl.edu/research/sparse/matrices/SNAP/p2p-Gnutella09.html` compute the PageRank score of the nodes in the input graph using python. More specifically you are asked to do the following:

- (**5 points**) In python implement the simple power method we discussed in class and is also described in Exercise 2.
- (**5 points**) In python implement the "clever" power method described in Exercise 2 above.
- (**10 points**) Plot scatterplots of the PageRank scores obtained from either of the algorithms with the degree of the nodes. For your implementation and for the scatterplots use values for $\alpha = 0.2, 0.4, 0.6$ and 0.8.

Exercise 5 (20 points) This problem shows the rate of convergence of the stationary-distribution computation for a symmetric stochastic matrix is exponential (aka very very fast) in the number of steps t. Let M be an $n \times n$ symmetric stochastic matrix and π its stationary distribution, which is the uniform distribution: i.e., $\pi = (1/n, \ldots, 1/n)$.

- (**5 points**) Show that if $\mathbf{e_1}, \ldots, \mathbf{e_n}$ are eigenvectors of M with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$, then $\mathbf{e_1}, \ldots, \mathbf{e_n}$ are also eigenvectors of M^t with corresponding eigenvalues $\lambda_1^t, \ldots, \lambda_n^t$.
- (**15 points**) Recall that when thinking about the computation of the stationary distribution (e.g., via the power method) we are interested in how well $\mathbf{x}M^t$ approximates the stationary probability distribution π. In other words, we are interested in the value of $||\mathbf{x}M^t - \pi||_2$. Given that $\mathbf{x}$ is a linear combination of the eigenvectors, i.e., $\mathbf{x} = \sum_{i\leq n} \alpha_i \mathbf{e_i}$, with $\alpha_i = \mathbf{x}\mathbf{e_i}^T$ (careful, this could be negative!), show that $||\mathbf{x}M^t - \pi||_2 \leq \sqrt{n}\lambda_2^t$. Don't forget that $||\mathbf{y}||_2 = (\mathbf{y}\mathbf{y}^T)^{1/2}$ for any row vector $\mathbf{y}$.

Hint: to show $\alpha_i \leq 1$, you will likely want to use the Cauchy-Schwarz inequality for inner product: $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}\mathbf{y}^\mathbf{T} \leq ||\mathbf{x}||_2 ||\mathbf{y}||_2$.