A Mathematical Introduction to Data Science **November 25, 2014**

Homework 5. Perron-Frobenius and Fiedler Theory

Instructor: Yuan Yao *Due: Tuesday November 18, 2014*

The problem below marked by * is optional with bonus credits.

1. *PageRank*: The following dataset contains Chinese (mainland) University Weblink during 12/2001-1/2002,

   `http://www.math.pku.edu.cn/teachers/yaoy/Fall2011/univ_cn.mat`

   where `rank_cn` is the research ranking of universities in that year, `univ_cn` contains the webpages of universities, and `W_cn` is the link matrix from university i to j.

   (a) Compute PageRank with Google's hyperparameter $\alpha = 0.85$;

   (b) Compute HITS authority and hub ranking using SVD of the link matrix;

   (c) Compare these rankings against the research ranking (you may consider Kendall's τ distance – as the number of pairwise mismatches between two orders – to compare different rankings);

   (d) Compute extended PageRank with various hyperparameters $\alpha \in (0, 1)$, investigate its effect on ranking stability.

   For your reference, an implementation of PageRank and HITs can be found at

   `http://www.math.pku.edu.cn/teachers/yaoy/Fall2011/pagerank.m`

2. *Perron Theorem:* Assume that $A > 0$. Consider the following optimization problem:

$$\begin{aligned} & \max \delta \\ s.t.\quad & Ax \geq \delta x \\ & x \geq 0 \\ & x \neq 0. \end{aligned}$$

   Let λ^* be optimal value with $\nu^* \geq 0, \quad 1^T \nu^* = 1$, and $A\nu^* \geq \lambda^* \nu^*$. Show that

   (a) $A\nu^* = \lambda^* \nu^*$, i.e. (λ^*, ν^*) is an eigenvalue-eigenvector pair of A;

   (b) $\nu^* > 0$;

   *(c) λ^* is unique and ν^* is unique;

*(d) For other eigenvalue λ $(\lambda z = Az \quad when \quad z \neq 0)$, $|\lambda| < \lambda^*$.

3. *Absorbing Markov Chain:*

   Let P be a row Markov matrix on $n+1$ states with non-absorbing state $\{1, \ldots, n\}$ and absorbing state $n+1$. Then P can be partitioned into

   $$P = \begin{bmatrix} Q & R \\ 0 & 1 \end{bmatrix}$$

   Assume that Q is primitive. Let $N(i,j)$ be the expected number of jumps starting from nonabsorbent state i and hitting state j, before reaching the absorbing state $n+1$. Show that

   (a) $N(i,i) = 1 + \sum_k N(i,k)Q(k,i)$, for $i = 1, \ldots, n$;

   (b) $N(i,j) = \sum_k N(i,k)Q(k,j)$, for $i \neq j$;

   (c) These identities together imply that $N = (I - Q)^{-1}$, called the fundamental matrix;

   (d) Show that the probability of absorption from state i, $B(i)$ $(i = 1 \ldots, n)$, is given by $B = NR$.