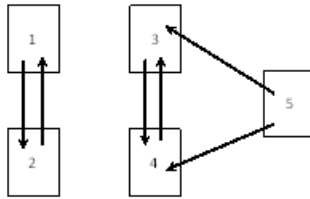


# 1 Challenges to the algorithm

In this worksheet, we explore some challenges to the PageRank algorithm.

## 1.1 Nonuniqueness

We would like our ranking to be unique, which means that we should have only one eigenvector representing the eigenvalue 1. It turns out that this is true if the web is a strongly connected digraph. We will show this later. However, if the web is disconnected, then we can have a higher dimensional eigenspace for eigenvalue 1. Consider the web in BL Figure 2.2, reproduced here:



The link matrix is

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \frac{1}{2} \\ 0 & 0 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

1) Show that the vectors  $[\frac{1}{2}, \frac{1}{2}, 0, 0, 0]^T$  and  $[0, 0, \frac{1}{2}, \frac{1}{2}, 0]^T$  are both eigenvectors with eigenvalue 1.

2) Show that any linear combination of these are eigenvectors with eigenvalue 1, and so we have vectors like  $[\frac{3}{8}, \frac{3}{8}, \frac{1}{8}, \frac{1}{8}, 0]^T$  as well as  $[\frac{1}{8}, \frac{1}{8}, \frac{3}{8}, \frac{3}{8}, 0]^T$ , which would give different rankings!

3) We will show the following:

**Proposition 1** *Let  $W$  be a web with  $r$  components  $W_1, W_2, \dots, W_r$ . Then the eigenspace of the eigenvalue 1 is at least  $r$ -dimensional.*

3A) Show that if we label the web by assigning the vertices in  $W_1$  first, then the vertices in  $W_2$ , etc., then the link matrix will have a block diagonal form like

$$A = \begin{bmatrix} A_1 & 0 & 0 & 0 \\ 0 & A_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & A_r \end{bmatrix},$$

where  $A_k$  is the link matrix for the web  $W_k$ .

3B) Show that if each is  $A_k$  column stochastic, each has an eigenvector  $v_k$  with eigenvalue 1, and that can be expanded into a eigenvector  $w_k$  for  $A$  by letting

$$w_1 = \begin{pmatrix} v_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad w_2 = \begin{pmatrix} 0 \\ v_2 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

etc. We can show that each of these is linearly independent and part of the eigenspace  $V_1$  of eigenvalue 1.

## 2 Dealing with disconnected webs

Let  $S$  be the matrix  $n \times n$  matrix with all entries  $1/n$ . Notice that this matrix is column stochastic. In terms of probabilities, this matrix represents equal probabilities of jumping to any page on the web (including the one you are already on).

4) Show that if  $A$  is a column stochastic matrix, then

$$M = (1 - m)A + mS$$

is column stochastic for all values of  $m$  between zero and one. Supposedly the original value for  $m$  used by Google was 0.15.

We will show that the matrix  $M$  has a one-dimensional eigenspace  $V_1(M)$  for the eigenvalue 1 as long as  $m > 0$ . Note that the matrix  $M$  has all positive entries. This motivates:

**Definition 2** A matrix  $M = (M_{ij})$  is positive if  $M_{ij} > 0$  for every  $i$  and  $j$ .

**Proposition 3** If  $M$  is a positive, column-stochastic matrix, then any eigenvector with eigenvalue 1 is a scalar multiple of a vector with all positive entries.

In the interest, we are omitting the proof, though it is in the notes and in BL and uses only elementary mathematics.

5) Explain why the matrix  $M$  can be used to produce unique rankings if the link matrix  $A$  is column stochastic.

## 2.1 Dangling nodes

**Definition 4** A dangling node is a vertex in the web with outdegree zero (i.e., with no links).

6) Show that a dangling node produces a column of zeroes in the link matrix. This means that the resulting link matrix is not column-stochastic, since some columns may sum to zero. This means that we may not use our theorem that 1 is an eigenvalue. In fact, it may not be true.

For dangling nodes, we have the following theorem of Perron:

**Theorem 5** If  $A$  is a positive matrix, then  $A$  contains a real, positive eigenvalue  $\rho$  such that

1. For any other eigenvalue  $\lambda$ , we have  $|\lambda| < \rho$  (recall that  $\lambda$  could be complex).
2. The eigenspace of  $\rho$  is one-dimensional and there is a unique eigenvector  $x = [x_1, x_2, \dots, x_p]^T$  with eigenvalue  $\rho$  such that  $x_i > 0$  for all  $i$  and

$$\sum_{i=1}^p x_i = 1.$$

This eigenvector is called the *Perron vector*. Thus, if we had a matrix with all positive entries, as we got in the last section, we can use the Perron vector as the ranking.

7) Explain how Perron's theorem can be used to give a unique ranking even if there are dangling nodes.