An internet search engine must perform two tasks: (i) decide which pages on the net are relevant in some way as responses to your query, and, (ii) produce a ranking of the relevant pages in order of importance. In this lecture we consider the second task: ranking by importance. It is reasonable to assume that each page j has an *importance coefficient* x_j which is independent of the query q, and a *relevance coefficient* r_{qj} which depends upon the query. Thus in response to a query q, we may perhaps present pages to the user in non-increasing order of $x_j \cdot r_{qj}$. In this lecture we will not worry about relevance, but merely importance.

One heuristic for this is: a page is as important as the number of other pages pointing to it. But then, clearly we should not count just how many pages point to a page, but give more weightage to whether the pages pointing to you are themselves pointed to by other pages, and so on. Also, perhaps if a page points to many pages, then it is not recommending anyone page very strongly, so its recommendation should be taken less seriously. These concerns can be summarized by saying that each x_j satisfy the following recurrence:

$$x_j = \sum_{i|i \text{ points to } j} x_i/d_{i+}$$

We are assuming in this that each page i has positive outdegree d_{i+}.

It is convenient to define a matrix W with $w_{ij} = 1/d_{i+}$ if $(i,j) \in E$, and 0 otherwise where the web is represented as a directed graph $G = (V, E)$. Then if x denotes the vector with components x_j, then the above recurrence is equivalent to:

$$x = xW$$

Thus we are asking that x be a left eigenvector of W of eigenvalue 1. W is called the *walk* matrix of G, why is explained shortly. In the rest of this lecture, we will see that under reasonable assumptions this requirement is satisfied nicely.

Example: Say our graph has vertex set $\{1, 2, 3\}$ and (directed) edge set

$\{(1,2),(1,3),(2,1),(3,2)\}$. Then its walk matrix is:

$$W = \begin{pmatrix} 0 & 0.5 & 0.5 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

Simply writing $x^T = x^T W$ will not enable us to solve for x; clearly if x is a solution so is any multiple. So we will assert that the sum of the entries of x be 1. With this additional equation, we should get the required eigenvector, (0.4 0.4 0.2). Even in this example, you should get some intuition: 1,2 in some ways are more important than 3: 2 is pointed to by 1,3, and 1 is pointed to by an important node 2.

1 Eigenvalues of a stochastic matrix

A matrix P is right stochastic if its entries are non-negative real numbers and rows (left if columns) add up to 1. Clearly, W defined above is right stochastic. The notion of stochasticity arises from the context of performing a random walk on a graph with n vertices, where entries p_{ij} represent probabilities of transiting from state i to state j. Our matrix W can indeed be interpreted in this manner; if x gives the current probability distribution of where the walker is, then xW gives the probabilities for the next step. The condition $xW = x$ is equivalent to asking for a stationary distribution. A stochastic matrix is simply a generalized walk matrix; we allow the walk to have different probabilities for each outgoing edge, and also allow self loops.

Although we are concerned with left eigenvectors and eigenvalues of P, it is useful to note the following for any right stochastic matrix:

$$P\mathbf{1} = \mathbf{1}$$

In this equation, by $\mathbf{1}$ we mean the vector consisting of all 1s. Equality clearly holds because rows sum to 1.

Note now that the left and right eigenvalues of any matrix M are the same (but not necessarily the eigenvectors). This is because the expression $xM = \lambda x$ maybe written as $x(M - \lambda I) = 0$, which requires that $M - \lambda I$ be singular, and thus the eigenvalues λ be solutions to $|M - \lambda I| = 0$. This is a polynomial in λ and the same polynomial results if we consider the right eigenvalues, because the determinant of a matrix is the same as that of its transpose.

In fact, right multiplication by a right stochastic matrix has a very simple interpretation. For $y = Wx$, $y_i = \sum_j w_{ij} x_j$. Thus each y_i may be viewed as a weighted average of x, because $\sum_j w_{ij} = 1$. We will work with right multiplication wherever possible because of this reason, say when we only want to consider properties of eigenvalues and not eigenvectors.

So we know that 1 is a (left) eigenvalue of any right stochastic matrix P. The question that now arises is: does only one eigenvector have eigenvalue 1, or are there many eigenvectors (to within scaling) with eigenvalue 1? In other words, does the eigenvalue have multiplicity 1?

Lemma 1 *If a graph G is strongly connected, then its walk matrix W has largest eigenvalue 1, of multiplicity 1.*

Proof: Since each term of Wx is a weighted average of the values in x, Thus the maximum entry of Wx cannot be bigger (or smaller) than the maximum (or minimum) entry in x. Thus all eigenvalues must be at most 1. With slightly more work, you should be able to show that the eigenvalues cannot be smaller than -1 either (exercise).

We next show that $\mathbf{1}$ is the only right eigenvector for eigenvalue 1. Suppose some other x is also an eigenvector of value 1. An eigenvector remains an eigenvector after uniform scaling, so we assume the largest entry in x_i is 1. But x_i is a weighted average of the values all its out neighbours, and these values are themselves at most 1. The only way the average can be 1 is if the values are all 1. So all out neighbours must have $x_v = 1$. But this applies to all their out neighbours, and so on to all the vertices since the graph is strongly connected. But then we have established that $x = \mathbf{1}$ as well. ■

Lemma 2 *For the walk matrix W of a strongly connected graph G, the unique (to within scaling) left eigenvector of eigenvalue 1 has all positive entries.*

A plausible line of proof is: suppose I start a random walk in the graph from some vertex. I keep walking, and eventually in the limit as the number of steps goes to infinity, the probability distribution should tend to a limit – and that limit would have to be a stationary distribution x, i.e. satisfying $x = xW$. Further, since the graph is strongly connected, every vertex should have a nonzero probability of being visited from any other vertex, and thus every vertex must appear with a positive probability in the stationary distribution.

Unfortunately, in general, a random walk in a graph may not tend to a stationary distribution. Consider just a directed 2 cycle. If the walk starts with the distribution (p, q) at step 0, then at all even time steps the distribution will be just this, and at all odd time steps, the distribution will be (q, p). So there is no single limit as the number of steps increases! The problem clearly is that our walk has a periodicity. If we could construct a matrix W' which has the same eigenvectors as W but which does not have this periodicity, perhaps we would be done. This indeed works out.

Proof: Consider the matrix $W^* = \frac{1}{n}\sum_{i=0}^{n-1} W^i$. Let x denote the left eigenvector of W of eigenvalue 1. Clearly x is a left eigenvector of eigenvalue 1 of any W^k. Thus $xW^* = \frac{1}{n}\sum_{i=0}^{n-1} x = x$. Hence x is also a left eigenvector of eigenvalue 1 of W^*. It should also be clear tht W^* is also stochastic.

We next establish that every entry w_{ij} of W^* is positive. Since G is strongly connected, there must be a path from i to j of length $k \leq n - 1$. Thus there is positive probability of visiting j in k steps starting at i. Thus the ij entry of W^k must be positive, and this positiveness is preserved in W^*.

Note that x cannot be identically zero, so after multiplying by -1 (which still keeps an eigenvector) there must be at least one $x_i > 0$. Suppose x_+, x_-, x_0 denote the subvectors of x with positive, negative and 0 values. Multiplying a vector x by a stochastic matrix causes x_i at any node to be sent to its out neighbours. Each neighbour receives such contributions and adds them up. Now the total contribution leaving the nodes in x_+ is $\sum_{i \in x_+} x_i$. But this total contribution leaves x_+ since every vertex has an edge to every vertex in W^*. The contribution received in x_+ from x_-, x_0 is non-negative. Thus at the end of the step $\sum_{i \in x_+} x_i$ must reduce. But this is not possible because x has eigenvalue 1. Thus if some x_i is positive, then all x_i must be positive.[1] ■

Note that while W might have been periodic, W^* is not. Indeed it is possible to show that starting from any initial distribution x, $x(W^*)^t$ tends to the eigenvector with eigenvalue 1.

[1] A different proof: Suppose x is not sign uniform. Since $x = xW^*$ we have $x_j = \sum_i w^*_{ij} x_i$. Since $w^*_{ij} > 0$, and x not sign uniform, it follows that $|x_j| < \sum_j w^*_{ij}|x_i|$. Thus $x' < x'W^*$ where x' is the vector obtained from x by taking absolute value of every entry. Now $x' \cdot \mathbf{1} < x'W^* \cdot \mathbf{1}$. But $\mathbf{1}$ is a right eigenvector, and hence $x' \cdot \mathbf{1} < x' \cdot \mathbf{1}$ and we have a contradiction. Thus, all elements of x' must have the same sign, say positive, or be zero.

But now suppose $x_j = 0$. Some such j must have an edge from i where $x_i > 0$. But this is not possible since $x_j = \sum_i x_i/d_{i+} \geq x_i/d_{i+} > 0$. Which proof do you like better?

2 The real Page rank algorithm

Our assumption of the graph being a strongly connected component does not hold for the web graph. It can be in several components, and there can be important pages with no outdegree. There may also be pages with no indegree, or pages all of whose in neighbours have zero indegree.

The original pagerank algorithm proposed by the founders of google handled this with an interesting variation. They augment the random walk model described above with "random restarts". The random walk can be thought of as modelling a user who keeps following random links from the page he is currently at. In the new model, at each step, the user tosses a biased coin. If a head appears, which say happens with probability α, then the user picks a completely random page on the web and jumps to that. If a tail appears, i.e. with probability $1-\alpha$ picks one of the outgoing links from the current page with equal probability. The recommended value is $\alpha = 0.15$. So if Y is the transition matrix of this new walk, we have:

$$\begin{aligned} y_{ij} &= \alpha/n && \text{if } (i,j) \notin G \\ &= \alpha/n + (1-\alpha)w_{ij} && \text{if } (i,j) \in G \end{aligned}$$

where W is as defined earlier. Define

$$Y = (1-\alpha)W + \alpha J/n$$

where J is a $n \times n$ matrix of all 1s. Then the importance coefficient x is simply the left eigenvector of eigenvalue 1 of Y, i.e. $x = xY$. Notice that Y already has all entries positive, and is stochastic. Thus it will have a unique eigenvector of eigenvalue 1. But once we know that uniqueness is guaranteed, we can solve for x more directly. Noting that $xJ = \mathbf{1}$, we get

$$x = xY = (1-\alpha)xW + (\alpha/n)\mathbf{1}$$

or alternatively

$$x(I - (1-\alpha)W) = (\alpha/n)\mathbf{1}$$

Thus we simply need to solve this linear system. The matrix inverse can be approximated, noting that $(I-M)^{-1} = \sum_{i=0}^{\infty} M^i$, as long as the sum converges. Thus we have

$$(I - (1-\alpha)W)^{-1} = \sum_{i=0}^{\infty} (1-\alpha)^t W^t$$

Since $(1-\alpha) = .85$ by our choice, the terms become small very rapidly. So we find an approximate inverse Z by taking just a few terms of the series, and then set $x = (\alpha/n)\mathbf{1}Z$. The approximation requires a few multiplications with the sparse matrix W, and will run in time $O(n)$ if the average degree is constant assuming the matrix is represented suitably. Performing the full Gaussian Elimination will invariably introduce fill-in, making the matrix dense, and thus requiring $O(n^3)$ time.

Exercises

1. Which of the following matrices are stochastic? Which are walk matrices? Draw the graphs associated with those that are. For all find the left and right eignvectors and eigenvalues.

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix}, \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$$

2. Suppose W is a walk matrix of some strongly connected graph. Is W^k also necessarily a walk matrix of a strongly connected graph? Is it also stochastic?

3. Argue that $W - I$ for a strongly connected graph must have rank $n-1$.

4. Suppose G consists of k disjoint strongly connected components. Show that W has eigenvalue 1 with multiplicity k.

5. Complete the arguement that the smallest eigenvalue of W cannot be smaller than -1.

6. Consider an $n \times n$ stochastic matrix W in which each $w_{ij} > \epsilon > 0$, for some constant ϵ. Assume that $w_{ii} > \epsilon$ as well. Let x be a (row) vector of average value 0, i.e. $x \cdot \mathbf{1} = 0$. For any vector x, Let $P(x) = \sum_{i|x_i>0} x_i$. Let $x(t) = xW^t$.

   (i) Show that $P(x(t+1)) \leq P(x(t))(1 - n\epsilon)$. Hint: How much positive charge is annihilated in each step?

   (ii) Suppose y is a probability (row) vector. Show that yW^t tends to z where z is the stationary distribution of W. Hint: Consider $x = y - z$.