Eigenvalues and Eigenvectors

Refs

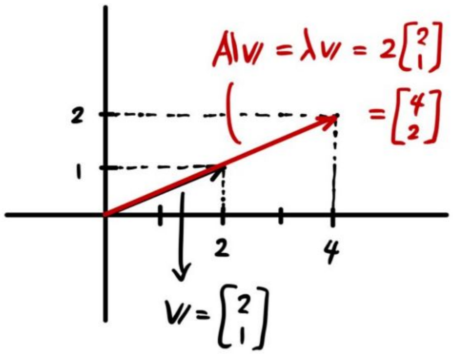
- https://medium.com/sho-jp/linear-algebra-part-6-eigenvalues-and-eigenvectors-35365dc4365a

Eigenvectors are the vectors that does not change its orientation when multiplied by the transition matrix, but it just scales by a factor of corresponding eigenvalues.

$$Av = \lambda v$$

$$\begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 2+2 \\ 6-4 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 2\begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$= \lambda_1 v$$



In diagonal Form:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_m \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix}$$

Steps to Obtain Eigen Vectors

$$A = \begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix}$$

Step 1

$$\det(A - \lambda I) = 0$$

$$\downarrow$$

$$\begin{vmatrix} 1-\lambda & 2 \\ 3 & -4-\lambda \end{vmatrix} = (1-\lambda)(-4-\lambda) - 2 \cdot 3$$

$$= -4 - \lambda + 4\lambda + \lambda^2 - 6$$

$$= \lambda^2 + 3\lambda - 10$$

$$= (\lambda - 2)(\lambda + 5) = 0$$

$$\therefore \lambda_1 = 2, \ \lambda_2 = -5$$

Step 2

(i) $\lambda_1 = 2$

$$(A - \lambda I)v = 0$$

$$\downarrow$$

$$\begin{bmatrix} 1-\lambda_1 & 2 \\ 3 & -4-\lambda_1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\downarrow$$

$$\begin{bmatrix} -1 & 2 \\ 3 & -6 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Nullspace!

$$\downarrow$$

$$\begin{cases} -v_1 + 2v_2 = 0 \\ 3v_1 - 6v_2 = 0 \end{cases}$$

$$\downarrow$$

$$\therefore v_1 = 2, \ v_2 = 1$$

(ii) $\lambda_2 = -5$

$$\begin{bmatrix} 1-\lambda_2 & 2 \\ 3 & -4-\lambda_2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\downarrow$$

$$\begin{bmatrix} 6 & 2 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\downarrow$$

$$\begin{cases} 6v_1 + 2v_2 = 0 \\ 3v_1 + v_2 = 0 \end{cases}$$

$$\downarrow$$

$$\therefore v_1 = -1, \ v_2 = 3$$

Power Method

The Power Method is **used to find a dominant eigenvalue (one having the largest absolute value), if one exists, and a corresponding eigenvector**.

Eigenvalues of an matrix A are obtained by solving its characteristic equation:

$$\lambda^n + c_{n-1}\lambda^{n-1} + c_{n-2}\lambda^{n-2} + \cdots + c_0 = 0.$$

- For large values of n, polynomial equations like this one are difficult and time-consuming to solve.

- Moreover, numerical techniques for approximating roots of polynomial equations of high degree are sensitive to rounding errors.

Alternative method:

- find the eigenvalue of A that is largest in absolute value—we call this eigenvalue the

  dominant eigenvalue of A.
- Although this restriction may seem severe, dominant eigenvalues are of primary interest in many physical applications.

Dominant Eigenvalue and Dominant Eigenvector

Let $\lambda_1, \lambda_2, \ldots,$ and λ_n be the eigenvalues of an $n \times n$ matrix A. λ_1 is called the **dominant eigenvalue** of A if

$$|\lambda_1| > |\lambda_i|, \quad i = 2, \ldots, n.$$

The eigenvectors corresponding to λ_1 are called **dominant eigenvectors** of A.

Not every matrix has a dominant eigenvalue. For instance, the matrix

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

(with eigenvalues of $\lambda_1 = 1$ and $\lambda_2 = -1$) has no dominant eigenvalue. Similarly, the matrix

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

(with eigenvalues of $\lambda_1 = 2$, $\lambda_2 = 2$, and $\lambda_3 = 1$) has no dominant eigenvalue.

The power method for approximating eigenvalues is iterative:

- First we assume that the matrix A has a dominant eigenvalue with corresponding dominant eigenvectors.
- Then we choose an initial approximation of one of the dominant eigenvectors of A. This initial approximation must be a nonzero vector in Rn

$$
\begin{aligned}
\mathbf{x}_1 &= A\mathbf{x}_0 \\
\mathbf{x}_2 &= A\mathbf{x}_1 = A(A\mathbf{x}_0) = A^2\mathbf{x}_0 \\
\mathbf{x}_3 &= A\mathbf{x}_2 = A(A^2\mathbf{x}_0) = A^3\mathbf{x}_0 \\
&\vdots \\
\mathbf{x}_k &= A\mathbf{x}_{k-1} = A(A^{k-1}\mathbf{x}_0) = A^k\mathbf{x}_0.
\end{aligned}
$$

For large powers of k, and by properly scaling this sequence, we will see that we obtain
a good approximation of the dominant eigenvector of A.

Approximating a Dominant Eigenvector by the Power Method

Complete six iterations of the power method to approximate a dominant eigenvector of

$$A = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}.$$

We begin with an initial nonzero approximation of

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

We then obtain the following approximations.

Iteration		Approximation
$\mathbf{x}_1 = A\mathbf{x}_0 = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -10 \\ -4 \end{bmatrix}$	$\rightarrow$	$-4\begin{bmatrix} 2.50 \\ 1.00 \end{bmatrix}$
$\mathbf{x}_2 = A\mathbf{x}_1 = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} -10 \\ -4 \end{bmatrix} = \begin{bmatrix} 28 \\ 10 \end{bmatrix}$	$\rightarrow$	$10\begin{bmatrix} 2.80 \\ 1.00 \end{bmatrix}$
$\mathbf{x}_3 = A\mathbf{x}_2 = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 28 \\ 10 \end{bmatrix} = \begin{bmatrix} -64 \\ -22 \end{bmatrix}$	$\rightarrow$	$-22\begin{bmatrix} 2.91 \\ 1.00 \end{bmatrix}$
$\mathbf{x}_4 = A\mathbf{x}_3 = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} -64 \\ -22 \end{bmatrix} = \begin{bmatrix} 136 \\ 46 \end{bmatrix}$	$\rightarrow$	$46\begin{bmatrix} 2.96 \\ 1.00 \end{bmatrix}$
$\mathbf{x}_5 = A\mathbf{x}_4 = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 136 \\ 46 \end{bmatrix} = \begin{bmatrix} -280 \\ -94 \end{bmatrix}$	$\rightarrow$	$-94\begin{bmatrix} 2.98 \\ 1.00 \end{bmatrix}$
$\mathbf{x}_6 = A\mathbf{x}_5 = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} -280 \\ -94 \end{bmatrix} = \begin{bmatrix} 568 \\ 190 \end{bmatrix}$	$\rightarrow$	$190\begin{bmatrix} 2.99 \\ 1.00 \end{bmatrix}$

From the characteristic polynomial of A:

$$\lambda^2 + 3\lambda + 2 = (\lambda + 1)(\lambda + 2).$$

Eigen values -1 and -2,

The dominant eigenvector: $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$

Eigenvalue from Eigenvector

Since $\mathbf{x}$ is an eigenvector of A, we know that $A\mathbf{x} = \lambda\mathbf{x}$, and we can write

$$\frac{A\mathbf{x} \cdot \mathbf{x}}{\mathbf{x} \cdot \mathbf{x}} = \frac{\lambda\mathbf{x} \cdot \mathbf{x}}{\mathbf{x} \cdot \mathbf{x}} = \frac{\lambda(\mathbf{x} \cdot \mathbf{x})}{\mathbf{x} \cdot \mathbf{x}} = \lambda.$$

$$A\mathbf{x} = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix} \begin{bmatrix} 2.99 \\ 1.00 \end{bmatrix} = \begin{bmatrix} -6.02 \\ -2.01 \end{bmatrix}$$

Then, since

$$A\mathbf{x} \cdot \mathbf{x} = (-6.02)(2.99) + (-2.01)(1) \approx -20.0$$

and

$$\mathbf{x} \cdot \mathbf{x} = (2.99)(2.99) + (1)(1) \approx 9.94,$$

we compute the Rayleigh quotient to be

$$\lambda = \frac{A\mathbf{x} \cdot \mathbf{x}}{\mathbf{x} \cdot \mathbf{x}} \approx \frac{-20.0}{9.94} \approx -2.01,$$

which is a good approximation of the dominant eigenvalue $\lambda = -2$.

We begin with an initial nonzero approximation of

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

We then obtain the following approximations.

Iteration		*Approximation*
$\mathbf{x}_1 = A\mathbf{x}_0 = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -10 \\ -4 \end{bmatrix}$	$\rightarrow$	$-4\begin{bmatrix} 2.50 \\ 1.00 \end{bmatrix}$
$\mathbf{x}_2 = A\mathbf{x}_1 = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} -10 \\ -4 \end{bmatrix} = \begin{bmatrix} 28 \\ 10 \end{bmatrix}$	$\rightarrow$	$10\begin{bmatrix} 2.80 \\ 1.00 \end{bmatrix}$
$\mathbf{x}_3 = A\mathbf{x}_2 = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 28 \\ 10 \end{bmatrix} = \begin{bmatrix} -64 \\ -22 \end{bmatrix}$	$\rightarrow$	$-22\begin{bmatrix} 2.91 \\ 1.00 \end{bmatrix}$
$\mathbf{x}_4 = A\mathbf{x}_3 = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} -64 \\ -22 \end{bmatrix} = \begin{bmatrix} 136 \\ 46 \end{bmatrix}$	$\rightarrow$	$46\begin{bmatrix} 2.96 \\ 1.00 \end{bmatrix}$
$\mathbf{x}_5 = A\mathbf{x}_4 = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 136 \\ 46 \end{bmatrix} = \begin{bmatrix} -280 \\ -94 \end{bmatrix}$	$\rightarrow$	$-94\begin{bmatrix} 2.98 \\ 1.00 \end{bmatrix}$
$\mathbf{x}_6 = A\mathbf{x}_5 = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} -280 \\ -94 \end{bmatrix} = \begin{bmatrix} 568 \\ 190 \end{bmatrix}$	$\rightarrow$	$190\begin{bmatrix} 2.99 \\ 1.00 \end{bmatrix}$

power method tends to produce approximations

with large entries. In practice it is best to “scale down” each approximation before pro-

The Power Method with Scaling

Calculate seven iterations of the power method with *scaling* to approximate a dominant eigenvector of the matrix

$$A = \begin{bmatrix} 1 & 2 & 0 \\ -2 & 1 & 2 \\ 1 & 3 & 1 \end{bmatrix}.$$

Use $\mathbf{x}_0 = (1, 1, 1)$ as the initial approximation.

One iteration of the power method produces

$$A\mathbf{x}_0 = \begin{bmatrix} 1 & 2 & 0 \\ -2 & 1 & 2 \\ 1 & 3 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix},$$

and by scaling we obtain the approximation

$$\mathbf{x}_1 = \frac{1}{5}\begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix} = \begin{bmatrix} 0.60 \\ 0.20 \\ 1.00 \end{bmatrix}.$$

A second iteration yields

$$A\mathbf{x}_1 = \begin{bmatrix} 1 & 2 & 0 \\ -2 & 1 & 2 \\ 1 & 3 & 1 \end{bmatrix}\begin{bmatrix} 0.60 \\ 0.20 \\ 1.00 \end{bmatrix} = \begin{bmatrix} 1.00 \\ 1.00 \\ 2.20 \end{bmatrix}$$

and

$$\mathbf{x}_2 = \frac{1}{2.20}\begin{bmatrix} 1.00 \\ 1.00 \\ 2.20 \end{bmatrix} = \begin{bmatrix} 0.45 \\ 0.45 \\ 1.00 \end{bmatrix}.$$

$\mathbf{x}_0$	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$	$\mathbf{x}_6$	$\mathbf{x}_7$
$\begin{bmatrix} 1.00 \\ 1.00 \\ 1.00 \end{bmatrix}$	$\begin{bmatrix} 0.60 \\ 0.20 \\ 1.00 \end{bmatrix}$	$\begin{bmatrix} 0.45 \\ 0.45 \\ 1.00 \end{bmatrix}$	$\begin{bmatrix} 0.48 \\ 0.55 \\ 1.00 \end{bmatrix}$	$\begin{bmatrix} 0.51 \\ 0.51 \\ 1.00 \end{bmatrix}$	$\begin{bmatrix} 0.50 \\ 0.49 \\ 1.00 \end{bmatrix}$	$\begin{bmatrix} 0.50 \\ 0.50 \\ 1.00 \end{bmatrix}$	$\begin{bmatrix} 0.50 \\ 0.50 \\ 1.00 \end{bmatrix}$

Convergence of Power Method

If A is an $n \times n$ diagonalizable matrix with a dominant eigenvalue, then there exists a nonzero vector $\mathbf{x}_0$ such that the sequence of vectors given by

$$A\mathbf{x}_0,\quad A^2\mathbf{x}_0,\quad A^3\mathbf{x}_0,\quad A^4\mathbf{x}_0,\quad \ldots,\quad A^k\mathbf{x}_0,\quad \ldots$$

approaches a multiple of the dominant eigenvector of A.

Proof

- Since A is diagonalizable, it has n linearly independent eigenvectors with corresponding eigenvalues.
- These eigenvalues are ordered so that is the dominant eigenvalue (with a corresponding eigenvector of x1).

- Because the n eigenvectors are linearly independent, they must form a basis for Rn

- For the initial approximation x0, we choose a nonzero vector with linear combination of those eigen vectors

$$\mathbf{x}_0 = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_n\mathbf{x}_n$$

has nonzero leading coefficients. (If $c_1 = 0$, the power method may not converge, and a different $\mathbf{x}_0$ must be used as the initial approximation. Now, multiplying both sides of this equation by A produces

$$\begin{aligned}
A\mathbf{x}_0 &= A(c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_n\mathbf{x}_n) \\
&= c_1(A\mathbf{x}_1) + c_2(A\mathbf{x}_2) + \cdots + c_n(A\mathbf{x}_n) \\
&= c_1(\lambda_1\mathbf{x}_1) + c_2(\lambda_2\mathbf{x}_2) + \cdots + c_n(\lambda_n\mathbf{x}_n).
\end{aligned}$$

Repeated multiplication of both sides of this equation by A produces

$$A^k\mathbf{x}_0 = c_1(\lambda_1{}^k\mathbf{x}_1) + c_2(\lambda_2{}^k\mathbf{x}_2) + \cdots + c_n(\lambda_n{}^k\mathbf{x}_n),$$

which implies that

$$A^k\mathbf{x}_0 = \lambda_1{}^k\left[c_1\mathbf{x}_1 + c_2\left(\frac{\lambda_2}{\lambda_1}\right)^k\mathbf{x}_2 + \cdots + c_n\left(\frac{\lambda_n}{\lambda_1}\right)^k\mathbf{x}_n\right].$$

Now, from our original assumption that λ_1 is larger in absolute value than the other eigenvalues it follows that each of the fractions

$$\frac{\lambda_2}{\lambda_1},\quad \frac{\lambda_3}{\lambda_1},\quad \ldots,\quad \frac{\lambda_n}{\lambda_1}$$

is less than 1 in absolute value. Therefore each of the factors

$$\left(\frac{\lambda_2}{\lambda_1}\right)^k,\quad \left(\frac{\lambda_3}{\lambda_1}\right)^k,\quad \ldots,\quad \left(\frac{\lambda_n}{\lambda_1}\right)^k$$

must approach 0 as k approaches infinity. This implies that the approximation

$$A^k\mathbf{x}_0 \approx \lambda_1{}^k c_1\mathbf{x}_1,\quad c_1 \neq 0$$

improves as k increases. Since $\mathbf{x}_1$ is a dominant eigenvector, it follows that any scalar multiple of $\mathbf{x}_1$ is also a dominant eigenvector. Thus we have shown that $A^k\mathbf{x}_0$ approaches a multiple of the dominant eigenvector of A.

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_n|,$$

then the power method will converge quickly if $|\lambda_2| / |\lambda_1|$ is small, and slowly if $|\lambda_2| / |\lambda_1|$ is close to 1. This principle is illustrated in Example 5.

The Rate of Convergence

(a) The matrix

$$A = \begin{bmatrix} 4 & 5 \\ 6 & 5 \end{bmatrix}$$

has eigenvalues of $\lambda_1 = 10$ and $\lambda_2 = -1$. Thus the ratio $|\lambda_2| / |\lambda_1|$ is 0.1. For this matrix, only four iterations are required to obtain successive approximations that agree when rounded to three significant digits. (See Table 10.7.)

TABLE 10.7

$\mathbf{x}_0$	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
$\begin{bmatrix} 1.000 \\ 1.000 \end{bmatrix}$	$\begin{bmatrix} 0.818 \\ 1.000 \end{bmatrix}$	$\begin{bmatrix} 0.835 \\ 1.000 \end{bmatrix}$	$\begin{bmatrix} 0.833 \\ 1.000 \end{bmatrix}$	$\begin{bmatrix} 0.833 \\ 1.000 \end{bmatrix}$

(b) The matrix

$$A = \begin{bmatrix} -4 & 10 \\ 7 & 5 \end{bmatrix}$$

has eigenvalues of $\lambda_1 = 10$ and $\lambda_2 = -9$. For this matrix, the ratio $|\lambda_2| / |\lambda_1|$ is 0.9, and the power method does not produce successive approximations that agree to three significant digits until sixty-eight iterations have been performed, as shown in Table 10.8.

TABLE 10.8

$\mathbf{x}_0$	$\mathbf{x}_1$	$\mathbf{x}_2$		$\mathbf{x}_{66}$	$\mathbf{x}_{67}$	$\mathbf{x}_{68}$
$\begin{bmatrix} 1.000 \\ 1.000 \end{bmatrix}$	$\begin{bmatrix} 0.500 \\ 1.000 \end{bmatrix}$	$\begin{bmatrix} 0.941 \\ 1.000 \end{bmatrix}$	$\begin{matrix} \cdots \\ \cdots \end{matrix}$	$\begin{bmatrix} 0.715 \\ 1.000 \end{bmatrix}$	$\begin{bmatrix} 0.714 \\ 1.000 \end{bmatrix}$	$\begin{bmatrix} 0.714 \\ 1.000 \end{bmatrix}$

Page Rank Search

- http://pi.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture3/lecture3.html

Just open your favorite search engine, like Google, AltaVista, Yahoo, type in the key words, and the search engine will display the pages relevant for your search. But how does a search engine really work?

At first glance, it seems reasonable to imagine that what a search engine does is to keep an index of all web pages, and when a user types in a query search, the engine browses through its index and counts the occurrences of the key words in each web file. The winners are the pages with the highest number of occurrences of the key words. These get displayed back to the user.

First Level Solution: Early 90s

Text based ranking systems

- it seems reasonable to imagine that what a search engine does is to keep an index of all web pages
- when a user types in a query search, the engine browses through its index and counts the occurrences of the key words in each web file.
- The winners are the pages with the highest number of occurrences of the key words. These get displayed back to the user.

Limitations

A search about a common term such as "Internet" was problematic. The first page displayed by one of the early search engines was written in Chinese, with repeated occurrences of the word "Internet" and containing no other information about the Internet.

Moreover, suppose we wanted to find some information about Cornell. We type in the word "Cornell" and expect that "www.cornell.edu" would be the most relevant site to our query. However there may be millions of pages on the web using the world Cornell, and www.cornell.edu may not be the one that uses it most often.

Suppose spammers decided to write a web site that contains the word "Cornell" a billion times and nothing else. Would it then make sense for our web site to be the first one displayed by a search engine?

The usefulness of a search engine depends on the *relevance* of the result set it gives back not the count.

Page Rank algorithm

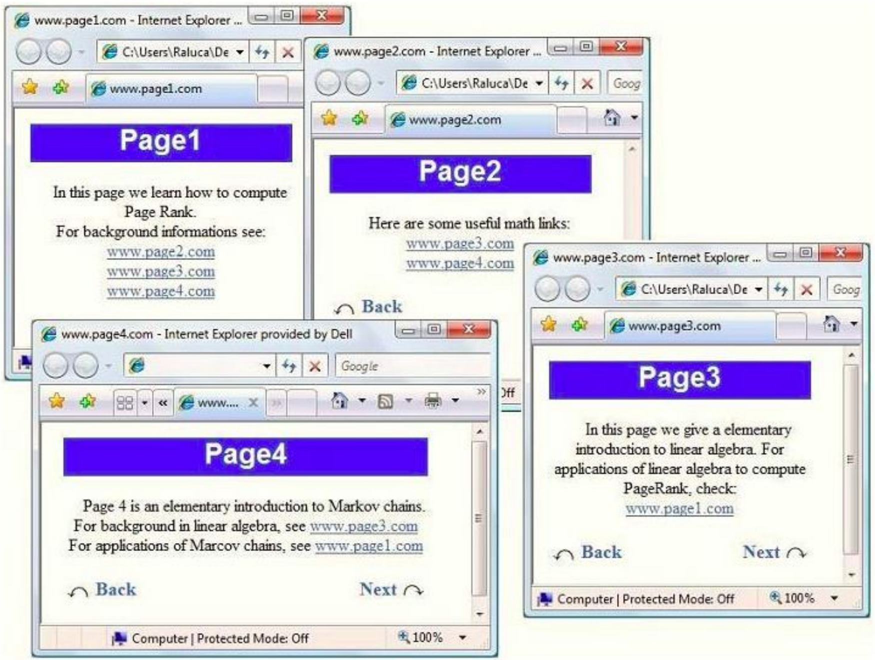
invented by Larry Page and Sergey Brin

became a Google trademark in 1998

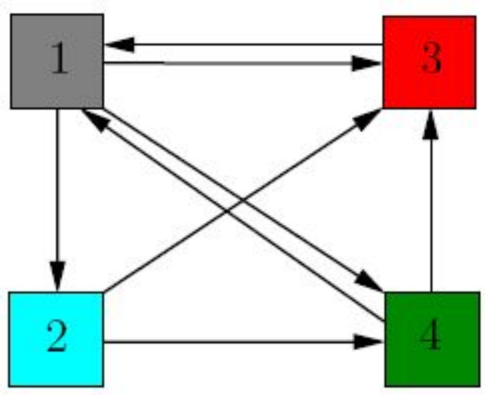
Basic idea:

- the importance of any web page can be judged by looking at the pages that link to it.
- If we create a web page i and include a hyperlink to the web page j, this means that we consider j important and relevant for our topic.
- If there are a lot of pages that link to *j*, this means that the common belief is that page *j* is important. If on the other hand, j has only one backlink, but that comes from an authoritative site *k*, (like www.google.com, www.cnn.com, www.cornell.edu) we say that *k* transfers its authority to *j*; in other words, *k* asserts that *j* is important. Whether we talk about popularity or authority, we can iteratively assign a rank to each web page, based on the ranks of the pages that point to it.

Consider small Internet consisting of just 4 web sites:

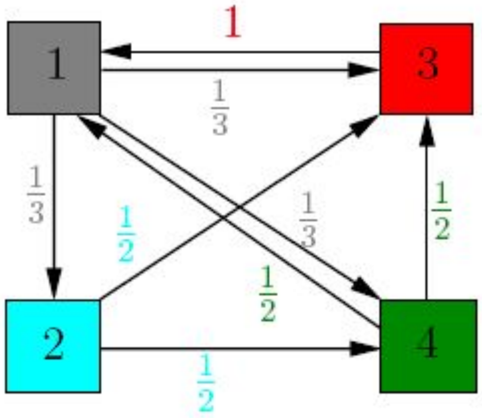


- "translate" the picture into a directed graph with 4 nodes, one for each web site.
- When web site i references j, we add a directed edge between node i and node j in the graph.
- For the purpose of computing their page rank, we ignore any navigational links such as back, next buttons, as we only care about the connections between different web sites.



Weighted Assignment

- Each page should transfer evenly its importance to the pages that it links to
- . Node 1 has 3 outgoing edges, so it will pass on of its importance to each of the other 3 nodes.
- Node 3 has only one outgoing edge, so it will pass on all of its importance to node 1.
- In general, if a node has k outgoing edges, it will pass on of its importance to each of the nodes that it links to. Let us better visualize the process by assigning weights to each edge.



Let us denote by A the transition matrix of the graph, $A = \begin{array}{cc} & \text{from} \\ \text{to} & \begin{bmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix} \end{array}$

What are the importance values of these pages?

Dynamical systems point of view:

- Suppose that initially the importance is uniformly distributed among the 4 nodes, each getting ¼.
- Denote by v the initial rank vector, having all entries equal to ¼.

Steps:

- Each incoming link increases the importance of a web page.
- So at step 1, we update the rank of each page by adding to the current value the importance of the incoming links.
- This is the same as multiplying the matrix A with v .
- At step 1, the new importance vector is $v_1 = Av$.
- Iterate the process.

$$\mathbf{v} = \begin{pmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{pmatrix}, \quad \mathbf{Av} = \begin{pmatrix} 0.37 \\ 0.08 \\ 0.33 \\ 0.20 \end{pmatrix}, \quad \mathbf{A}^2\,\mathbf{v} = \mathbf{A}\,(\mathbf{Av}) = \mathbf{A}\begin{pmatrix} 0.37 \\ 0.08 \\ 0.33 \\ 0.20 \end{pmatrix} = \begin{pmatrix} 0.43 \\ 0.12 \\ 0.27 \\ 0.16 \end{pmatrix}$$

$$\mathbf{A}^3\,\mathbf{v} = \begin{pmatrix} 0.35 \\ 0.14 \\ 0.29 \\ 0.20 \end{pmatrix}, \quad \mathbf{A}^4\,\mathbf{v} = \begin{pmatrix} 0.39 \\ 0.11 \\ 0.29 \\ 0.19 \end{pmatrix}, \quad \mathbf{A}^5\,\mathbf{v} = \begin{pmatrix} 0.39 \\ 0.13 \\ 0.28 \\ 0.19 \end{pmatrix}$$

$$\mathbf{A}^6\,\mathbf{v} = \begin{pmatrix} 0.38 \\ 0.13 \\ 0.29 \\ 0.19 \end{pmatrix}, \quad \mathbf{A}^7\,\mathbf{v} = \begin{pmatrix} 0.38 \\ 0.12 \\ 0.29 \\ 0.19 \end{pmatrix}, \quad \mathbf{A}^8\,\mathbf{v} = \begin{pmatrix} 0.38 \\ 0.12 \\ 0.29 \\ 0.19 \end{pmatrix}$$

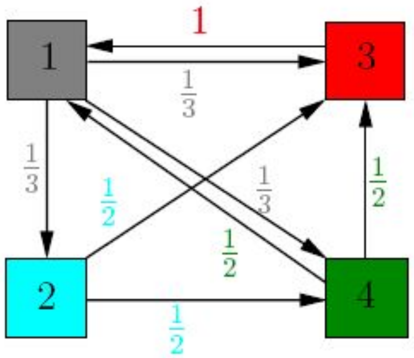
We notice that the sequences of iterates $v, Av, \ldots, A^k v$ tends to the equilibrium value $v^* = \begin{pmatrix} 0.38 \\ 0.12 \\ 0.29 \\ 0.19 \end{pmatrix}$. We call this the PageRank vector of our web graph.

Linear algebra point of view

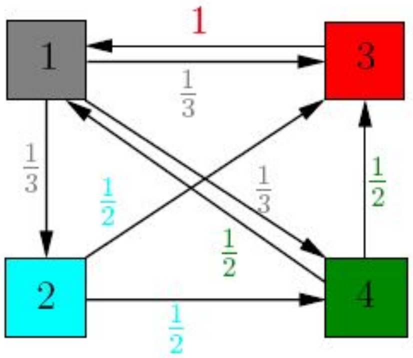
Let us denote by x_1, x_2, x_3, and x_4 the importance of the four pages. Analyzing the situation at each node we get the system:

$$\begin{cases} x_1 = 1 \cdot x_3 + \frac{1}{2} \cdot x_4 \\ x_2 = \frac{1}{3} \cdot x_1 \\ x_3 = \frac{1}{3} \cdot x_1 + \frac{1}{2} \cdot x_2 + \frac{1}{2} \cdot x_4 \\ x_4 = \frac{1}{3} \cdot x_1 + \frac{1}{2} \cdot x_2 \end{cases}$$

This is equivalent to asking for the solutions of the equations $A \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$.



Linear algebra point of view

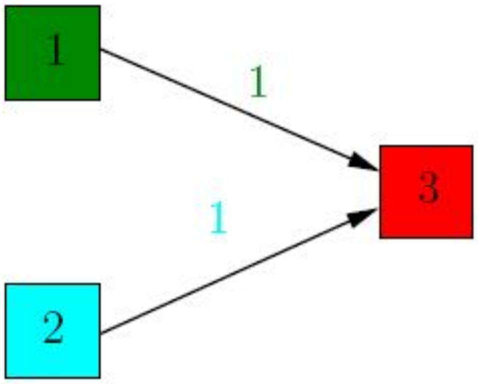


This is equivalent to asking for the solutions of the equations $A \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$.

Choose v^* to be the unique eigenvector with the sum of all entries equal to 1. (We will sometimes refer to it as the probabilistic eigenvector corresponding to the eigenvalue 1).

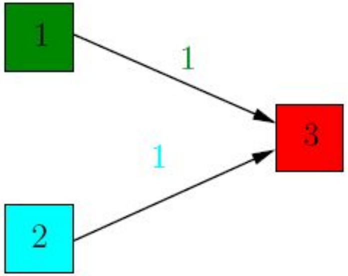
Identify the challenge with dangling node

Compute the rank of 3 pages:



Identify the challenge with dangling node

Nodes with no outgoing edges (dangling nodes)



We iteratively compute the rank of the 3 pages:

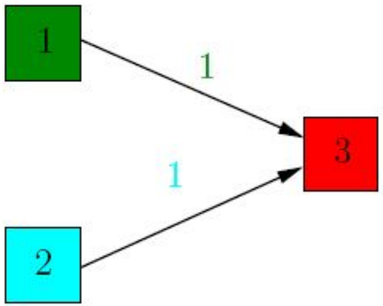
$$v_0 = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}, \quad v_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \frac{2}{3} \end{bmatrix}, \quad v_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ \frac{2}{3} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

So in this case the rank of every page is 0. This is counterintuitive, as page 3 has 2 incoming links, so it must have some importance!

replace with [1/n,/1n...]

Identify the challenge with dangling node

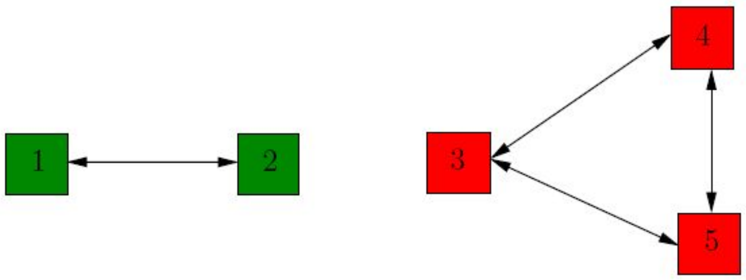
Compute the rank of 3 pages:



Solution:

An easy fix for this problem would be to replace the column corresponding to the dangling node 3 with a column vector with all entries 1/3. In this way, the importance of node 3 would be equally redistributed among the other nodes of the graph, instead of being lost.

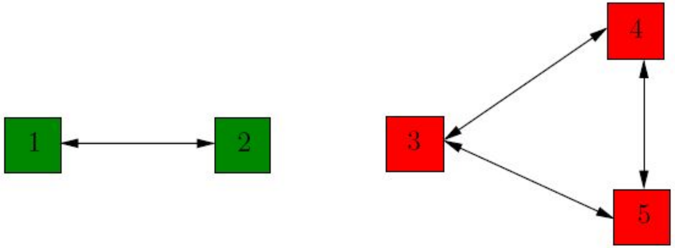
Disconnected components



A random surfer that starts in the first connected component has no way of getting to web page 5 since the nodes 1 and 2 have no links to node 5 that he can follow.

Disconnected components

The transition matrix for this graph is $A = \left[\begin{array}{cc|ccc} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{array}\right]$



u and v both are eigenvectors of eigenvalue 1, so that is confusing

$$: v = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad u = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

The solution of Page and Brin:

In order to overcome these problems, fix a positive constant p between 0 and 1, which we call the damping factor (a typical value for p is 0.15). Define the Page Rank matrix (also known as the Google matrix) of the graph by $M = (1-p) \cdot A + p \cdot B$ where $B = \frac{1}{n} \cdot \begin{bmatrix} 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix}$.

The matrix M models the random surfer model as follows: most of the time, a surfer will follow links from a page: from a page i the surfer will follow the outgoing links and move on to one of the neighbors of i. A smaller, but positive percentage of the time, the surfer will dump the current page and choose arbitrarily a different page from the web and "teleport" there. The damping factor p reflects the probability that the surfer quits the current page and "teleports" to a new one. Since he/she can teleport to any web page, each page has $\frac{1}{n}$ probability to be chosen. This justifies the structure of the matrix B.

Assignment

. Prove that M remains a column stochastic matrix. Prove that M has only positive entries.

Problem 2. Redo the computations for the Page Rank with the transition matrix A replaced with the matrix M, for the graphs representing the Dangling nodes, respectively Disconnected components. Do the problems mentioned in there still occur?

To Conclude

Fact: The PageRank vector for a web graph with transition matrix A, and damping factor p, is the unique probabilistic eigenvector of the matrix M, corresponding to the eigenvalue 1.

- **From the mathematical point of view, once we have M, computing the eigenvectors corresponding to the eigenvalue 1 is, at least in theory, a straightforward task. Solve the system Ax = x! But when the matrix M has size 30 billion (as it does for the real Web graph), even mathematical software such as Matlab or Mathematica are clearly overwhelmed.**

- **computing the probabilistic eigenvector corresponding to the eigenvalue 1 - use Power Method**