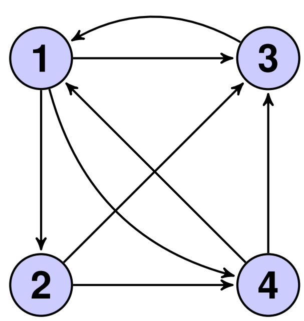
Page Rank: Recap

Recall that Google's Page Rank algorithm uses a web crawler to estimate the popularity of webpages. We assume that the more popular a website, the more incoming links it has. We use the following simulation to model the problem: suppose we have a large population of web surfers scattered at random around the internet. At each time step, all surfers randomly choose an outgoing link on their current page to arrive at a new webpage. After running the simulation for a large number of time steps, we expect the number of viewers on each page to reflect that page's popularity.

It's important to realize that this is a simplified model that leaves out certain characteristics of the real world: for example, people do not click links at random, not everyone visits outgoing links at once, browsing sessions vary in length, and people often seek out entirely new pages over clicking outgoing links from their current page. Page Rank as it's used in practice, while still a simplified model, has many more intricacies than we cover here. However, even this first order approximation is a very powerful tool.

Let's revisit the example from last week of an internet consisting of four webpages:



Let's say we started off with 6 people on each webpage. Then one time step ticks and everyone randomly selects an outgoing link to visit. How many people do we expect to be on each webpage now?

Page 1: Page 3 only links to page 1, so all 6 people currently on 3 will move to 1. Page 4 links to pages 1 and 3, so on average about half the people on page 4 will go to page 1. There are no other incoming links, so we expect page 1 to have about $6 \times 1 + 6 \times \frac{1}{2} = 9$ people in the next time step.

Page 2: Page 1 is the only page that links to page 2, with a probability of $\frac{1}{3}$. We expect about a third of people from page 1 to move to page 2, so we expect page 2 to have $6 \times \frac{1}{3} = 2$ people in the next time step.

Page 3: We can arrive at page 3 from all the other pages. At page 1, we move to page 3 with probability $\frac{1}{3}$, at page 2 we move to page 3 with probability $\frac{1}{2}$, and at page 4 we move to page 3 with probability $\frac{1}{2}$. We expect $6 \times \frac{1}{3} + 6 \times \frac{1}{2} + 6 \times \frac{1}{2} = 8$ people to be on page 3 in the next time step.

Page 4: We can arrive at page 4 from either page 1 or page 2. If we're at page 1, we move to page 4 with probability $\frac{1}{3}$; if we're at page 2, we move to 4 with probability $\frac{1}{2}$. We expect $6 \times \frac{1}{3} + 6 \times \frac{1}{2} = 5$ people to be on page 4 in the next time step.

Notice that to find the number of people we expect to be on page i, we added up the number of people we expected would move from every other page to page i. In general, if we have n pages, $x_i(k)$ is the number of people on page i at time step k, and $p(x,y)$ is the probability that you will jump from page x to page y, then

$$x_i(k+1) = \sum_{j=1}^{n} x_j(k)\, p(j,i)$$

If we want to compute the expected number of viewers at time $k+1$ for all pages simultaneously, we can use a handy matrix notation. Define $\vec{x}(k) = \begin{bmatrix} x_1(k) & x_2(k) & \dots & x_n(k) \end{bmatrix}^T$ as the vector encoding the number of viewers on each page at time k. Then

$$\vec{x}(k+1) = \begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ \vdots \\ x_n(k+1) \end{bmatrix} = \begin{bmatrix} p(1,1) & p(2,1) & \dots & p(n,1) \\ p(1,2) & p(2,2) & \dots & p(n,2) \\ \vdots & \vdots & \vdots & \vdots \\ p(1,n) & p(2,n) & \dots & p(n,n) \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ \vdots \\ x_n(k) \end{bmatrix}$$

We call the matrix of probabilities P. Check that the entry $x_i(k+1)$ does in fact match the summation above. For our specific example, the matrix equation looks like this:

$$\begin{bmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} 6 \\ 6 \\ 6 \\ 6 \end{bmatrix} = \begin{bmatrix} 9 \\ 2 \\ 8 \\ 5 \end{bmatrix}$$

If we start with initial counts $\vec{x}(0)$ and want to find the expected number of viewers on each page at time k, we compute

$$\vec{x}(k) = P\vec{x}(0) = \begin{bmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix}^k \vec{x}(0)$$

We usually think of $\vec{x}(k)$ as probabilities or frequencies of viewers, so we initialize $\vec{x}(0)$ as $\begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}^T$. After running the simulation for many time steps, we hope the frequencies will reflect the popularity of the webpages. This system is only useful to us if the values of $\vec{x}(k)$ *converge* to some stable frequencies.

Once our frequencies exactly match these "stable frequencies", running the simulation for another time step should not change the values. This means that if the frequencies converge, at some point we'll reach frequencies $\vec{x}^*$ such that

$$\vec{x}^* = P\vec{x}^*$$

Eigenvectors and Eigenvalues

In our Page Rank example, $\vec{x}^*$ is an *eigenvector* of P. An $n \times 1$ vector $\vec{x}$ is called an eigenvector of an $n \times n$ square matrix A if

$$A\vec{x} = \lambda\vec{x},\ \vec{x} \neq \vec{0},\ \lambda \in \mathbb{R}$$

where λ is a scalar value, called the *eigenvalue* of $\vec{x}$.

How do we solve this equation for both $\vec{x}$ and λ, knowing only A? It seems we don't have enough information. We start by bringing everything over to one side:

$$A\vec{x} - \lambda\vec{x} = \vec{0}$$

We would like to factor out $\vec{x}$ on the left side, but currently the dimensions don't agree. $\vec{x}$ is an $n \times 1$ vector, while A is an $n \times n$ matrix. To fix this, we replace $\lambda\vec{x}$ with $\lambda I_n \vec{x}$, where I_n is the $n \times n$ identity matrix:

$$(A - \lambda I_n)\vec{x} = \vec{0}$$

Remember, in the definition of an eigenvector, we explicitly excluded the $\vec{0}$ vector. We know at least one element of $\vec{x}$ is nonzero, yet $(A - \lambda I_n)\vec{x} = \vec{0}$. To see what this means, let's rewrite the product in terms of the columns of $A - \lambda I_n$. The i^{th} column is the vector $\vec{a}_i$:

$$\begin{bmatrix} \\ \vec{a}_1 & \vec{a}_2 & \dots & \vec{a}_n \\ \\ \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} \\ v_1\vec{a}_1 & v_2\vec{a}_2 & \dots & v_n\vec{a}_n \\ \\ \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

(Check to see that this column formulation is a valid way to represent matrix multiplication.) From this formulation, we can see that some nonzero linear combination of the columns of $A - \lambda I_n$ results in 0. This means that the columns of this matrix must be linearly dependent, and therefore $A - \lambda I_n$ is not invertible.

Determinants

Enter the determinant. The determinant is a scalar value associated with a square matrix that concisely encodes many important properties of the matrix. For our purposes, the most important part is that the determinant of M is 0 is M is not invertible.

For this class, we only need to know how to compute the determinant of a 2×2 matrix. This is given by

$$\det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad - bc$$

With this, we can revisit our previous equation:

$$(A - \lambda I_n)\vec{x} = \vec{0} \implies \det(A - \lambda I_n) = 0$$

Consider the example 2×2 matrix

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

We now expand the expression above:

$$A - \lambda I_n = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 1-\lambda & 2 \\ 4 & 3-\lambda \end{bmatrix}$$

We now take the determinant of this matrix and set it equal to 0:

$$(1-\lambda)(3-\lambda) - 4 \times 2 = 0$$

Expanding this expression, we get

$$\lambda^2 - 4\lambda - 5 = (\lambda+1)(\lambda-5) = 0$$

We find that there are two eigenvalues: $\lambda = -1, \lambda = 5$. Each eigenvalue will have its own corresponding eigenvector. To find these, we simply plug in the values of λ into the original equation:

$$(A - 5I_2)\vec{x} = \begin{bmatrix} -4 & 2 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \vec{0}$$

We see that both of the rows provide redundant information: $4x_1 - 2x_2 = 0$. The eigenvectors associated with $\lambda = 5$ are all of the form

$$\alpha \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \alpha \in \mathbb{R}$$

Now we plug in $\lambda = -1$:

$$(A + I_2)\vec{x} = \begin{bmatrix} 2 & 2 \\ 4 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \vec{0}$$

Both rows provide the information $x_1 = -x_2$. The eigenvectors associated with $\lambda = -1$ are of the form

$$\alpha \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \alpha \in \mathbb{R}$$