Math 443/543 Graph Theory Notes 5: Graphs as matrices, spectral graph theory, and PageRank

David Glickenstein

November 3, 2014

1 Representing graphs as matrices

It will sometimes be useful to represent graphs as matrices. This section is taken from C-10.1.

Let G be a graph of order p. We denote the vertices by $v_1, \ldots, v_p$. We can then find an adjacency matrix $A = A(G) = [a_{ij}]$ defined to be the $p \times p$ matrix such that $a_{ij} = 1$ if $v_i v_j \in E(G)$. This matrix will be symmetric for an undirected graph. We can easily consider the generalization to directed graphs and multigraphs.

Note that two isomorphic graphs may have different adjacency matrices. However, they are related by permutation matrices.

Definition 1 *A permutation matrix is a matrix gotten from the identity by permuting the columns (i.e., switching some of the columns).*

Proposition 2 *The graphs G and G' are isomorphic if and only if their adjacency matrices are related by*

$$A = P^T A' P$$

for some permutation matrix P.

Proof (sketch). Given isomorphic graphs, the isomorphism gives a permutation of the vertices, which leads to a permutation matrix. Similarly, the permutation matrix gives an isomorphism. ■

Now we see that the adjacency matrix can be used to count uv-walks.

Theorem 3 *Let A be the adjacency matrix of a graph G, where $V(G) = \{v_1, v_2, \ldots, v_p\}$. Then the (i, j) entry of A^n, where $n \geq 1$, is the number of different $v_i v_j$-walks of length n in G.*

Proof. We induct on n. Certainly this is true for $n = 1$. Now suppose $A^n = \left(a_{ij}^{(n)}\right)$ gives the number of $v_i v_j$-walks of length n. We can consider the entries of $A^{n+1} = A^n A$. We have

$$a_{ij}^{(n+1)} = \sum_{k=1}^{p} a_{ik}^{(n)} a_{kj}.$$

This is the sum of all walks of length n between v_i and v_k followed by a walk from v_k to v_j of length 1. All walks of length $n + 1$ are generated in this way, and so the theorem is proven. ■

2 Spectral graph theory

Recall that an eigenvalue of a matrix M is a number λ such that there is a vector v (called the corresponding eigenvector) such that

$$Mv = \lambda v.$$

It turns out that symmetric $n \times n$ matrices have n eigenvalues. Since adjacency matrices of two isomorphic graphs are related by permutation matrices as above, and so the set of eigenvalues of A is an invariant of a graph.

We will actually use the Laplacian matrix instead of the adjacency matrix. The Laplacian matrix is defined to be

$$L = A - D$$

where D is the diagonal matrix whose entries are the degrees of the vertices (called the degree matrix). The Laplacian matrix is also symmetric, and thus it has a complete set of eigenvalues. The set of these eigenvalues is called the spectrum of the Laplacian. Notice the following.

Proposition 4 *Let G be a finite graph. The eigenvalues of the matrix L are all nonpositive. Moreover, the constant vector $\vec{1} = (1, 1, 1, \ldots, 1)$ is an eigenvector with eigenvalue zero.*

Proof. It is clear that $\vec{1}$ is an eigenvector with eigenvalue 0 since the sum of the entries in each row must be zero. Now, notice that we can write

$$\begin{aligned}
v^T L v &= \sum v_i (Lv)_i \\
&= \sum_i v_i \sum_j L_{ij} v_j \\
&= \sum_{v_i v_j \in E} v_i (v_j - v_i) \\
&= \frac{1}{2} \left[\sum_{v_i v_j \in E} v_i (v_j - v_i) + \sum_{v_i v_j \in E} v_j (v_i - v_j) \right] \\
&= -\frac{1}{2} \sum_{v_i v_j \in E} (v_i - v_j)^2 \leq 0.
\end{aligned}$$

(The sums over i are over all vertices, but the sums over $v_i v_j \in E$ is the sum over the edges.) Now note that if v is an eigenvector of L with eigenvalue λ, then $Lv = \lambda v$, and

$$v^T L v = \lambda v^T v = \lambda \sum_i v_i^2.$$

Thus we have that

$$\lambda = \frac{-\frac{1}{2} \sum_{v_i v_j \in E} (v_i - v_j)^2}{\sum_i v_i^2} \leq 0.$$

■

Definition 5 *The eigenvalues of $-L$ can be arranged $0 = \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{p-1}$, where p is the order of the graph. The collection $(\lambda_0, \lambda_1, \ldots, \lambda_p)$ is called the* spectrum *of the Laplacian.*

Remark 6 *Sometimes the Laplacian is taken to be $D^{-1/2} L D^{-1/2}$. If there are no isolated vertices, these are essentially equivalent.*

Remark 7 *Note that the Laplacian matrix, much like the adjacency matrix, depends on the ordering of the vertices and must be considered up to conjugation by permutation matrices. Since eigenvalues are independent of conjugation by permutation matrices, the spectrum is an isomorphism invariant of a graph.*

The following is an easy fact about the spectrum:

Proposition 8 *For a graph G of order p,*

$$\sum_{i=0}^{p-1} \lambda_i = 2q.$$

Proof. The sum of the eigenvalues is equal to the trace, which is the sum of the degrees. ■

We will be able to use the eigenvalues to determine some geometric properties of a graph.

3 Connectivity and spanning trees

Recall that $\lambda_0 = 0$, which means that the matrix L is singular and its determinant is zero. Recall the definition of the adjugate of a matrix.

Definition 9 *If M is a matrix, the adjugate is the matrix $M^\dagger = \left[M_{ij}^\dagger\right]$ where $M_{ij}^\dagger$ is equal to $(-1)^{i+j} \det\left(\hat{M}_{ij}\right)$, where $\hat{M}_{ij}$ is the matrix with the ith row and jth column removed.*

The adjugate has the property that

$$M\left(M^\dagger\right)^T = (\det M)\, I,$$

where I is the identity matrix. Applying this to L (which is symmetric) gives that

$$LL^\dagger = 0.$$

Now, the $p \times p$ matrix L has rank less than p. If it is less than or equal to $p-2$, then all determinants of $(p-1) \times (p-1)$ submatrices are zero, and hence $L^\dagger = 0$. If L has rank $p-1$, then it has only one zero eigenvalue, which must be $(1,1,\ldots,1)^T$. Since $LL^\dagger = 0$, all columns of $L^\dagger$ must be a multiple of $(1,1,\ldots,1)^T$. But L is symmetric, so that means that $L^\dagger$ must be a multiple of the matrix of all ones. This motivates the following definition.

Definition 10 *We define $t(G)$ by*

$$t(G) = (-1)^{i+j} \det\left(-\hat{L}_{ij}\right)$$

for any i and j (it does not matter since all are the same).

Remark 11 *It follows that $t(G)$ is an integer.*

Proposition 12 $t(G) = \frac{1}{p}\lambda_1\lambda_2\cdots\lambda_{p-1}$.

Proof. In general for a matrix A with eigenvalues $\lambda_0, \ldots, \lambda_{p-1}$ we have that

$$\sum_{k=0}^{p-1} \frac{\lambda_0\lambda_1\lambda_2\cdots\lambda_{p-1}}{\lambda_k} = \sum_{i=0}^{p-1} \det \hat{A}_{ii}.$$

In our case, $\lambda_0 = 0$ and the right sum is the sum of p of the same entries, so the result follows. ∎

Remark 13 *It follows that $t(G) \geq 0$.*

Recall that a spanning tree of G is a subgraph containing all of the vertices of G and is a tree.

Theorem 14 (Matrix Tree Theorem) *The number $t(G)$ is equal to the number of spanning trees of G.*

Proof. Omitted, for now. ■

We can apply this, however, as follows.

Example 1, consider the graph K_3. Clearly this has Laplacian matrix

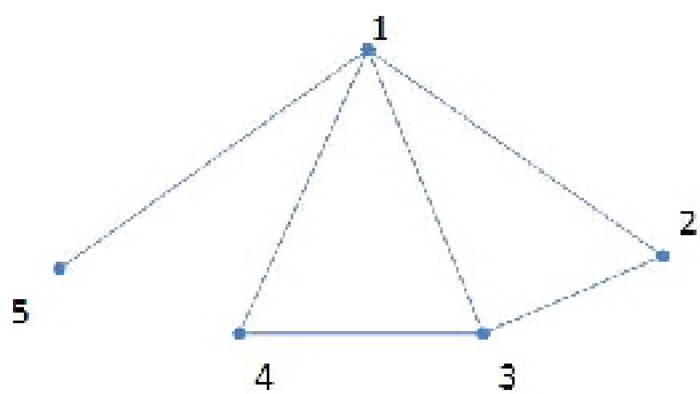
$$L(G) = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}.$$

The number of spanning trees are equal to

$$\det \begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix} = 3.$$

It is clear that each spanning tree is given by omitting one edge, so it is clear there are 3.

Example 2: Consider the following graph.



Its Laplacian matrix is

$$L(G) = \begin{pmatrix} 4 & -1 & -1 & -1 & -1 \\ -1 & 2 & -1 & 0 & 0 \\ -1 & -1 & 3 & -1 & 0 \\ -1 & 0 & -1 & 2 & 0 \\ -1 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

The number of spanning trees are equal to

$$
\begin{aligned}
t(G) &= \det\begin{pmatrix} -2 & 1 & 0 & 0 \\ 1 & -3 & 1 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \\
&= \det\begin{pmatrix} 2 & -1 & 0 \\ -1 & 3 & -1 \\ 0 & -1 & 2 \end{pmatrix} \\
&= 2(6-1) + (-2) = 8
\end{aligned}
$$

One can check directly that it has eight spanning trees.

Corollary 15 $\lambda_1 \neq 0$ *if and only if* G *is connected.*

Proof. $\lambda_1 = 0$ if and only if $t(G) = 0$ since $t(G)$ is the product of the eigenvalues $\lambda_1\lambda_2 \cdots \lambda_{p-1}$ and λ_1 is the minimal eigenvalue after λ_0. But $t(G) = 0$ means that there are no spanning trees, so G is not connected. ■

Now we can consider the different components.

Definition 16 *The disjoint union of two graphs* $G = G_1 \sqcup G_2$ *is the graph gotten by taking* $V(G) = V(G_1) \sqcup V(G_2)$ *and* $E(G) = E(G_1) \sqcup E(G_2)$ *where* $\sqcup$ *is the disjoint union of sets.*

It is not hard to see that if we number the vertices in G by first numbering the vertices of G_1 and then numbering the vertices of G_2, that the Laplacian matrix takes the form

$$
L(G) = \begin{pmatrix} L(G_1) & 0 \\ 0 & L(G_2) \end{pmatrix}.
$$

This means that the eigenvalues of $L(G)$ are the union of the eigenvalues of $L(G_1)$ and $L(G_2)$. This implies the following.

Corollary 17 *If* $\lambda_n = 0$, *then there are at least* $n+1$ *connected components of* G.

Proof. Induct on n. We already know this is true for $n = 1$. Suppose $\lambda_n = 0$. We know there must be at least n components, since $\lambda_n = 0$ implies $\lambda_{n-1} = 0$. We can then write the matrix $L(G)$ in the block diagonal form with $L(G_i)$ along the diagonal for some graphs G_i. Since $\lambda_n = 0$, one of these graphs must have $\lambda_1(G_i) = 0$. But that means that there is another connected component, completing the induction. ■

4 PageRank problem and idea of solution

We will generally follow the paper by Bryan and Leise, denoted BL.

Search engines generally do three things:

1. Locate all webpages on the web.
2. Index the data so that it can be searched efficiently for relevent words.
3. Rate the importance of each page so that the most important pages can be shown to the user first.

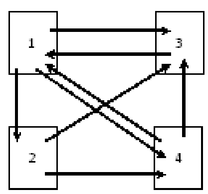
We will discuss this third step.

We will assign a nonnegative score to each webpage such that more important pages have higher scores. The first idea is:

- Derive the score for a page by the number of links to that page from other pages (called the "backlinks" for the page).

In this sense, other pages vote for the page. The linking of pages produces a digraph. Denote the vertices by v_k and the score of vertex v_k by x_k.

Approach 1: Let x_k equal the number of backlinks for page v_k. See example in BL Figure 1 reproduced here:



We see that $x_1 = 2$, $x_2 = 1$, $x_3 = 3$, and $x_4 = 2$. Here are two problems with this ranking:

Problem 1: Links from more important pages should increase the score more. For instance, the scores of v_1 and v_4 are the same, but v_1 has a link from x_3, which is a more important page, so maybe it should be ranked higher. We will deal with this by, instead of letting x_i equal the total number of links to it, we will have it be equal to the sum of the scores of the pages linking to it, so more important pages count more. Thus we get the relations

$$\begin{aligned} x_1 &= x_3 + x_4 \\ x_2 &= x_1 \\ x_3 &= x_1 + x_2 + x_4 \\ x_4 &= x_1 + x_2. \end{aligned}$$

This doesn't quite work as stated, since to solve this linear system, we see that we get $x_1 = x_2 = \frac{1}{2}x_4 = \frac{1}{4}x_3$, which means that if we look at the first equality, we must have that they are all equal to zero. However, a slight modification in regard to the next problem will fix this.

Problem 2: One site should not be able to *significantly* affect the rankings by creating lots of links. Of course, creating links should affect the rankings, but by creating thousands of links from one site, one should not be able to boost the importance too much. So instead of giving one vote for each link out, we will give equal votes to each outlink from a particular page, but the total votes is equal to one. This changes the above system to

$$\begin{aligned} x_1 &= x_3 + \frac{1}{2}x_4 \\ x_2 &= \frac{1}{3}x_1 \\ x_3 &= \frac{1}{3}x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_4 \\ x_4 &= \frac{1}{3}x_1 + \frac{1}{2}x_2. \end{aligned}$$

This can be solved as follows.

$$\begin{aligned} x_2 &= \frac{1}{3}x_1 \\ x_4 &= \frac{1}{3}x_1 + \frac{1}{2}x_2 = \frac{1}{2}x_1 \\ x_3 &= \frac{1}{3}x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_4 = \frac{1}{3}x_1 + \frac{1}{6}x_1 + \frac{1}{4}x_1 = \frac{3}{4}x_1. \end{aligned}$$

Thus we can have a score of $x_1 = 1$, $x_2 = \frac{1}{3}$, $x_3 = \frac{3}{4}$, $x_4 = \frac{1}{2}$. Notice that x_1 has the highest ranking! This is because x_3 threw its whole vote to x_1 and so that even though x_3 got votes from three different sites, they still do not total as much as what x_1 gets. Note, usually we will rescale so that the sum is equal to 1, and so we get

$$x_1 = \frac{12}{31}, \;\; x_2 = \frac{4}{31}, \;\; x_3 = \frac{9}{31}, \;\; x_4 = \frac{6}{31}.$$

5 General formulation

First, we introduce some terminology for directed graphs.

Definition 18 *If $e = (v, w) \in E_+$ is a directed edge (we use E_+ to denote the edges in a directed graph, and the ordered pair to denote that the edge is from v to w, denoted in a picture as an arrow from v to w), we say v is* adjacent to *(or* links to*) w and w is* adjacent from *(or* links from*) v.*

Definition 19 *The* indegree *of a vertex v, denoted $i\deg(v)$, is the number of vertices adjacent to v. The* outdegree *of v, denoted $o\deg(v)$, is the number of vertices adjacent from v.*

We can state the Page Rank algorithm in a more general way. We want to assign scores so that

$$x_i = \sum_{j \in L_i} \frac{x_j}{n_j}$$

where L_i are the indices such that v_j links to v_i if $j \in L_i$, and n_j is equal to outdegree of v_j. Note that L_i contains $i\deg(v_i)$ elements. The set L_i is called the set of *backlinks* of vertex v_i. This can be rewritten as a vector equation

$$x = Ax,$$

where A is the matrix $A = (a_{ij})$ given by

$$a_{ij} = \begin{cases} \frac{1}{n_j} & \text{if } j \in L_i \\ 0 & \text{otherwise} \end{cases}.$$

This matrix is called the *link matrix*. We note that in the example, the matrix A was

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

The problem of solving for the scores x then amounts to finding an eigenvector with eigenvalue 1 for the matrix A.

We can consider the link matrix as giving the probabilities of traversing a link from the page represented by the column to the page representing the row. Thus it makes sense that the sum of the values of the columns are equal to one.

Definition 20 *A matrix is called a* column stochastic *matrix if all of its entries are positive and the sum of the elements in each column are equal to* 1.

Now the question is whether we can find an eigenvector for a column stochastic matrix, and the answer is yes.

Proposition 21 *If A is a column stochastic matrix, then* 1 *is an eigenvalue.*

Proof. Let e be the column vector of all ones. Since A is column stochastic, we clearly have that

$$e^T A = e^T.$$

Thus

$$A^T e = e$$

and e is an eigenvector with eigenvalue 1 for A^T. However, A and A^T have the same eigenvalues (not eigenvectors, though), so A must have an eigenvalue 1, too. ■

Remark 22 *Do you remember why A and A^T have the same eigenvalues? The eigenvalues of A are the solutions λ of* $\det(A - \lambda I) = \det(A^T - \lambda I)$.

6 Challenges to the algorithm

See the worksheet.

7 Computing the ranking

The basic idea is that we can try to compute an eigenvector iteratively like

$$x_{k+1} = Mx_k = M^k x_0.$$

Certainly, if $Mx_0 = x_0$, then this procedure fixes x_0. In general, if we replace this method with

$$x_{k+1} = \frac{Mx_k}{\|Mx_k\|}$$

for any vector norm, we will generally find an eigenvector for the largest eigenvalue.

Before we proceed, let's define the one-norm of a vector.

Definition 23 *The one-norm of a vector $v = (v_i) \in \mathbb{R}^n$ is equal to*

$$\|v\|_1 = \sum_{i=1}^{n} |v_i|,$$

where $|v_i|$ is the absolute value of the ith component of v.

Proposition 24 *Let M be a positive column-stochastic $n \times n$ matrix and let V denote the subspace of $\mathbb{R}^n$ consisting of vectors v such that*

$$\sum_{i=1}^{n} v_i = 0.$$

Then for any $v \in V$ we have $Mv \in V$ and

$$\|Mv\|_1 \leq c \|v\|_1,$$

where $c < 1$.

Corollary 25 *In the situation in the proposition,*

$$\left\|M^k v\right\|_1 \leq c^k \|v\|_1.$$

Proof. This is a simple induction on k, using the fact that $Mv \in V$ and

$$\left\|M^k v\right\|_1 \leq c \left\|M^{k-1} v\right\|_1.$$

■

This is essentially showing that the iteration is a contraction mapping, and that will allow us to show that the method works.

Proposition 26 *Every positive column-stochastic matrix M has a unique vector q with positive components such that $Mq = q$ and $\|q\|_1 = 1$. The vector can be computed as*

$$q = \lim_{k \to \infty} M^k x_0$$

for any initial guess x_0 with positive componets such that $\|x_0\|_1 = 1$.

Proof. We already know that M has 1 as an eigenvalue and that the subspace $V_1(M)$ is one-dimensional. All eigenvectors have all positive or all negative components, so we can choose a unique representative q with positive components and norm 1 by rescaling. Now let x_0 be any vector in $\mathbb{R}^n$ with positive components and $\|x_0\| = 1$. We can write

$$x_0 = q + v$$

for some vector v. We note that if we sum the components of x_0 or the components of q, we get one since both have positive components and 1-norm equal to one. Thus $v \in V$ as in the previous proposition. Now we see that

$$\begin{aligned} M^k x_0 &= M^k q + M^k v \\ &= q + M^k v. \end{aligned}$$

Thus

$$\left\| M^k x_0 - q \right\|_1 = \left\| M^k v \right\|_1 \leq c^k \|v\|_1 .$$

Since $c < 1$, we get that $\left\| M^k x_0 - q \right\|_1 \to 0$ as $k \to \infty$. ■

We now go back and prove Proposition 24.

Proof of Proposition 24. It is pretty clear that $Mv \in V$ since

$$\begin{aligned} \sum (Mv)_j &= \sum_j \sum_i M_{ji} v_i \\ &= \sum_i \sum_j M_{ji} v_i \\ &= \sum_i v_i = 0 \end{aligned}$$

since M is column-stochastic. Now we consider

$$\begin{aligned} \|Mv\|_1 &= \sum_j \left| \sum_i M_{ji} v_i \right| \\ &= \sum_j e_j \sum_i M_{ji} v_i \\ &= \sum_i a_i v_i \end{aligned}$$

where

$$e_j = \operatorname{sgn}\left(\sum_i M_{ji} v_i \right)$$

and

$$a_i = \sum_j e_j M_{ji}.$$

Note that if $|a_i| \leq c$ for all i, then

$$\|Mv\|_1 \leq c \|v\|_1 .$$

We can see that

$$\begin{aligned} |a_i| &= \left| \sum_j e_j M_{ji} \right| \\ &= \left| \sum_j M_{ji} + \sum_j (e_j - 1) M_{ji} \right| \\ &= \left| 1 + \sum_j (e_j - 1) M_{ji} \right| . \end{aligned}$$

Each term in the sum is nonpositive, and since M_{ji} are positive and e_j are not all the same sign, the largest this can be is if most e_j are 1 except for a single e_j which is negative and corresponds to the smallest M_{ji}. Thus we see that

$$|a_i| \leq 1 - 2 \min_j M_{ji} \leq 1 - 2 \min_{i,j} M_{ji} < 1.$$

■

8 Appendix: Approximating the Laplacian on a lattice

Recall that the Laplacian is the operator

$$\triangle = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

acting on functions $f(x, y, z)$, with analogues in other dimension. Let's first consider a way to approximate the one-dimensional Laplacian. Suppose $f(x)$ is a function and I want to approximate the second derivative $\frac{d^2 f}{dx^2}(x)$. We can take a centered difference approximation to the get this as

$$\begin{aligned} \frac{d^2 f}{dx^2}(x) &\approx \frac{f'\left(x + \frac{1}{2}\Delta x\right) - f'\left(x - \frac{1}{2}\Delta x\right)}{\Delta x} \\ &\approx \frac{1}{\Delta x}\left[\frac{f(x + \Delta x) - f(x)}{\Delta x} - \frac{f(x) - f(x - \Delta x)}{\Delta x}\right] \\ &= \frac{1}{(\Delta x)^2}\left[f(x + \Delta x) - f(x) + (f(x - \Delta x) - f(x))\right] \\ &= \frac{1}{(\Delta x)^2}\left[f(x + \Delta x) + f(x - \Delta x) - 2f(x)\right] \end{aligned}$$

Note that if we take $\Delta x = 1$, then this only depends on the value of the function at the integer points.

Now consider the graph consisting of vertices on the integers of the real line and edges between consecutive integers. Give an function f on the vertices, we can compute the Laplacian as

$$\triangle f(v_i) = f(v_{i+1}) + f(v_{i-1}) - 2f(v_i)$$

for any vertex v_i. Notice that the Laplacian is an infinite matrix of the form

$$\triangle f = \begin{pmatrix} \cdots & \cdots & & & & & \\ \cdots & -2 & 1 & 0 & & & \\ & 1 & -2 & 1 & 0 & & \\ & 0 & 1 & -2 & 1 & 0 & \\ & & 0 & 1 & -2 & 1 & 0 \\ & & & 0 & 1 & -2 & \cdots \\ & & & & & \cdots & \cdots \end{pmatrix} \begin{pmatrix} \cdots \\ f(v_2) \\ f(v_1) \\ f(v_0) \\ f(v_{-1}) \\ f(v_{-2}) \\ \cdots \end{pmatrix}.$$

Also note that that matrix is exactly equal to the adjacency matrix minus twice the identity. The number 2 is the degree of each vertex, so we can write the matrix, which is called the Laplacian matrix, as

$$L = A - D$$

where A is the adjacency matrix and D is the diagonal matrix consisting of degrees (called the degree matrix).

Note that something similar can be done for a two-dimensional grid. We see that

$$\begin{aligned}
\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) f(x,y) &\approx \frac{f_x\left(x + \frac{1}{2}\Delta x, y\right) - f_x\left(x - \frac{1}{2}\Delta x, y\right)}{\Delta x} + \frac{f_y\left(x, y + \frac{1}{2}\Delta y\right) - f_y\left(x, y - \frac{1}{2}\Delta y\right)}{\Delta y} \\
&\approx \frac{1}{\Delta x} \frac{f(x + \Delta x, y) - f(x,y) - [f(x,y) - f(x - \Delta x, y)]}{\Delta x} \\
&+ \frac{1}{\Delta y} \frac{f(x, y + \Delta y) - f(x,y) - [f(x,y) - f(x, y - \Delta y)]}{\Delta y} \\
&= \frac{1}{(\Delta x)^2} [f(x + \Delta x, y) - f(x,y) + (f(x - \Delta x, y) - f(x,y))] \\
&+ \frac{1}{(\Delta y)^2} [f(x, y + \Delta y) - f(x,y) + (f(x, y - \Delta y) - f(x,y))] \\
&= \frac{1}{(\Delta x)^2} [f(x + \Delta x, y) + (f(x - \Delta x, y) - 2f(x,y))] \\
&+ \frac{1}{(\Delta y)^2} [f(x, y + \Delta y) + (f(x, y - \Delta y) - 2f(x,y))].
\end{aligned}$$

If we let $\Delta x = \Delta y = 1$, then we get

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) f(x,y) \approx f(x+1,y) + f(x-1,y) + f(x,y+1) + f(x,y-1) - 4f(x,y).$$

Note that on the integer grid, this translates to the sum of the value of f for the four vertices neighboring the vertex, minus 4 times the value at the vertex. This is precisely the same as the last time, and we see that this operator can again be written as

$$L = A - D.$$

In general we will call this matrix the Laplacian matrix. It can be thought of as a linear operator on functions on the vertices. Sometimes the Laplacian will denote the negative of this operator (which gives positive eigenvalues instead of negative ones), and sometimes a slight variation is used in the graph theory literature.

9 Appendix: Electrical networks

One finds applications of the Laplacian in the theory of electrical networks. Recall that the current through a circuit is proportional to the change in voltage, and that constant of proportionality is called the conductance (or resistance, depending on where the constant is placed. Thus, for a wire with conductance C between points v and w with voltages $f(v)$ and $f(w)$ respectively, the current from v to w is $C(f(v) - f(w))$. Kirchoff's law says that if we have a network of wires, each with conductance C, the total current through any given point is zero. Thus, we get that

$$C \sum_{v \text{ adjacent to } w} (f(w) - f(v)) = 0$$

which is the same as $\triangle f = 0$. Note that if the conductances are different, then we get an equation like

$$\sum_{v \text{ adjacent to } w} c_{vw} (f(w) - f(v)) = 0,$$

which is quite similar to the Laplacian. Hence we can use the Laplacian to understand graphs by attaching voltages to some of the vertices and seeing what happens at the other vertices. This is very much like solving a boundary value problem for a partial differential equation!

10 Appendix: Proof of the matrix tree theorem

In order to prove the Matrix Tree Theorem, we need another characterization of the Laplacian.

Definition 27 *Let G be a directed (p,q)-graph. The oriented vertex-edge incidence graph is a $p \times q$ matrix $Q = [q_{ir}]$, such that $q_{ir} = 1$ if $e_r = (v_i, v_j)$ for some j and $q_{ir} = -1$ if $e_r = (v_j, v_i)$ for some j.*

Proposition 28 *The Laplacian matrix L satisfies*

$$-L = QQ^T$$

for any oriented vertex-edge incidence graph (so, given an undirected graph, we can take any orientation), i.e.,

$$L_{ij} = -\sum_{r=1}^{q} q_{ir} q_{jr}.$$

Proof. If $i = j$, then we see that $L_{ii} = -\deg v_i$. Notice that q_{ir} is nonzero if v_i is in edge e_r. Thus

$$\sum_{r=1}^{q} (q_{ir})^2 = \deg v_i.$$

Now for $i \neq j$, we have that $q_{ir} q_{jr} = -1$ if $e_r = v_i v_j$, giving the result. ∎

Remark 29 *This observation can be used to give a different proof that L_{ij} has all nonpositive eigenvalues.*

The proof of the matrix tree theorem proceeds as follows. We need only consider the case

$$\det\left(-\hat{L}_{11}\right).$$

A property of determinants allows us to use the fact that $-L = QQ^T$ to compute this determinant in terms of determinants of submatrices of Q.

Proposition 30 (Binet-Cauchy Formula) *Let A be an $n \times m$ matrix and B be an $n \times m$ matrix (usually we think $n < m$). We can compute*

$$\det(AB) = \sum_S \det A_S \det B_S^T$$

where A_S is matrix consisting of the columns of A specified by S, and S ranges over all choices of n columns.

We will not prove this theorem, but it is quite geometric in the case that $B = A^T$, which is our case. If we consider the rows of A, then $\det AA^T$ is equal to the square of the volume of the parallelopiped determines by these rows (since we can rotate A to be in $\mathbb{R}^n$). The formula says that the volume squared is equal to the sum of the squares of the volumes when projected onto the coordinate planes. These are generalizations of the pythagorean theorem!

Thus, we have that

$$-L = QQ^T$$

for a choice of directions. To compute $t(G)$, we have that

$$\begin{aligned} t(G) &= \det\left(-\hat{L}_{11}\right) = \det\left(\hat{Q}\hat{Q}^T\right) \\ &= \sum_S \left(\det \hat{Q}_S\right)^2 \end{aligned}$$

where $\hat{Q}$ is Q with the first row removed and S ranges over collections of $p-1$ edges in G. We need to understand what $\det \hat{Q}_S$ represents. The claim is that $\det Q_S = \pm 1$ if the edges represented by S form a spanning tree for G and zero otherwise. If S forms a disconnected graph, then $\hat{Q}_S$ looks like

$$\begin{pmatrix} M & 0 \\ 0 & M' \end{pmatrix}$$

and since M' consists of columns with exactly one 1 and one -1, we see that $(0,0,\ldots,0,1,1,1,\ldots 1)^T$ (respecting the form above) is a zero eigenvector, and hence the determinant is equal to 0. Now, if S contains a cycle, then since it has $p-1$ edges, the corresponding graph must be disconnected, and the previous comment follows.The problem of showing that $\det \hat{Q}_S = \pm 1$ if the edges represented by S form a spanning tree will be part of your next homework assignment.

It follows that

$$t(G) = \sum_S \left(\det \hat{Q}_S\right)^2$$

is equal to the number of spanning trees.

11 Appendix: Challenges to the PageRank algorithm

There are two issues we will have to deal with.

11.1 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. The link matrix is

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

It is easy to see that the vectors $\left[\frac{1}{2},\frac{1}{2},0,0,0\right]^T$ and $\left[0,0,\frac{1}{2},\frac{1}{2},0\right]^T$ have eigenvalue 1. However, we also have that any linear combination of these have eigenvalue 1, and so we have vectors like $\left[\frac{3}{8},\frac{3}{8},\frac{1}{8},\frac{1}{8},0\right]^T$ as well as $\left[\frac{1}{8},\frac{1}{8},\frac{3}{8},\frac{3}{8},0\right]^T$, which give different rankings!

Proposition 31 *Let W be a web with r components $W_1, W_2, \ldots, W_r$. Then the eigenspace of the eigenvalue 1 is at least r-dimensional.*

Proof (Sketch). A careful consideration shows 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. If each is 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. Each of these is linearly independent and part of the eigenspace V_1 of eigenvalue 1. ■

We will figure out a way to deal with this soon.

11.2 Dangling nodes

Definition 32 *A* dangling node *is a vertex in the web with outdegree zero (i.e., with no links).*

The problem with a dangling node is that it produces a column of zeroes. 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. We will sketch how to deal with this later.

12 Appendix: Solving the problems of the PageRank algorithm

12.1 Dealing with multiple eigenspaces

Recall that we seemed to be okay if we had a strongly connected graph (web). We will now take our webs that are not strongly connected and make them strongly connected by adding a little bit of an edge between any two vertices. From a probabilistic perspective, we are adding on a possibility of randomly jumping to any page on the entire web. We will make this probability small compared with the probability to navigate from a page.

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). Also notice 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 33 *A matrix $M = (M_{ij})$ is positive if $M_{ij} > 0$ for every i and j.*

For future use, we define the following.

Definition 34 *Given a matrix M, we write $V_\lambda(M)$ for the eigenspace of eigenvalue λ.*

We will show that a positive column-stochastic matrix has a one dimensional eigenspace V_1 for eigenvalue 1.

Proposition 35 *If M is a positive, column-stochastic matrix, then $V_1(M)$ has dimension 1.*

Proof. Suppose v and w are in $V_1(M)$. Then we know that $sv + tw \in V_1(M)$ for any real numbers s and t. We will now show that (1) any eigenvector in $V_1(M)$ has all positive or all negative components and that (2) if x and y are any two linearly independent vectors, then there is some s and some t such that $sx + ty$ has both positive and negative components. This would imply that $sv + tw$ has all positive or all negative components, and thus v and w must be linearly dependent. ■

First we prove the following:

Proposition 36 *Any eigenvector in $V_1(M)$ has all positive or all negative components.*

Proof. Suppose $Mv = v$. Since M is column-stochastic, we know that

$$\sum_{i=1}^{n} M_{ij} = 1$$

for each j, and since M is positive, we know that

$$|M_{ij}| = M_{ij}$$

for each i and j. Therefore, we see that

$$\begin{aligned}
\|v\|_1 &= \|Mv\|_1 \\
&= \sum_{i=1}^{n} \left| \sum_{j=1}^{n} M_{ij} v_j \right| \\
&\le \sum_{i=1}^{n} \sum_{j=1}^{n} |M_{ij}| \, |v_j| \\
&= \sum_{j=1}^{n} \sum_{i=1}^{n} M_{ij} \, |v_j| \\
&= \sum_{j=1}^{n} |v_j| = \|v\|_1 .
\end{aligned}$$

That means that the inequality must be an equality, meaning that

$$\left| \sum_{j=1}^{n} M_{ij} v_j \right| = \sum_{j=1}^{n} |M_{ij}| \, |v_j| .$$

This is only true if $M_{ij} v_j \ge 0$ for each i and j (or $M_{ij} v_j \le 0$ for each i and j). However, since $M_{ij} > 0$, this implies that $v_j \ge 0$ for each j (or $v_j \le 0$ for each j). Furthermore, since

$$v_i = \sum_{j=1}^{n} M_{ij} v_j$$

with $v_j \ge 0$ ($v_j \le 0$)and $M_{ij} > 0$, we must have that either all v are zero or all are positive (negative). Since v is an eigenvector, it is not the zero vector. ■

Remark 37 *A similar argument shows that for any positive, column-stochastic matrix, all eigenvalues λ satisfy $|\lambda| \le 1$.*

Proposition 38 *For any linearly independent vectors x and $y \in \mathbb{R}^n$, there are real values of s and t such that $sx+ty$ has both negative and positive components.*

Proof. Certainly this is true if either x or y have both postive and negative components, so we may assume both have only positive components (the other cases of both negative or one positive and one negative are handled by adjusting the signs of s and t appropriately). We may now consider the vector

$$x = \left(\sum_{i=1}^{n} w_i \right) v - \left(\sum_{i=1}^{n} v_i \right) w.$$

Both the sums in the above expression are nonzero by assumption (in fact, positive). Also x is nonzero since v and w are linearly independent. Notice that

$$\sum_{i=1}^{n} x_i = 0.$$

Since x is not the zero vector, this implies that x must have both positive and negative components. ■

Thus, the matrix M can be used to produce unique rankings if there no dangling nodes.

12.2 Dealing with Dangling nodes

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

Theorem 39 *If A is a matrix with all positive entries, then A contains a real, positive eigenvalue ρ such that*

1. *For any other eigenvalue λ, we have $|\lambda| < \rho$ (recall that λ could be complex).*
2. *The eigenspace of ρ is one-dimensional and there is a unique eigenvector $x = [x_1, x_2, \ldots, x_p]^T$ with eigenvalue ρ 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 postive entries, as we got in the last section, we can use the Perron vector as the ranking.