Random Walks and Spectral Graph Drawing

Lecturer: Daniel A. Spielman September 18, 2007

4.1 Introduction

My main goals for today are to finish relating the mixing rate of random walks with the eigenvalues of the walk matrix, and to show how the eigenvectors of this matrix can be used to draw graphs.

Let's begin by recalling what we found out last lecture. The lazy walk matrix was defined by

$$(1/2)\left(\boldsymbol{I} + \boldsymbol{A}\boldsymbol{D}^{-1}\right), \tag{4.1}$$

If $\boldsymbol{p}_0$ is the initial distribution of a random walk, then the distribution after t steps is given by

$$\boldsymbol{p}_t \stackrel{\text{def}}{=} \boldsymbol{W}^t \boldsymbol{p}_0.$$

We saw that $\boldsymbol{W}$ has an eigenvalue of 1, with

$$\mathbf{1}\boldsymbol{W} = \mathbf{1},$$

and

$$\boldsymbol{W}\boldsymbol{d} = \boldsymbol{d}.$$

We also proved that all eigenvalues of $\boldsymbol{W}$ lie between 0 and 1.

We defined the stable distribution to be the vector

$$\boldsymbol{\pi} = \boldsymbol{d} / \sum_i d(i),$$

where $\boldsymbol{d}$ is the vector of degrees.

4.2 Uniqueness of the stable distribution

Lemma 4.2.1. *If the graph G is connected, then $\boldsymbol{x}\boldsymbol{W} = \boldsymbol{x}$ implies that $\boldsymbol{x} = c\mathbf{1}$ for some constant c.*

Proof. First, note that $\boldsymbol{x}\boldsymbol{W} = \boldsymbol{x}$ implies $\boldsymbol{x}\boldsymbol{A}\boldsymbol{D}^{-1} = \boldsymbol{x}$. Without loss of generality, assume that $\boldsymbol{x}(1) = 1$ and $\boldsymbol{x}(u) \leq 1$ for all $u \in V$. Let

$$S = \{u : \boldsymbol{x}(u) = 1\}.$$

If it were not the case that $\boldsymbol{x} = \mathbf{1}$, then there would be some vertex v not in S. As the graph is connected, this would imply that there would be a vertex $u \in S$ and a vertex $v \notin S$ for which $(u, v) \in E$. This would give a contradiction, as we would then have

$$\begin{aligned}
\boldsymbol{x}(u) &= (1/d(u)) \sum_{(u,w) \in E} \boldsymbol{x}(w), \\
&= (1/d(u)) \left(\boldsymbol{x}(v) + \sum_{(u,w) \in E, w \neq v} \boldsymbol{x}(w) \right), \\
&\leq (1/d(u)) \left(\boldsymbol{x}(v) + \sum_{(u,w) \in E, w \neq v} 1 \right), \\
&< (1/d(u)) \left(1 + \sum_{(u,w) \in E, w \neq v} 1 \right), \\
&< (1/d(u))(d(u)) \\
&= 1,
\end{aligned}$$

a contradiction. □

Corollary 4.2.2. *Every right-eigenvector of the lazy walk matrix of a connected graph is a multiple of* $\boldsymbol{\pi}$.

Proof. The dimension of the eigenspace of any eigenvalue, and in particular the eigenvalue 1, is the same from the left as from the right. Lemma 4.2.1 proves that this dimension is 1. □

4.3 Convergence

We will finish this lecture by proving that every lazy random walk on a connected graph converges to $\boldsymbol{\pi}$, and that the speed at which it converges is related to the eigenvalues of the matrix $\boldsymbol{W}$. Let $1 = \mu_1 > \mu_2 \geq \mu_3 \cdots \geq \mu_n$ be the eigenvalues of $\boldsymbol{W}$. We define the *spectral gap* of $\boldsymbol{W}$ to be

$$\lambda \stackrel{\text{def}}{=} 1 - \mu_2.$$

We will frequently use the fact that $\mu_i \leq (1 - \lambda)$ for $i \geq 2$.

We will prove:

Lemma 4.3.1. *For every initial distribution* $\boldsymbol{p}_0$, *and every* $t \geq 0$ *and every* $v \in V$

$$|\boldsymbol{p}_t(v) - \boldsymbol{\pi}(v)| \leq \sqrt{\frac{d(u)}{\min_v d(v)}} (1 - \lambda)^t.$$

We will first prove the lemma in the case in which the graph is regular. This case is easier, as then $\boldsymbol{W}$ is symmetric, but it contains the main points of the argument. This argument demonstrates what eigenvectors are really for—understanding what happens when one multiplies a vector by a matrix.

Proof of Lemma 4.3.1 in the regular case. When G is regular, the stable distribution is $\boldsymbol{\pi} = \frac{1}{n}\mathbf{1}$. So, we need to prove that

$$|\boldsymbol{p}_t(u) - 1/n| \leq (1-\lambda)^t. \tag{4.2}$$

We have that

$$|\boldsymbol{p}_t(u) - 1/n| \leq \|\boldsymbol{p}_t - \boldsymbol{\pi}\|,$$

so it suffices to upper bound this later term. Let $1 = \mu_1 > \mu_2 \geq \mu_3 \cdots \geq \mu_n$ be the eigenvalues of $\boldsymbol{W}$, and let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ be a corresponding orthnormal set of eigenvectors. Note that $\boldsymbol{v}_1 = \frac{1}{\sqrt{n}}\mathbf{1}$. The matrix $\boldsymbol{W}^t$ has the same eigenvectors, with the eigenvalues $\mu_1^t, \ldots, \mu_n^t$. So,

$$\begin{aligned}
\boldsymbol{p}_t = \boldsymbol{W}^t \boldsymbol{p}_0 &= \sum_{i=1}^n \mu_i^t \boldsymbol{v}_i (\boldsymbol{v}_i^T \boldsymbol{p}_0) \\
&= \boldsymbol{v}_1 (\boldsymbol{v}_1^T \boldsymbol{p}_0) + \sum_{i=2}^n \mu_i^t \boldsymbol{v}_i (\boldsymbol{v}_i^T \boldsymbol{p}_0).
\end{aligned}$$

Observe that

$$\boldsymbol{v}_1^T \boldsymbol{p}_0 = (1/\sqrt{n}) \sum_u \boldsymbol{p}_0(u) = 1/\sqrt{n},$$

so

$$\boldsymbol{v}_1 (\boldsymbol{v}_1^T \boldsymbol{p}_0) = \frac{1}{n}\mathbf{1} = \boldsymbol{\pi}.$$

Thus,

$$\boldsymbol{p}_t - \boldsymbol{\pi} = \sum_{i=2}^n \mu_i^t \boldsymbol{v}_i (\boldsymbol{v}_i^T \boldsymbol{p}_0), \quad \text{and}$$

$$\|\boldsymbol{p}_t - \boldsymbol{\pi}\| = \left\| \sum_{i=2}^n \mu_i^t \boldsymbol{v}_i (\boldsymbol{v}_i^T \boldsymbol{p}_0) \right\|.$$

Since the vectors $\boldsymbol{v}_i$ are orthonormal,

$$\begin{aligned}
\left\| \sum_{i=2}^n \mu_i^t \boldsymbol{v}_i (\boldsymbol{v}_i^T \boldsymbol{p}_0) \right\| &= \sqrt{\sum_{i=2}^n \left(\mu_i^t (\boldsymbol{v}_i^T \boldsymbol{p}_0)\right)^2} \\
&\leq \sqrt{\sum_{i=2}^n \left((1-\lambda)^t (\boldsymbol{v}_i^T \boldsymbol{p}_0)\right)^2} \\
&= (1-\lambda)^t \sqrt{\sum_{i=2}^n \left(\boldsymbol{v}_i^T \boldsymbol{p}_0\right)^2}.
\end{aligned}$$

Again using the orthonormality of the $\boldsymbol{v}_i$s, we find

$$\sqrt{\sum_{i=2}^{n} \left((\boldsymbol{v}_i^T \boldsymbol{p}_0) \right)^2} \leq \|\boldsymbol{p}_0\| \, ,$$

and since $\boldsymbol{p}_0$ is a non-negative vector that sums to 1, we have

$$\|\boldsymbol{p}_0\| \leq 1.$$

Putting all these inequalities together, we obtain

$$\|\boldsymbol{p}_t - \boldsymbol{\pi}\| \leq (1-\lambda)^t.$$

□

To figure out how large t has to be before this becomes small, we recall the inequality

$$(1-\lambda) \leq e^{-\lambda}.$$

So, we have

$$(1-\lambda)^t \leq e^{-\lambda t}.$$

If we now set $t = 2\ln n/\lambda$, then we get

$$(1-\lambda)^t \leq e^{-2\ln n} = n^{-2}.$$

So, by this time, the walk will be very close to the stable distribution.

While I won't have time to cover it in class, I will present the proof of Lemma 4.3.1 in the general case so that you can contemplate it at your leisure. The idea behind the proof is to apply the above reasoning to $\boldsymbol{N}$, and then to figure out the impact of multiplying by $\boldsymbol{D}^{-1/2}$ and $\boldsymbol{D}^{1/2}$ on either side.

Proof of Lemma 4.3.1. We will derive an upper bound on

$$\frac{1}{\sqrt{d(u)}} \left| \boldsymbol{p}_t(u) - \boldsymbol{\pi}(u) \right| \leq \left\| \boldsymbol{D}^{-1/2} \boldsymbol{W}^t \boldsymbol{p}_0 - D^{-1/2} \boldsymbol{\pi} \right\| .$$

First, note that

$$\boldsymbol{W} = \boldsymbol{D}^{1/2} \boldsymbol{N} \boldsymbol{D}^{-1/2},$$

implies

$$\boldsymbol{W}^t = \boldsymbol{D}^{1/2} \boldsymbol{N}^t \boldsymbol{D}^{-1/2},$$

and

$$\boldsymbol{D}^{-1/2} \boldsymbol{W}^t = \boldsymbol{N}^t \boldsymbol{D}^{-1/2}.$$

Thus, we need to upper bound

$$\left\| \boldsymbol{N}^t \boldsymbol{D}^{-1/2} \boldsymbol{p}_0 - D^{-1/2} \boldsymbol{\pi} \right\| .$$

Let $\boldsymbol{q}_0 = \boldsymbol{D}^{-1/2}\boldsymbol{p}_0$. As $\|\boldsymbol{p}_0\| \leq 1$, we have

$$\|\boldsymbol{q}_0\| \leq \frac{1}{\min_u d(u)}.$$

Let $\boldsymbol{v}_1$ be the eigenvector of eigenvalue 1 of $\boldsymbol{N}$. Fortunately, we will discover that

$$D^{-1/2}\boldsymbol{\pi} = \boldsymbol{v}_1 \left(\boldsymbol{v}_1^T \boldsymbol{q}_0\right). \tag{4.3}$$

To see this, recall from the correspondance described earlier that

$$\boldsymbol{v}_1(u) = \frac{\sqrt{d(u)}}{\sqrt{\sum_v d(v)}}.$$

Thus,

$$\boldsymbol{v}_1^T \boldsymbol{q}_0 = \frac{1}{\sqrt{\sum_u d(u)}} \sum_u d^{1/2}(u) d^{-1/2}(u) \boldsymbol{p}_0(u) = \frac{1}{\sqrt{\sum_v d(v)}},$$

and so

$$\left(\boldsymbol{v}_1 \left(\boldsymbol{v}_1^T \boldsymbol{q}_0\right)\right)(u) = \frac{\sqrt{d(u)}}{\sum_v d(v)} = \left(D^{-1/2}\boldsymbol{\pi}\right)(u).$$

Using the technique we applied in the regular case, we can show that

$$\left\|N^t \boldsymbol{q}_0 - \boldsymbol{D}^{-1/2}\boldsymbol{\pi}\right\| = \left\|N^t \boldsymbol{q}_0 - \boldsymbol{v}_1\left(\boldsymbol{v}_1^T \boldsymbol{q}_0\right)\right\| \leq (1-\lambda)^t \|\boldsymbol{q}_0\|.$$

The lemma now follows by combining this with the bound on the norm of $\boldsymbol{q}_0$.

□

4.4 Spectral Graph Drawing

We now know that the largest eigenvalue of the walk matrix is 1, and that it is unique if the graph is connected. We know the left and right eigenvectors of 1 ($\mathbf{1}$ and $\boldsymbol{\pi}$), and we know that the gap between the next eigenvalue and 1 controls the mixing rate of random walks. It's now time to learn a little about the eigenvectors of the eigenvalues close to 1. Let μ_i be the ith largest eigenvalue of W, and let $\boldsymbol{y}_i$ be the corresponding left-eigenvector.

$$\boldsymbol{y}_i W = \mu \boldsymbol{y}_i. \tag{4.4}$$

To consider the case in which μ_i is close to 1, let's write

$$\mu_i = 1 - \lambda_i,$$

and understand that in this case λ_i is small. Breaking equation (4.4) down vertex-by-vertex, we find that for each vertex $u \in V$,

$$(1/2)\boldsymbol{y}_i(u) + (1/2d(u)) \sum_{v:(u,v)\in E} \boldsymbol{y}_i(v) = (1-\lambda_i)\boldsymbol{y}_i(u),$$

which implies

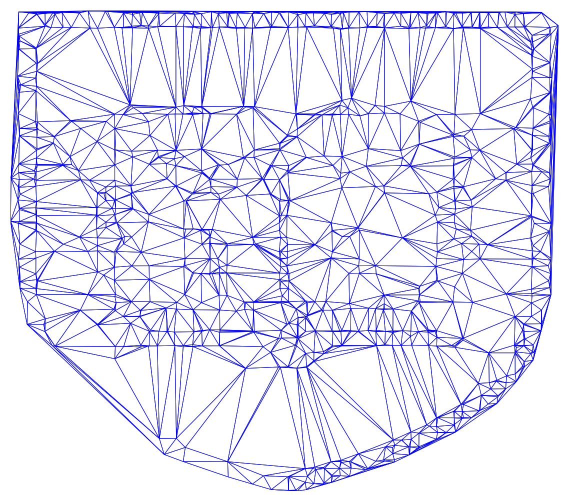
$$(1/d(u)) \sum_{v:(u,v)\in E} \boldsymbol{y}_i(v) = (1 - 2\lambda_i)\boldsymbol{y}_i(u).$$

The left-hand-side of this equation is the averge of the values assigned by $\boldsymbol{y}_i$ to the neighbors of u. So, if λ_i is small, this equation tells us that the value assigned to u by $\boldsymbol{y}_i$ is close to the average of the values assigned to its neighbors.

Intuitively, this is useful is we want to draw a picture of a graph. To draw a picture, we need two coordinates for each vertex. We will use $\boldsymbol{y}_2(u)$ and $\boldsymbol{y}_3(u)$. If both λ_2 and λ_3 are small, then each vertex will be drawn close to the average of the locations of its neigbors.

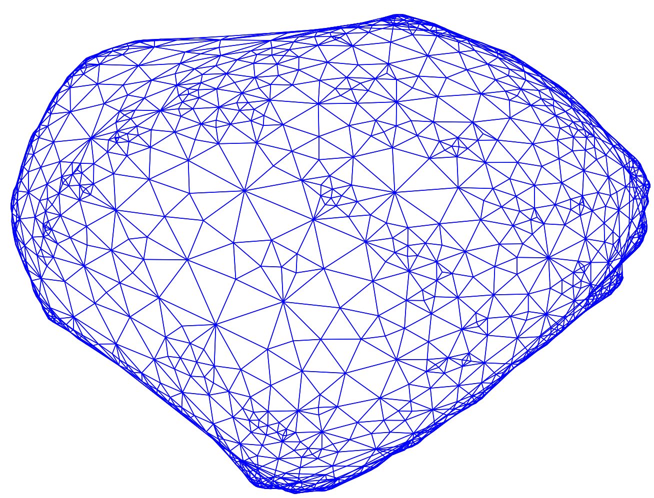
First, let's begin with a graph. This graph is a mesh that I drew on points sampled from the Yale shield.

```
load yaleShieldBig
gplot(a,xy)
```



Here is a drawing that I made from this graph by using the largest eigenvectors of the walk matrix.

```
s = sum(a);
n = length(a);
Dinv = diag(1./s);
W = (1/2)*(speye(n) + a*Dinv);
[v,D] = eigs(W',3);
gplot(a,v(:,2:3))
```

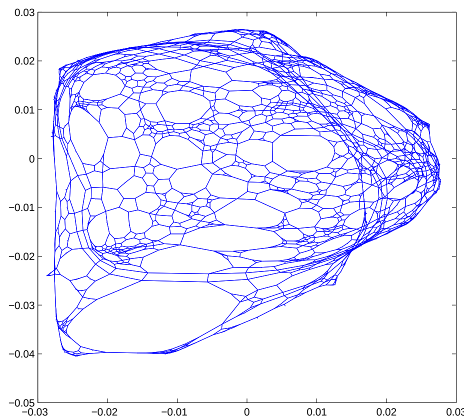


Let me point out a couple of things about this drawing. The first, and most important, is that it was made using only the adjacency structure of the graph. Yet, it makes a very nice picture! Also, note the picture is almost planar. Very few edges cross, and those that do are mostly lie near the boundaries.

Let's do some more examples.

The next graph is one I found surfing the web. The vertices are roads in Rome, and they have an edge between them if they meet at an intersection.

```
load rome
s = sum(a);
n = length(a);
Dinv = diag(1./s);
W = (1/2)*(speye(n) + a*Dinv);
[V,D] = eigs(W',3);
gplot(a,V(:,2:3))
```



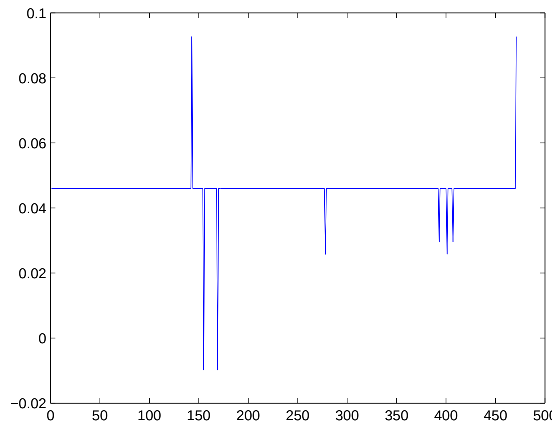
In the next graph, the vertices are everyone who co-authored a paper with Erdos, with edges between each pair who co-authored a paper with each other. Here is the beginning of the transcript of my Matlab session.

```
>> load erdosGraph
>> length(a)

ans =

   471

>> s = sum(a);
>> n = length(a);
>> Dinv = diag(1./s);
>> W = (1/2)*(speye(n) + a*Dinv);
>> [V,D] = eigs(W',10);

. . .

>> D

D =

    1.0000         0         0         0         0         0         0         0         0
         0    0.9753         0         0         0         0         0         0         0
         0         0    0.9646         0         0         0         0         0         0
```

```
     0         0         0    0.9501         0         0         0         0         0
     0         0         0         0    0.9436         0         0         0         0
     0         0         0         0         0    0.9416         0         0         0
     0         0         0         0         0         0    1.0000         0         0
     0         0         0         0         0         0         0    0.9384         0
     0         0         0         0         0         0         0         0    1.0000
     0         0         0         0         0         0         0         0         0
```

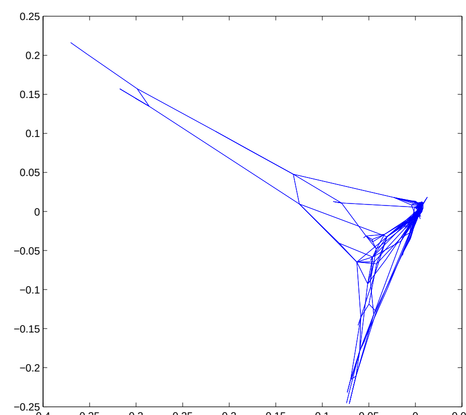
So, this graph has many eigenvalues of 1. This means that it has many connected components. In the following code, I will extract the largest component by looking at an eigenvector of eigenvalue 1. I will do it by grabing the vertices with the most common value. If you try this at home, the most common value might be different.

```
>> plot(V(:,1))
>> m = median(V(:,1));
>> ind = find(abs(V(:,1)-m) < .001);
>> length(ind)

ans =

   463
```



Now, let's plot the graph induced on this component.

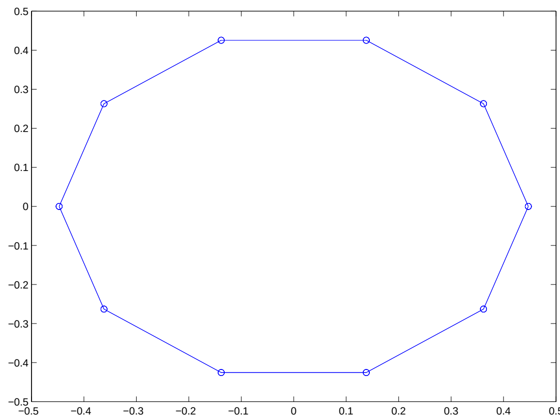
```
a = a(ind,ind);
s = sum(a);
Dinv = diag(1./s);
n = length(a);
W = (1/2)*(speye(n) + a*Dinv);
[V,D] = eigs(W',3);
D
gplot(a,V(:,2:3))
```



This graph just does not look too good. Don't worry about it. This technique won't always work. Some graphs just can't be drawn nicely, and some that can do not have nice pictures in their eigenvalues. Unfortunately, I don't know any good theorems explaining which graphs have nice pictures, and I don't know what such a theorem should look like.

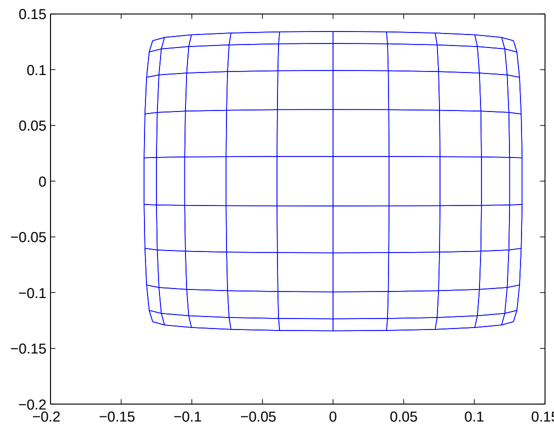
Let's see some graphs that do have nice spectral images. Here's the ring graph:

```
clear
clf
a = diag(ones(1,9),1);
a(1,10) = 1;
a = a + a';
s = sum(a);
Dinv = diag(1./s);
n = length(a);
W = (1/2)*(speye(n) + a*Dinv);
[V,D] = eigs(W',3);
gplot(a,V(:,2:3))
hold on
gplot(a,V(:,2:3),'o')
```



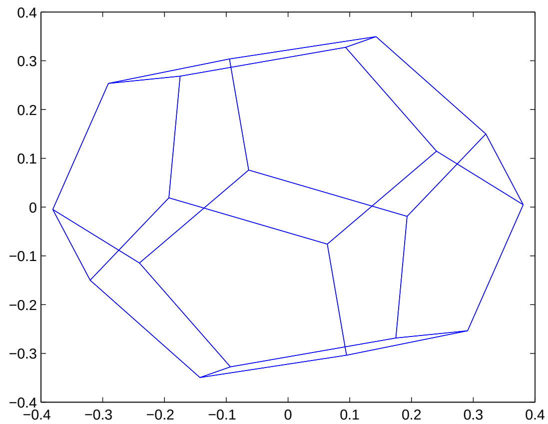
Now, let's try a 10-by-11 grid.

```
a = grid2(10,11);
s = sum(a);
Dinv = diag(1./s);
n = length(a);
W = (1/2)*(speye(n) + a*Dinv);
[V,D] = eigs(W',3);
gplot(a,V(:,2:3))
```



Here's another favorite: the dodecahedron.

```
load dodec.mat
s = sum(a);
n = length(a);
Dinv = diag(1./s);
W = (1/2)*(speye(n) + a*Dinv);
[V,D] = eigs(W',3);
D
gplot(a,V(:,2:3))
```

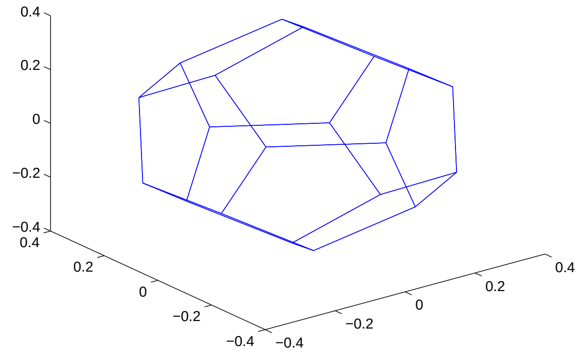


It looks like a squashed dodecahedron. The reason is that the dodecahedron is really represented by its first 3 non-trivial eigenvectors, all of which have the same eigenvalues. Here are the eigenvalues of the dodecahedron:

```
>> eig(full(W))

ans =

    0.1273
    0.1273
    0.1273
    0.1667
    0.1667
    0.1667
    0.1667
    0.5000
    0.5000
    0.5000
    0.5000
    0.6667
    0.6667
    0.6667
    0.6667
    0.6667
    0.8727
    0.8727
    0.8727
    1.0000
```

To properly look at this graph in 3d, try this.

```
[V,D] = eigs(W',4);
gplot3(a,V(:,2:4))
```



4.5 Spectral Gaps of Some Common Graphs

To finish, let me give you the orders of magnitudes of the spectral gaps of some common graphs.

Graph	λ_2
Path on n vertices	$\Theta(1/n^2)$
Ring on n vertices	$\Theta(1/n^2)$
Complete binary tree on n vertices	$\Theta(1/n)$
$\sqrt{n}$-by-$\sqrt{n}$ grid	$\Theta(1/n)$
Hypercube on n vertices	$\Theta(1/\log n)$
Expander graph	$\Theta(1)$

I should also point out that Teng and I proved (Linear Algebra and its Applications, March 2007) that for every planar graph $\lambda_2 \leq \frac{8d_{max}}{d_{min}n}$, where d_{max} and d_{min} are the maximum and minimum degrees of vertices, respectively.