Eigenvalues and Random Walks
(section 1.5 of "Spectral Graph Theory" by Fan Chung)

Weighted (undirected) graphs: we have a weight function w: $V \times V \to \mathbb{R}$ satisfying

(a) $w(u, v) = w(v, u)$
(b) $w(u, v) \geq 0$ and if u and v are not adjacent, $w(u, v) = 0$.

Then the degree d_v of a vertex v is defined by:

$d_v = \sum_u w(u, v)$
vol $G = \sum_v d_v$

The normalized Laplacian of G is still defined as $T^{-1/2} L T^{-1/2}$, so we have

$$\mathcal{L}(u, v) = \begin{cases} 1 - w(v, v) / d_v, & \text{if } u = v \\ - w(u, v) / (d_u d_v)^{1/2}, & \text{otherwise} \end{cases}$$

The Rayleigh quotient characterization for the eigenvalues can be readily generalized for weighted graphs.

Given a weighted graph G, we can define a random walk on it in a natural way, by specifying the transition probabilities:

$$P(u, v) = \mathrm{w}(u, v) / d_u$$

Clearly, $\sum_v P(u, v) = 1$.

For our Markov chain, if the initial distribution is $f: V \to \mathrm{R}$, $\sum_v f(v) = 1$, the distribution after k steps is given by fP^k.

We call a Markov chain ergodic if there is a unique stationary distribution $\pi(v)$ satisfying

$$\lim fP^k(v) = \pi(v)$$

for any initial distribution f.

From the theory of finite Markov chains we know that the necessary and sufficient conditions for the ergodicity of P are:

(i) irreducibility (for any u, v, there exists s, such that $P^s(u, v) > 0$)
(ii) aperiodicity (GCD$\{s: P^s(u, u) > 0\} = 1$, which is a communicating class property)

It is easy to see that these two properties can be conveniently stated in the spectral terms. This is what we have for undirected graphs:

irreducibility is equivalent to the condition that G is connected, that is, $\lambda_1 > 0$.

aperiodicity is equivalent to the condition that G is not bipartite, that is, $\lambda_{n-1} < 2$.

A major problem of interest is the convergence rate of a given ergodic random walk:

Given an arbitrary initial distribution f, how many steps s are required for fP^s to be close to the stationary distribution?

As we will see shortly, the answer can be given in the terms of “spectral gap”, determined by λ_1 and λ_{n-1}.

We start with the L_2-norm convergence: $\| fP^s - \pi \|_2$.

Note that $P = T^{-1}A = T^{-1/2}(I - \mathcal{L})T^{1/2}$.

The stationary distribution must satisfy $\pi P = \pi$ (if $a_s \to \pi$, then $a_s P \to \pi P$).

Since $(1T)P = 1A = (d_1, \ldots, d_n) = 1T$, we see that $\pi = (1/\mathrm{vol}\ G)\ (d_1, \ldots, d_n)$ is a natural candidate for the stationary distribution.

Let $\{\varphi_i\}$ be the orthonormal eigenbasis of $\mathcal{L}$, where φ_i is associated with λ_i. Given an initial distribution f, we write $f\, T^{-1/2} = \sum_i a_i\, \varphi_i$.

We know that $\varphi_0 = (\mathrm{vol}\ G)^{-1/2}\ 1T^{1/2}$, so $a_0 = (\mathrm{vol}\ G)^{-1/2}$.

Then we have:

$$\| fP^s - \pi \| = \| fP^s - (1/\mathrm{vol}\ G)\ 1T \| = \| fP^s - a_0\, \varphi_0\, T^{1/2} \| =$$

$$\| f\, T^{-1/2}(I - \mathcal{L})^s T^{1/2} - a_0\, \varphi_0\, T^{1/2} \| = \| (\textstyle\sum_i a_i\, \varphi_i)\ (I - \mathcal{L})^s T^{1/2} - a_0\, \varphi_0\, T^{1/2} \| =$$

(since $\mathcal{L}$ is symmetric, $\varphi_i \mathcal{L} = \lambda_i \varphi_i$)

$= \| \sum_{i>0} a_i (1 - \lambda_i)^s \varphi_i T^{1/2} \| = \| (\sum_{i>0} a_i (1 - \lambda_i)^s \varphi_i) T^{1/2} \| \leq$

$(\|(x_1, \ldots, x_n) T^{1/2} \| = \|(x_1 d_1^{1/2}, \ldots, x_n d_n^{1/2})\| \leq d_{\max}^{1/2} \|(x_1, \ldots, x_n)\|)$

$\leq d_{\max}^{1/2} \| \sum_{i>0} a_i (1 - \lambda_i)^s \varphi_i \| \leq d_{\max}^{1/2} (\sum_{i>0} a_i^2 (1 - \lambda_i)^{2s})^{1/2} \leq$

$\leq d_{\max}^{1/2} (\sum_{i>0} a_i^2)^{1/2} \max_{i>0} | 1 - \lambda_i |^s =$

(define λ' as λ_1 if $(1 - \lambda_1) \geq (\lambda_{n-1} - 1)$ or as $(2 - \lambda_{n-1})$ otherwise)

$= d_{\max}^{1/2} (\sum_{i>0} a_i^2)^{1/2} | 1 - \lambda' |^s \leq$

$(\sum_i a_i^2 = \| f T^{-1/2} \|^2 = \sum_i (f_i^2/d_i) \leq (\sum_i f_i) / d_{\min} = 1/d_{\min})$

$\leq (d_{\max}/d_{\min})^{1/2} | 1 - \lambda' |^s \leq \exp(-s\lambda') (d_{\max}/d_{\min})^{1/2}$ (since $(1 - \lambda') \leq \exp(-\lambda')$ on $[0, 1]$)

How many steps do we need to guarantee $\| fP^s - \pi \| \leq \varepsilon$?

$s \geq (1/\lambda') \ln(1/\varepsilon \cdot (d_{\max}/d_{\min})^{1/2})$

We can eliminate the dependence on $\lambda_{n\text{-}1}$ by modifying the random walk slightly. Let's modify the weights in the following way:

$w'(v, v) = w(v, v) + c\, d_v$, where $c = (\lambda_1 + \lambda_{n\text{-}1}) / 2 - 1$
(note that we have $c > 0$ if $(1 - \lambda_1) < (\lambda_{n\text{-}1} - 1)$)
and $w'(u, v) = w(u, v)$ if $u \neq v$.

Then we have $\lambda_k' = \lambda_k / (1 + c)$.

So, $1 - \lambda_1' = \lambda_{n\text{-}1}' - 1 = (\lambda_{n\text{-}1} - \lambda_1) / (\lambda_1 + \lambda_{n\text{-}1})$.

This is called a *lazy* random walk.

A stronger notion of convergence: the *relative pointwise distance*.

We know that every row of matrix P of an ergodic random walk converges to π.

We define the relative pointwise distance as:

$$\Delta(s) = \max_{x,y} | P^s(y, x) - \pi(x) | / \pi(x)$$

Similar to the above, we can show that

$$\Delta(s) \leq \exp(-s\lambda') (\mathrm{vol}\, G / d_{\min})$$

Why is the relative pointwise distance a stronger notion of convergence than the L_2-norm one?
Given an initial distribution f, we have

$$| fP^s(x) - \pi(x) | / \pi(x) \leq \sum_y f(y) (| P^s(y, x) - \pi(x) | / \pi(x)) \leq \sum_y f(y) \Delta(s) \leq \Delta(s)$$

So, we obtained $\| fP^s - \pi \|_2 \leq \Delta(s) \|\pi\|_2$

The Total Variation distance.

$\Delta_{TV}(s) = 1/2 \max_y \sum_x | P^s(y, x) - \pi(x) |$

(half of the L_1-distance)

Easy to see that $\Delta_{TV}(s) \leq \Delta(s) / 2$.

The case of directed graphs.

We can show the following:

If G is a strongly connected directed graph on n vertices, then we can define a lazy walk, such that after at most $t \geq 2/\lambda_1 ((-\ln \varphi_{min}) + 2c)$ steps, we have

$\Delta_{TV}(t) \leq \exp(-c)/2$,

where φ is the normalized Perron vector.

A subtlety: for directed graphs, the Perron vector components and eigenvalues can be exponentially small in n. However, for certain “well-behaving” classes of graphs those values are of the order of $1 / \mathrm{poly}(n)$. For instance, that holds true for *Eulerian* graphs (in-degree of each vertex is equal to its out-degree).