Analysis and Geometry on Graphs
Part 1. Laplace operator on weighted graphs

Alexander Grigor'yan

Tsinghua University, Winter 2012

1 Weighted graphs and Markov chains

The notion of a graph. A graph is a couple (V, E) where V is a set of *vertices*, that is, an arbitrary set, whose elements are called vertices, and E is a set of *edges*, that is, E consists of some couples (x, y) where $x, y \in V$. We write $x \sim y$ (x is connected to y, or x is joint to y, or x is *adjacent* to y, or x is a *neighbor* of y) if $(x, y) \in E$. The edge (x, y) will be normally denoted by $\overline{xy}$. In this section we assume that the graphs are undirected so that $\overline{xy} \equiv \overline{yx}$. The vertices x, y are called the *endpoints* of the edge $\overline{xy}$. The edge $\overline{xx}$ with the same endpoints (should it exist) is called a *loop*. Generally we allow loops although in the most examples our graphs do not have loops.

A graph (V, E) is called *finite* if the number $\#V$ of vertices is finite. In this course all graphs are finite unless otherwise stated. For each vertex x, define its degree

$$\deg(x) = \#\{y \in V : x \sim y\},$$

that is, $\deg(x)$ is the number of neighbors of x. A graph is called *regular* if $\deg(x)$ is the same for all $x \in V$.

Example. Consider some examples of graphs.

1. A *complete graph* K_n. The set of vertices is $V = \{1, 2, ..., n\}$, and the edges are defined as follows: $i \sim j$ for any two distinct $i, j \in V$.

2. A *complete bipartite* graph $K_{n,m}$. The set of vertices is $V = \{1, .., n, n+1, ..., n+m\}$, and the edges are defined as follows: $i \sim j$ if either $i < n$ and $j \geq n$ or $i \geq n$ and $j < n$. That is, the set of vertices is split into two groups: $V_1 = \{1, ..., n\}$ and $V_2 = \{n+1, ..., m\}$, and the vertices are connected if and only if they belong to the different groups.

3. A *cycle graph,* denoted by $\mathbb{Z}_m$. The set of vertices is the set of residues $\bmod m$ that is, $V = \{0, 1, ..., m-1\}$, and $i \sim j$ if $i - j = \pm 1 \bmod m$.

4. A *path graph* P_m. The set of vertices is $V = \{0, 1, ..., m-1\}$, and $i \sim j$ if $|i - j| = 1$.

For example,

$$K_2 = K_{1,1} = \mathbb{Z}_2 = P_2 = \bullet - \bullet$$

$$\mathbb{Z}_3 = \begin{array}{ccc} & & \bullet \\ & / & | \\ \bullet & - & \bullet \end{array}, \quad \mathbb{Z}_4 = K_{2,2} = \begin{array}{ccc} \bullet & - & \bullet \\ | & & | \\ \bullet & - & \bullet \end{array}$$

Product of graphs. More interesting examples of graphs can be constructed using the operation of product of graphs.

Definition. Let (X, E_1) and (Y, E_2) be two graphs. Their *Cartesian product* is defined as follows:

$$(V, E) = (X, E_1) \,\square\, (Y, E_2)$$

where $V = X \times Y$ is the set of pairs (x, y) where $x \in X$ and $y \in Y$, and the set E of edges is defined by

$$(x, y) \sim (x', y) \text{ if } x' \sim x \quad \text{and} \quad (x, y) \sim (x, y') \quad \text{if } y \sim y', \tag{1.1}$$

which is illustrated on the following diagram:

$$\begin{array}{cccccc}
y'\bullet & \dots & \overset{(x,y')}{\bullet} & - & \overset{(x',y')}{\bullet} & \dots \\
| & & | & & | & \\
y\bullet & \dots & \overset{(x,y)}{\bullet} & - & \overset{(x',y)}{\bullet} & \dots \\
\vdots & & \vdots & & \vdots & \\
{}^{Y}\diagup_{X} & \dots & x\bullet & - & x'\bullet & \dots
\end{array}$$

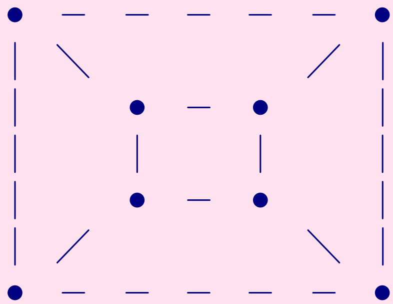
Clearly, we have $\#V = (\#X)\,(\#Y)$ and $\deg\,(x,y) = \deg\,(x) + \deg\,(y)$ for all $x \in X$ and $y \in Y$.

For example, we have

$$\mathbb{Z}_2 \square \mathbb{Z}_2 = \mathbb{Z}_4 = \begin{array}{ccc}
\bullet & - & \bullet \\
| & & | \\
\bullet & - & \bullet
\end{array}, \qquad P_4 \square P_3 = \begin{array}{ccccccc}
\bullet & - & \bullet & - & \bullet & - & \bullet \\
| & & | & & | & & | \\
\bullet & - & \bullet & - & \bullet & - & \bullet \\
| & & | & & | & & | \\
\bullet & - & \bullet & - & \bullet & - & \bullet
\end{array}$$

This definition can be iterated to define the product of a finite sequence of graphs. The graph $\mathbb{Z}_2^n := \underbrace{\mathbb{Z}_2 \square \mathbb{Z}_2 \square ... \square \mathbb{Z}_2}_{n}$ is called the n-dimensional binary cube. For example,

$\mathbb{Z}_2^3 =$



The graph distance.

Definition. A finite sequence $\{x_k\}_{k=0}^n$ of vertices on a graph is called a *path* if $x_k \sim x_{k+1}$ for all $k = 0, 1, ..., n-1$. The number n of edges in the path is referred to as the *length* of the path.

Definition. A graph (V, E) is called *connected* if, for any two vertices $x, y \in V$, there is a path connecting x and y, that is, a path $\{x_k\}_{k=0}^n$ such that $x_0 = x$ and $x_n = y$. If (V, E) is connected then define the *graph distance* $d(x,y)$ between any two distinct vertices x, y as follows: if $x \neq y$ then $d(x,y)$ is the minimal length of a path that connects x and y, and if $x = y$ then $d(x,y) = 0$.

The connectedness here is needed to ensure that $d(x,y) < \infty$ for any two points. It is easy to see that on any connected graph, the graph distance is a metric, so that (V, d) is a metric space.

Weighted graphs.

Definition. A *weighted graph* is a couple $((V,E), \mu)$ where (V,E) is a graph and μ_{xy} is a non-negative function on $V \times V$ such that

1. $\mu_{xy} = \mu_{yx}$;
2. $\mu_{xy} > 0$ if and only if $x \sim y$.

The weighted graph can also be denoted by (V, μ) because the weight μ contains all information about the set of edges E.

Example. Set $\mu_{xy} = 1$ if $x \sim y$ and $\mu_{xy} = 0$ otherwise. Then μ_{xy} is a weight. This specific weight is called *simple*.

Any weight μ_{xy} gives rise to a function on vertices as follows:

$$\mu(x) = \sum_{\{y \in V, y \sim x\}} \mu_{xy}. \tag{1.2}$$

Then $\mu(x)$ is called the weight of a vertex x. It gives rise to a measure of subsets: for any subset $A \subset V$, define its measure by $\mu(A) = \sum_{x \in A} \mu(x)$.

For example, if the weight μ_{xy} is simple then $\mu(x) = \deg(x)$ and $\mu(A) = \sum_{x \in A} \deg(x)$.

Markov chains. Let V be a finite set and $P(x,y)$ be a *Markov kernel* on V, that is, a non-negative function on $V \times V$ with the property that

$$\sum_{y \in V} P(x,y) = 1 \text{ for all } x \in V. \tag{1.3}$$

Any Markov kernel gives rise to a *Markov chain* $\{X_n\}_{n=0}^{\infty}$ that is a *random walk* on V. It is determined by a family $\{\mathbb{P}_x\}_{x \in V}$ of probability measures on the set of all paths starting from x (that is, $X_0 = x$), that satisfies the following property: for all positive integers n and all $x, x_1, ..., x_n \in V$,

$$\mathbb{P}_x \left(X_1 = x_1, X_2 = x_2, ..., X_n = x_n\right) = P\left(x, x_1\right) P\left(x_1, x_2\right) ... P\left(x_{n-1}, x_n\right). \tag{1.4}$$

In other words, if $X_0 = x$ then the probability that the walk $\{X_n\}$ visits successively the vertices $x_1, x_2, ..., x_n$ is equal to the right hand side of (1.4).

$$\begin{array}{ccccccccccc} & & \bullet x_1 & & & & x_3\bullet & & & & x_n\bullet \\ & \nearrow & & \searrow & & \nearrow & & \ddots \quad \dots \quad \ddots & & \nearrow & \\ x\bullet & & & & \bullet x_2 & & & & x_{n-1}\bullet & & \end{array}$$

For any positive integer n and any $x \in V$, set

$$P_n\left(x, y\right) = \mathbb{P}_x\left(X_n = y\right).$$

The function $P_n\left(x, y\right)$ is called the *transition function* or the *transition probability* of the Markov chain. For a fixed n and $x \in V$, the function $P_n\left(x, \cdot\right)$ nothing other than the distribution measure of the random variable X_n.

For $n = 1$ we obtain from (1.4) $P_1(x, y) = P(x, y)$. Let us state without proofs some easy consequences of (1.4).

1. For all $x, y \in V$ and positive integer n,

$$P_{n+1}(x, y) = \sum_{z \in V} P_n(x, z) P(z, y). \tag{1.5}$$

2. Moreover, for all positive integers n, k,

$$P_{n+k}(x, y) = \sum_{z \in V} P_n(x, z) P_k(z, y). \tag{1.6}$$

3. $P_n(x, y)$ is also a Markov kernel, that is,

$$\sum_{y \in V} P_n(x, y) = 1. \tag{1.7}$$

Reversible Markov chains. Let (V, μ) be a finite weighted graph without isolated vertices (the latter is equivalent to $\mu(x) > 0$). The weight μ induces a natural Markov kernel

$$P(x,y) = \frac{\mu_{xy}}{\mu(x)}. \tag{1.8}$$

Since $\mu(x) = \sum_{y \in V} \mu_{xy}$, we see that $\sum_y P(x,y) \equiv 1$ so that $P(x,y)$ is indeed a Markov kernel. For example, if μ is a simple weight then $\mu(x) = \deg(x)$ and

$$P(x,y) = \begin{cases} \frac{1}{\deg(x)}, & y \sim x, \\ 0, & y \not\sim x. \end{cases} \tag{1.9}$$

The Markov kernel (1.8) has an additional specific property

$$P(x,y)\,\mu(x) = P(y,x)\,\mu(y), \tag{1.10}$$

that follows from $\mu_{xy} = \mu_{yx}$.

Definition. An arbitrary Markov kernel $P(x,y)$ is called *reversible* if there is a positive function $\mu(x)$ with the property (1.10). Function μ is called then the *invariant measure* of P.

It follows easily from (1.5) that if $P(x,y)$ is reversible then $P_n(x,y)$ is also reversible with the same invariant measure.

Hence, the Markov kernel (1.8) is reversible. Conversely, any reversible Markov chain on V gives rise to a weighted graph structure on V as follows. Indeed, define μ_{xy} by

$$\mu_{xy} = P(x,y)\,\mu(x),$$

so that μ_{xy} is symmetric by (1.10).

From now on, we stay in the following setting: we have a finite weighted graph (V,μ) without isolated vertices, the associated reversible Markov kernel $P(x,y)$, and the corresponding random walk (= Markov chain) $\{X_n\}$. Fix a point $x_0 \in V$ and consider the functions

$$v_n(x) = \mathbb{P}_{x_0}(X_n = x) = P_n(x_0,x) \quad \text{and} \quad u_n(x) = \mathbb{P}_x(X_n = x_0) = P_n(x,x_0)$$

The function $v_n(x)$ is the distribution of X_n at time $n \geq 1$. By (1.7), we have

$$\sum_{x \in V} v_n(x) = 1.$$

Function $u_n(x)$ is somewhat more convenient to be dealt with. Using the reversibility of P_n, we see that the function v_n and u_n are related follows:

$$v_n(x) = \frac{u_n(x)\,\mu(x)}{\mu(x_0)}. \tag{1.11}$$

Extend u_n and v_n to $n=0$ by setting $u_0 = v_0 = \mathbf{1}_{\{x_0\}}$, where $\mathbf{1}_A$ denotes the indicator function of a set $A \subset V$, that is, the function that has value 1 at any point of A and value 0 outside A. It follows easily from (1.5) that v_n satisfies the following recursive equation:

$$v_{n+1}(x) = \sum_y \frac{1}{\mu(y)} v_n(y)\,\mu_{xy} \tag{1.12}$$

called the *forward equation*. Substituting here v_n from (1.11), we obtain the equation for u_n

$$u_{n+1}(x) = \frac{1}{\mu(x)} \sum_y u_n(y)\,\mu_{xy} \tag{1.13}$$

that is called the *backward equation*.

In particular, for a simple random walk we have $\mu_{xy} = 1$ for $x \sim y$ and $\mu(x) = \deg(x)$ so that we obtain the following equations:

$$v_{n+1}(x) = \sum_{y \sim x} \frac{1}{\deg(y)} v_n(y).$$

$$u_{n+1}(x) = \frac{1}{\deg(x)} \sum_{y \sim x} u_n(y).$$

The last identity means that $u_{n+1}(x)$ is the mean-value of $u_n(y)$ taken at the points $y \sim x$. Note that in the case of a regular graph, when $\deg(x) \equiv \text{const}$, we have $u_n \equiv v_n$ by (1.11).

Example. On graph $\mathbb{Z}_m$ we have

$$u_{n+1}(x) = \frac{1}{2}\left(u_n(x-1) + u_n(x+1)\right).$$

The following table contains computation of $u_n(x)$ in $\mathbb{Z}_3$ with $x_0 = 1$

$x \diagdown n$	0	1	2	3	4	5	6
0	0	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{5}{16}$	$\frac{11}{32}$	$\frac{21}{64}$
1	1	0	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{5}{16}$	$\frac{11}{32}$
2	0	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{5}{16}$	$\frac{11}{32}$	$\frac{21}{64}$

Here one can observe that the function $u_n(x)$ converges to a constant function $1/3$ as $n \to \infty$ and later we will prove this. Hence, for large n, the probability that X_n visits a given point is nearly $1/3$, which should be expected.

Here are the values of $u_n(x)$ in $\mathbb{Z}_5$ with $x_0 = 2$:

$x \diagdown n$	0	1	2	3	4	...	25
0	0	0	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{4}$	...	0.199
1	0	$\frac{1}{2}$	0	$\frac{3}{8}$	$\frac{1}{16}$	...	0.202
2	1	0	$\frac{1}{2}$	0	$\frac{3}{8}$	...	0.198
3	0	$\frac{1}{2}$	0	$\frac{3}{8}$	$\frac{1}{16}$	...	0.202
4	0	0	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{4}$	...	0.199

Here $u_n(x)$ approaches to $\frac{1}{5}$ as $n \to 5$ but the convergence is slower than in the case of $\mathbb{Z}_3$.

Example. On a complete graph K_m we have

$$u_{n+1}(x) = \frac{1}{m-1} \sum_{y \neq x} u_n(y).$$

Here are the values of $u_n(x)$ on K_5 with $x_0 = 2$:

$x \diagdown n$	0	1	2	3	4
0	0	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{13}{64}$	0.199
1	0	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{13}{64}$	0.199
2	1	0	$\frac{1}{4}$	$\frac{3}{16}$	0.203
3	0	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{13}{64}$	0.199
4	0	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{13}{64}$	0.199

We see that $u_n(x) \to \frac{1}{5}$ but the rate of convergence is much faster than for $\mathbb{Z}_5$. Although $\mathbb{Z}_5$ and K_5 has the same number of vertices, the extra edges in K_5 allow a quicker mixing than in the case of $\mathbb{Z}_5$.

As we will see, for finite graphs it is typically the case that the transition function $u_n(x)$ converges to a constant as $n \to \infty$. For the function v_n this means that

$$v_n(x) = \frac{u_n(x)\,\mu(x)}{\mu_0(x)} \to c\mu(x) \quad \text{as } n \to \infty$$

for some constant c. The constant c is determined by the requirement that $c\mu(x)$ is a probability measure on V, that is, from the identity

$$c\sum_{x\in V}\mu(x)=1.$$

Hence, $c\mu(x)$ is asymptotically the distribution of X_n as $n\to\infty$. . The function $c\mu(x)$ on V is called the *stationary measure* or the *equilibrium measure* of the Markov chain. One of the problems for finite graphs that will be discussed in this course, is the rate of convergence of $v_n(x)$ to the equilibrium measure. The point is that X_n can be considered for large n as a random variable with the distribution function $c\mu(x)$ so that we obtain a natural generator of a random variable with a prescribed law. However, in order to be able to use this, one should know for which n the distribution of X_n is close enough to the equilibrium measure. The value of n, for which this is the case, is called the *mixing time.*

2 The Laplace operator

The definition of the Laplace operator. The Laplace operator on functions in $\mathbb{R}^2$ is defined by

$$\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}.$$

Using the approximation of the second derivative

$$\begin{aligned} \frac{\partial^2 f}{\partial x^2}(x,y) &\approx \frac{f(x+h,y) - 2f(x,y) + f(x-h,y)}{h^2} \\ \frac{\partial^2 f}{\partial x^2}(x,y) &\approx \frac{f(x,y+h) - 2f(x,y) + f(x,y-h)}{h^2}, \end{aligned}$$

we obtain

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

Restricting f to the grid $h\mathbb{Z}^2$ and define the edges in $h\mathbb{Z}^2$ as on the product graph, we see that

$$\Delta f(x,y) \approx \frac{4}{h^2}\left(\frac{1}{4}\sum_{(x',y')\sim(x,y)} f(x',y') - f(x,y)\right).$$

The expression in the parenthesis is called the discrete Laplace operator on $h\mathbb{Z}^2$.

This notion can be defined on any weighted graph as follows.

Definition. Let (V,μ) be a finite weighted graph without isolated points. For any function $f: V \to \mathbb{R}$, define the function $\Delta_\mu f$ by

$$\Delta_\mu f(x) = \frac{1}{\mu(x)}\sum_{y\sim x} f(y)\,\mu_{xy} - f(x). \tag{2.1}$$

The operator Δ_μ is called the (*weighted*) *Laplace operator* of (V,μ).

This operator can also be written in equivalent forms as follows:

$$\Delta_\mu f(x) = \frac{1}{\mu(x)}\sum_{y\in V} f(y)\,\mu_{xy} - f(x) = \frac{1}{\mu(x)}\sum_{y\in V}\left(f(y) - f(x)\right)\mu_{xy}. \tag{2.2}$$

Example. If μ is a simple weight then we obtain the Laplace operator of the graph (V, E):

$$\Delta f(x) = \frac{1}{\deg(x)} \sum_{y \sim x} f(y) - f(x).$$

Denote by $\mathcal{F}$ the set of all real-valued functions on V. Then $\mathcal{F}$ is obviously a linear space with respect to addition of functions and multiplication by a constant. It is easy to see that $\dim \mathcal{F} = \# V$.

The Laplace operator Δ_μ can be regarded as a linear operator in $\mathcal{F}$. Another useful property that follows from (2.2) is

$$\Delta_\mu \operatorname{const} = 0.$$

In terms of the corresponding reversible Markov kernel $P(x, y) = \frac{\mu_{xy}}{\mu(x)}$ we can write

$$\Delta_\mu f(x) = \sum_y P(x, y) f(y) - f(x).$$

Defining the Markov operator P on $\mathcal{F}$ by

$$Pf(x) = \sum_y P(x, y) f(y),$$

we see that the Laplace operator Δ_μ and the Markov operator P are related by a simple identity $\Delta_\mu = P - \mathrm{id}$, where id is the identity operator in $\mathcal{F}$.

Green's formula. Let us consider the *difference operator* ∇_{xy} that is defined for any two vertices $x, y \in V$ and maps $\mathcal{F}$ to $\mathbb{R}$ as follows:

$$\nabla_{xy} f = f(y) - f(x).$$

The relation between the Laplace operator Δ_μ and the difference operator is given by

$$\Delta_\mu f(x) = \frac{1}{\mu(x)} \sum_y (\nabla_{xy} f) \mu_{xy} = \sum_y P(x, y) (\nabla_{xy} f)$$

The following theorem is one of the main tools when working with the Laplace operator. For any subset Ω of V, denote by Ω^c the complement of Ω, that is, $\Omega^c = V \setminus \Omega$.

Theorem 2.1 (Green's formula) *Let (V, μ) be a finite weighted graph without isolated points, and let Ω be a non-empty finite subset of V. Then, for any two functions f, g on V,*

$$\sum_{x\in\Omega} \Delta_\mu f(x) g(x) \mu(x) = -\frac{1}{2} \sum_{x,y\in\Omega} (\nabla_{xy} f)(\nabla_{xy} g)\, \mu_{xy} + \sum_{x\in\Omega, y\in\Omega^c} (\nabla_{xy} f)\, g(x) \mu_{xy} \tag{2.3}$$

The formula (2.3) is analogous to the Green formula for the Laplace operator in $\mathbb{R}^2$: if Ω is a bounded domain in $\mathbb{R}^2$ with smooth boundary then, for all smooth enough functions f, g on $\overline{\Omega}$

$$\int_\Omega (\Delta f)\, g dx = -\int_\Omega \nabla f \cdot \nabla g dx + \int_{\partial\Omega} \frac{\partial f}{\partial \nu} g d\ell$$

where ν is the unit normal vector field on $\partial\Omega$ and $d\ell$ is the length element on $\partial\Omega$.

If $\Omega = V$ then Ω^c is empty so that the last "boundary" term in (2.3) vanishes, and we obtain

$$\sum_{x\in V} \Delta_\mu f(x) g(x) \mu(x) = -\frac{1}{2} \sum_{x,y\in V} (\nabla_{xy} f)(\nabla_{xy} g)\, \mu_{xy}. \tag{2.4}$$

Proof. We have

$$\begin{aligned}
\sum_{x\in\Omega}\Delta_\mu f(x)g(x)\mu(x) &= \sum_{x\in\Omega}\left(\frac{1}{\mu(x)}\sum_{y\in V}(\nabla_{xy}f)\,\mu_{xy}\right)g(x)\mu(x) \\
&= \sum_{x\in\Omega}\sum_{y\in V}(\nabla_{xy}f)\,g(x)\mu_{xy} \\
&= \sum_{x\in\Omega}\sum_{y\in\Omega}(\nabla_{xy}f)\,g(x)\mu_{xy} + \sum_{x\in\Omega}\sum_{y\in\Omega^c}(\nabla_{xy}f)\,g(x)\mu_{xy} \\
&= \sum_{y\in\Omega}\sum_{x\in\Omega}(\nabla_{yx}f)\,g(y)\mu_{xy} + \sum_{x\in\Omega}\sum_{y\in\Omega^c}(\nabla_{xy}f)\,g(x)\mu_{xy},
\end{aligned}$$

where in the last line we have switched notation of the variables x and y in the first sum using $\mu_{xy}=\mu_{yx}$. Adding together the last two lines and dividing by 2, we obtain

$$\sum_{x\in\Omega}\Delta_\mu f(x)g(x)\mu(x) = \frac{1}{2}\sum_{x,y\in\Omega}(\nabla_{xy}f)\,(g(x)-g(y))\,\mu_{xy} + \sum_{\substack{x\in\Omega\\ y\in\Omega^c}}(\nabla_{xy}f)\,g(x)\mu_{xy},$$

which was to be proved. ■

Eigenvalues of the Laplace operator. As was already mentioned, the Laplace operator Δ_μ is a linear operator in a N-dimensional vector space $\mathcal{F}$ where $N = \#V$. Let us investigate the spectral properties of this operator. In fact, it will be more convenient to speak about the spectrum of the operator $\mathcal{L} = -\Delta_\mu$ that is called the *positive definite Laplace operator* (for the reason that will be made clear below).

Given a linear operator A in a vector space $\mathcal{V}$, a vector $v \in \mathcal{V} \setminus \{0\}$ is called an *eigenvector* of A if $Av = \lambda v$ for some scalar λ; the latter is called an eigenvalue of A. The set of all (complex) values eigenvalues is called the *spectrum* of A and is denoted by $\operatorname{spec} A$.

In the case when the underlying vector space is the space $\mathcal{F}$ of functions on the graph V, the eigenvectors are also referred to as *eigenfunctions*. Let us give some examples of explicit calculation of the eigenvalues of the operator $\mathcal{L} = -\Delta_\mu$ on finite graphs with simple weight μ. Recall that for a simple weight we have

$$\mathcal{L}f(x) = f(x) - \frac{1}{\deg(x)} \sum_{y \sim x} f(y).$$

Example. 1. For graph $\mathbb{Z}_2$ we have

$$\begin{aligned}\mathcal{L}f(0) &= f(0) - f(1) \\ \mathcal{L}f(1) &= f(1) - f(0)\end{aligned}$$

so that the equation $\mathcal{L}f = \lambda f$ becomes

$$\begin{aligned}(1-\lambda) f(0) &= f(1) \\ (1-\lambda) f(1) &= f(0)\end{aligned}$$

whence $(1-\lambda)^2 f(k) = f(k)$ for both $k = 0, 1$. Since $f \not\equiv 0$, we obtain the equation $(1-\lambda)^2 = 1$ whence we find two eigenvalues $\lambda = 0$ and $\lambda_1 = 2$. Alternatively, considering a function f as a column-vector $\binom{f(0)}{f(1)}$, we can represent the action of $\mathcal{L}$ as a matrix multiplication:

$$\begin{pmatrix}\mathcal{L}f(0) \\ \mathcal{L}f(1)\end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} f(0) \\ f(1) \end{pmatrix},$$

so that the eigenvalues of $\mathcal{L}$ coincide with those of the matrix $\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$. Its characteristic equation is $(1-\lambda)^2 - 1 = 0$, whence we obtain again the same two eigenvalues $\lambda = 0$ and $\lambda = 2$.

2. For $\mathbb{Z}_3$ we have then

$$\mathcal{L}f(x) = f(x) - \frac{1}{2}(f(x-1) + f(x+1))$$

The action of $\mathcal{L}$ can be written as a matrix multiplication:

$$\begin{pmatrix} \mathcal{L}f(0) \\ \mathcal{L}f(1) \\ \mathcal{L}f(2) \end{pmatrix} = \begin{pmatrix} 1 & -1/2 & -1/2 \\ -1/2 & 1 & -1/2 \\ -1/2 & -1/2 & 1 \end{pmatrix} \begin{pmatrix} f(0) \\ f(1) \\ f(2) \end{pmatrix}.$$

The characteristic polynomial of the above 3×3 matrix is $-\left(\lambda^3 - 3\lambda^2 + \frac{9}{4}\lambda\right)$. Evaluating its roots, we obtain the following eigenvalues of $\mathcal{L}$: $\lambda = 0$ (simple) and $\lambda = 3/2$ with multiplicity 2.

3. For the path graph P_3 with vertices $\{0, 1, 2\}$ and edges $0 \sim 1 \sim 2$ we have

$$\begin{aligned} \mathcal{L}f(0) &= f(0) - f(1) \\ \mathcal{L}f(1) &= f(1) - \frac{1}{2}(f(0) + f(2)) \\ \mathcal{L}f(2) &= f(2) - f(1) \end{aligned}$$

so that the matrix of $\mathcal{L}$ is

$$\left(\begin{array}{ccc} 1 & -1 & 0 \\ -1/2 & 1 & -1/2 \\ 0 & -1 & 1 \end{array}\right).$$

The characteristic polynomial is $-\left(\lambda^3 - 3\lambda^2 + 2\lambda\right)$, and the eigenvalues are $\lambda = 0$, $\lambda = 1$, and $\lambda = 2$.

Coming back to the general theory, assume now that $\mathcal{V}$ is an inner product space, that is, an inner product (u, v) is defined for all $u, v \in \mathcal{V}$, that is a bilinear, symmetric, positive definite function on $\mathcal{V} \times \mathcal{V}$. Assume that the operator A is *symmetric* (or self-adjoint) with respect to this inner product, that is, $(Au, v) = (u, Av)$ for all $u, v \in \mathcal{V}$. It is known from Linear Algebra that all the eigenvalues of A are real. In particular, the eigenvalues can be enumerated in increasing order as $\lambda_1 \leq ... \leq \lambda_N$ where $N = \dim \mathcal{V}$ and each eigenvalue is counted with multiplicity. Furthermore, there is an orthonormal basis $\{v_k\}_{k=1}^N$ in $\mathcal{V}$ such that each v_k is an eigenvector of A with the eigenvalue λ_k, that is $Av_k = \lambda_k v_k$ (equivalently, the matrix of A in the basis $\{v_k\}$ is $\operatorname{diag}(\lambda_1, ..., \lambda_N)$).

Set $\mathcal{R}(v) = \frac{(Av,v)}{(v,v)}$. The function $\mathcal{R}(v)$, that is defined on $\mathcal{V} \setminus \{0\}$, is called the *Rayleigh quotient* of A. The following identities are true for all $k = 1, ..., N$:

$$\lambda_k = \mathcal{R}(v_k) = \inf_{v \perp v_1, ..., v_{k-1}} \mathcal{R}(v) = \sup_{v \perp v_N, v_{N-1}, ..., v_{k+1}} \mathcal{R}(v)$$

(where $v \perp u$ means that u and v are orthogonal, that is, $(v, u) = 0$). In particular,

$$\lambda_1 = \inf_{v \neq 0} \mathcal{R}(v) \quad \text{and} \quad \lambda_N = \sup_{v \neq 0} \mathcal{R}(v).$$

We will apply these results to the Laplace $\mathcal{L}$ in the vector space $\mathcal{F}$ of real-valued functions on V. Consider in $\mathcal{F}$ the following inner product: for any two functions $f, g \in \mathcal{F}$, set

$$(f, g) := \sum_{x \in V} f(x) g(x) \mu(x),$$

which can be considered as the integration of fg against measure μ on V.

Lemma 2.2 *The operator $\mathcal{L}$ is symmetric with respect to the above inner product, that is,*

$$(\mathcal{L} f, g) = (f, \mathcal{L} g)$$

for all $f, g \in \mathcal{F}$.

Proof. Indeed, by the Green formula (2.4), we have

$$(\mathcal{L}f, g) = -(\Delta_\mu f, g) = -\sum_{x \in V} \Delta_\mu f(x)\, g(x)\, \mu(x) = \frac{1}{2} \sum_{x,y \in V} (\nabla_{xy} f)(\nabla_{xy} g)\, \mu_{xy},$$

and the last expression is symmetric in f, g so that it is equal also to $(\mathcal{L}g, f)$.

Alternatively, since $\mathcal{L} = \mathrm{id} - P$, it suffices to prove that P is symmetric, which follows from

$$\begin{aligned} (Pf, g) &= \sum_x Pf(x)\, g(x)\, \mu(x) = \sum_x \sum_y P(x,y)\, f(y)\, g(x)\, \mu(x) \\ &= \sum_x \sum_y P(y,x)\, f(y)\, g(x)\, \mu(y) = (Pg, f) \end{aligned}$$

where we have used the reversibility of P. ∎

By the Green formula, the Rayleigh quotient of $\mathcal{L}$ is

$$\mathcal{R}(f) = \frac{(\mathcal{L}f, f)}{(f, f)} = \frac{1}{2} \frac{\sum_{x,y \in V} (\nabla_{xy} f)(\nabla_{xy} g)\, \mu_{xy}}{\sum_{x \in V} f^2(x)\, \mu(x)}.$$

To state the next theorem about the spectrum of $\mathcal{L}$, we need a notion of a bipartite graph.

Definition. A graph (V, E) is called *bipartite* if V admits a partition into two non-empty disjoint subsets V_1, V_2 such that $x, y \in V_i \Rightarrow x \not\sim y$.

In terms of coloring, one can say that a graph is bipartite if its vertices can be colored by two colors, so that the vertices of the same color are not connected by an edge.

Example. Here are some examples of bipartite graphs.

1. A complete bipartite graph $K_{n,m}$ is bipartite.

2. The graphs $\mathbb{Z}_m$ and P_m is bipartite provided m is even.

3. Product of bipartite graphs is bipartite. In particular, $\mathbb{Z}_m^n$ and P_m^n are bipartite provided m is even.

Theorem 2.3 *For any finite, connected, weighted graph* (V, μ) *with* $N = \#V > 1$, *the following is true.*

(a) *Zero is a simple eigenvalue of* $\mathcal{L}$.

(b) *All the eigenvalues of* $\mathcal{L}$ *are contained in* $[0,2]$.

(c) *If* (V,μ) *is not bipartite then all the eigenvalues of* $\mathcal{L}$ *are in* $[0,2)$.

Proof. (a) Since $\mathcal{L}1=0$, the constant function is an eigenfunction with the eigenvalue 0. Assume now that f is an eigenfunction of the eigenvalue 0 and prove that $f\equiv$ const, which will imply that 0 is a simple eigenvalue. If $\mathcal{L}f=0$ then it follows from (2.4) with $g=f$ that

$$\sum_{\{x,y\in V: x\sim y\}}\left(f(y)-f(x)\right)^2\mu_{xy}=0.$$

In particular, $f(x)=f(y)$ for any two neighboring vertices x,y. The connectedness of the graph means that any two vertices $x,y\in V$ can be connected to each other by a path $\{x_k\}_{k=0}^m$ where

$$x=x_0\sim x_1\sim \ldots \sim x_m=y$$

whence it follows that $f(x_0)=f(x_1)=\ldots=f(x_m)$ and $f(x)=f(y)$. Since this is true for all couples $x,y\in V$, we obtain $f\equiv$ const.

(b) Let λ be an eigenvalue of $\mathcal{L}$ with an eigenfunction f. Using $\mathcal{L}f = \lambda f$ and the Green formula (2.4), we obtain

$$\begin{aligned}\lambda \sum_{x \in V} f^2(x)\, \mu(x) &= \sum_{x \in V} \mathcal{L}f(x)\, f(x)\, \mu(x) \\ &= \frac{1}{2} \sum_{\{x,y \in V : x \sim y\}} \left(f(y) - f(x)\right)^2 \mu_{xy}. \qquad (2.5)\end{aligned}$$

It follows from (2.5) that $\lambda \geq 0$. Using $(a+b)^2 \leq 2\left(a^2+b^2\right)$, we obtain

$$\begin{aligned}\lambda \sum_{x \in V} f^2(x)\, \mu(x) &\leq \sum_{\{x,y \in V : x \sim y\}} \left(f(y)^2 + f(x)^2\right) \mu_{xy} \\ &= \sum_{x,y \in V} f(y)^2 \mu_{xy} + \sum_{x,y \in V} f(x)^2 \mu_{xy} \\ &= \sum_{y \in V} f(y)^2 \mu(y) + \sum_{x \in V} f(x)^2 \mu(x) \\ &= 2 \sum_{x \in V} f(x)^2 \mu(x). \qquad (2.6)\end{aligned}$$

It follows from (2.6) that $\lambda \leq 2$.

Alternatively, one can first prove that $\|P\| \leq 1$, which follows from $\sum_y P(x,y) = 1$ and which implies $\operatorname{spec} P \subset [-1,1]$, and then conclude that $\operatorname{spec} \mathcal{L} = 1 - \operatorname{spec} P \subset [0,2]$.

(c) We need to prove that $\lambda = 2$ is not an eigenvalue. Assume from the contrary that $\lambda = 2$ is an eigenvalue with an eigenfunction f, and prove that (V,μ) is bipartite. Since $\lambda = 2$, all the inequalities in the above calculation (2.6) must become equalities. In particular, we must have for all $x \sim y$ that

$$(f(x) - f(y))^2 = 2\left(f(x)^2 + f(y)^2\right),$$

which is equivalent to

$$f(x) + f(y) = 0.$$

If $f(x_0) = 0$ for some x_0 then it follows that $f(x) = 0$ for all neighbors of x_0. Since the graph is connected, we obtain that $f(x) \equiv 0$, which is not possible for an eigenfunction. Hence, $f(x) \neq 0$ for all $x \in \Gamma$. Then V splits into a disjoint union of two sets:

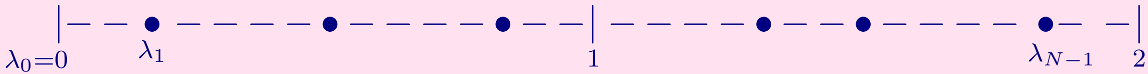
$$V^+ = \{x \in V : f(x) > 0\} \text{ and } V^- = \{x \in V : f(x) < 0\}.$$

The above argument shows that if $x \in V^+$ then all neighbors of x are in V^-, and vice versa. Hence, (V, μ) is bipartite, which finishes the proof. $\blacksquare$

Hence, we can enumerate all the eigenvalues of $\mathcal{L}$ in the increasing order as follows:

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_{N-1}.$$

Note that the smallest eigenvalue is denoted by λ_0 rather than by λ_1. Also, we have always $\lambda_{N-1} \leq 2$ and the latter inequality is strict if the graph is non-bipartite. Below is a diagram of the interval $[0, 2]$ with marked eigenvalues:



Example. As an example of application of Theorem 2.3, let us investigate the solvability of the equation $\mathcal{L}u = f$. Since by the Green formula

$$\sum_x (\mathcal{L}u)(x)\,\mu(x) = 0,$$

a necessary condition for solvability is

$$\sum_x f(x)\,\mu(x) = 0. \tag{2.7}$$

Assuming that, let us show that the equation $\mathcal{L}u = f$ has a solution. Indeed, condition (2.7) means that $f \perp 1$. Consider the subspace $\mathcal{F}_0$ of $\mathcal{F}$ that consists of all functions orthogonal to 1. Since 1 is the eigenfunction of $\mathcal{L}$ with eigenvalue $\lambda_0 = 0$, the space $\mathcal{F}_0$ is invariant for the operator $\mathcal{L}$, and the spectrum of $\mathcal{L}$ in $\mathcal{F}_0$ is $\lambda_1, ...\lambda_{N-1}$. Since all $\lambda_j > 0$, we see that $\mathcal{L}$ is invertible in $\mathcal{F}_0$, that is, the equation $\mathcal{L}u = f$ has for any $f \in \mathcal{F}_0$ a unique solution $u \in \mathcal{F}_0$ given by $u = \mathcal{L}^{-1}f$.

3 Convergence to equilibrium

The next theorem is one of the main results of this Section. We use the notation

$$\|f\| = \sqrt{(f, f)}.$$

Theorem 3.1 *Let (V, μ) be a finite, connected, weighted graph with $N = \#V > 1$. and P be its Markov operator. For any function $f \in \mathcal{F}$, set*

$$\overline{f} = \frac{1}{\mu(V)} \sum_{x \in V} f(x) \mu(x).$$

Then, for any positive integer n, we have

$$\left\| P^n f - \overline{f} \right\| \leq \rho^n \|f\| \tag{3.8}$$

where

$$\rho = \max(|1 - \lambda_1|, |1 - \lambda_{N-1}|). \tag{3.9}$$

Consequently, if the graph (V, μ) is non-bipartite then

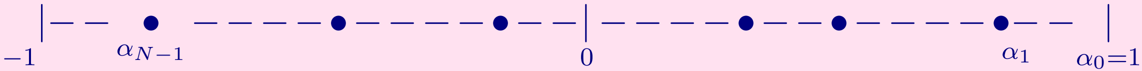
$$\left\| P^n f - \overline{f} \right\| \to 0 \text{ as } n \to \infty, \tag{3.10}$$

that is, $P^n f$ converges to a constant $\overline{f}$ as $n \to \infty$.

The estimate (3.8) gives the rate of convergence of $P^n f$ to the constant $\overline{f}$: it is decreasing exponentially in n provided $\rho < 1$. The constant

ρ is called the *spectral radius* of the Markov operator $P = \mathrm{id} - \mathcal{L}$. Note that the eigenvalues of P are $\alpha_k = 1 - \lambda_k$ where λ_k are the eigenvalues of $\mathcal{L}$. Hence, we have

$$-1 \leq \alpha_{N-1} \leq ... \leq \alpha_1 < \alpha_0 = 1$$



and $\rho = \max\left(|\alpha_1|, |\alpha_{N-1}|\right)$, so that all the eigenvalues of P except for α_0 are contained in $[-\rho, \rho]$.

Proof of Theorem 3.1. If the graph (V, μ) is non-bipartite then by Theorem 2.3 we have $\lambda_{N-1} < 2$ whence

$$-1 < \alpha_{N-1} \leq \alpha_1 < 1,$$

which implies that $\rho < 1$. Therefore, $\rho^n \to 0$ as $n \to \infty$ and (3.8) implies (3.10).

To prove (3.8), choose an orthonormal basis $\{v_k\}_{k=0}^{N-1}$ of the eigenfunctions of P so that $Pv_k = \alpha_k v_k$ and, hence,

$$Pv_k = \alpha_k v_k.$$

Any function $f \in \mathcal{F}$ can be expanded in the basis v_k as follows:

$$f = \sum_{k=0}^{N-1} c_k v_k$$

where $c_k = (f, v_k)$. By the Parseval identity, we have

$$\|f\|^2 = \sum_{k=0}^{N-1} c_k^2.$$

We have

$$Pf = \sum_{k=0}^{N-1} c_k P v_k = \sum_{k=0}^{N-1} \alpha_k c_k v_k,$$

whence, by induction in n,

$$P^n f = \sum_{k=0}^{N-1} \alpha_k c_k v_k.$$

On the other hand, recall that $v_0 \equiv c$ for some constant c. It can be determined from the normalization condition $\|v_0\| = 1$, that is,

$$\sum_{x \in V} c^2 \mu(x) = 1$$

whence $c = \frac{1}{\sqrt{\mu(V)}}$. It follows that

$$c_0 = (f, v_0) = \frac{1}{\sqrt{\mu(V)}} \sum_{x \in V} f(x) \mu(x)$$

and

$$c_0 v_0 = \frac{1}{\mu(V)} \sum_{x \in V} f(x) \mu(x) = \overline{f}.$$

Hence, we obtain

$$
\begin{aligned}
P^n f - \overline{f} &= \sum_{k=0}^{N-1} \alpha_k^n c_k v_k - c_0 v_0 \\
&= \alpha_0^n c_0 v_0 + \sum_{k=1}^{N-1} \alpha_k^n c_k v_k - c_0 v_0 \\
&= \sum_{k=1}^{N-1} \alpha_k^n c_k v_k .
\end{aligned}
$$

By the Parseval identity, we have

$$
\begin{aligned}
\left\| P^n f - \overline{f} \right\|^2 &= \sum_{k=1}^{N-1} \alpha_k^{2n} c_k^2 \\
&\leq \max_{1 \leq k \leq N-1} |\alpha_k|^{2n} \sum_{k=1}^{N-1} c_k^2 .
\end{aligned}
$$

As was already explained before the proof, all the eigenvalues α_k of P with $k \geq 1$ are contained in the interval $[-\rho, \rho]$, so that $|\alpha_k| \leq \rho$. Observing also that

$$\sum_{k=1}^{N-1} c_k^2 \leq \|f\|^2 ,$$

we obtain

$$\left\|P^n f - \overline{f}\right\|^2 \leq \rho^{2n} \|f\|^2 ,$$

which finishes the proof. ■

Let us show that if $\lambda_{N-1} = 2$ then (3.10) is not the case (as we will see later, for bipartite graphs one has exactly $\lambda_{N-1} = 2$). Indeed, if f is an eigenfunction of $\mathcal{L}$ with the eigenvalue 2 then f is the eigenfunction of P with the eigenvalue -1, that is, $Pf = -f$. Then we obtain that $P^n f = (-1)^n f$ so that $P^n f$ does not converge to any function as $n \to \infty$.

Corollary 3.2 *Let (V, μ) be a finite, connected, weighted graph that is non-bipartite, and let $\{X_n\}$ be the associated random walk. Fix a vertex $x_0 \in V$ and consider the distribution function of X_n:*

$$v_n (x) = \mathbb{P}_{x_0} (X_n = x) .$$

Then

$$v_n(x) \to \frac{\mu(x)}{\mu(V)} \text{ as } n \to \infty, \tag{3.11}$$

where $\mu(V) = \sum_{x \in V} \mu(x)$. *Moreover, we have*

$$\sum_{x \in V} \left(v_n(x) - \frac{\mu(x)}{\mu(V)} \right)^2 \frac{\mu(x_0)}{\mu(x)} \leq \rho^{2n}. \tag{3.12}$$

It follows from (3.12) that, for any $x \in V$,

$$\left| v_n(x) - \frac{\mu(x)}{\mu(V)} \right| \leq \rho^n \sqrt{\frac{\mu(x)}{\mu(x_0)}} \tag{3.13}$$

Proof. Recall that the operator P was defined by

$$Pf(x) = \sum_y P(x, y) f(y).$$

It follows by induction in n that

$$P^n f(x) = \sum_{y \in V} P_n(x, y) f(y) \tag{3.14}$$

where $P_n(x,y)$ is the n-step transition function. Since the graph is not bipartite, we have $\rho \in (0,1)$, so that (3.11) follows from (3.12) (or from (3.13)). To prove (3.12), consider also the backward distribution function

$$u_n(x) = \mathbb{P}_x(X_n = x_0) = \frac{v_n(x)\,\mu(x_0)}{\mu(x)}$$

(cf. (1.11)). Since

$$u_n(x) = P_n(x, x_0) = \sum_{y \in V} P_n(x,y)\,\mathbf{1}_{\{x_0\}}(y) = P^n \mathbf{1}_{\{x_0\}}(x),$$

we obtain by Theorem 3.1 with $f = \mathbf{1}_{\{x_0\}}$ that

$$\left\| u_n - \overline{f} \right\|^2 \leq \rho^{2n} \left\| \mathbf{1}_{\{x_0\}} \right\|^2 .$$

Since for this function f

$$\overline{f} = \frac{1}{\mu(V)}\mu(x_0) \quad \text{and} \quad \|f\|^2 = \mu(x_0),$$

we obtain that

$$\sum_x \left(\frac{v_n(x)\,\mu(x_0)}{\mu(x)} - \frac{\mu(x_0)}{\mu(V)} \right)^2 \mu(x) \leq \rho^{2n} \mu(x_0)$$

whence (3.11) follows. ■

A random walk is called *ergodic* if (3.11) is satisfied. We have seen that a random walk on a finite, connected, non-bipartite graph is ergodic. Given a small number $\varepsilon > 0$, define the *mixing time* $T = T(\varepsilon)$ by the condition $\rho^T = \varepsilon$, that is

$$\boxed{T = \frac{\ln \frac{1}{\varepsilon}}{\ln \frac{1}{\rho}}.} \tag{3.15}$$

Then, for any $n \geq T$, we obtain from (3.13) that

$$\left| v_n(x) - \frac{\mu(x)}{\mu(V)} \right| \leq \varepsilon \sqrt{\frac{\mu(x)}{\mu(x_0)}}.$$

To ensure a good approximation $v_n(x) \approx \frac{\mu(x)}{\mu(V)}$, the value of ε should be chosen so that

$$\varepsilon \sqrt{\frac{\mu(x)}{\mu(x_0)}} << \frac{\mu(x)}{\mu(V)},$$

which is equivalent to

$$\boxed{\varepsilon << \min_x \frac{\mu(x)}{\mu(V)}}.$$

In many examples of graphs, λ_1 is close to 0 and λ_{N-1} is close to 2. In this case, we have

$$\ln \frac{1}{|\alpha_1|} = \ln \frac{1}{1-\lambda_1} \approx \lambda_1$$

and

$$\ln \frac{1}{|\alpha_{N-1}|} = \ln \frac{1}{|1-\lambda_{N-1}|} = \ln \frac{1}{1-(2-\lambda_{N-1})} \approx 2-\lambda_{N-1},$$

whence

$$T = \frac{\ln \frac{1}{\varepsilon}}{\ln \frac{1}{\rho}} \approx \ln \frac{1}{\varepsilon} \max \left(\frac{1}{\lambda_1}, \frac{1}{2-\lambda_{N-1}} \right). \tag{3.16}$$

In the next sections, we will estimate the eigenvalues on specific graphs and, consequently, provide some explicit values for the mixing time.

4 Eigenvalues of bipartite graphs

The next statement contains an additional information about the spectrum of $\mathcal{L}$ in some specific cases.

Theorem 4.1 *Let (V, μ) be a finite, connected, weighted graph with $N := \#V > 1$. If (V, μ) is a bipartite and λ is an eigenvalue of $\mathcal{L}$ then $2 - \lambda$ is also an eigenvalue of $\mathcal{L}$, with the same multiplicity as λ. In particular, 2 is a simple eigenvalue of $\mathcal{L}$.*

Hence, we see that a graph is bipartite if and only if $\lambda_{N-1} = 2$.

Proof. Since the eigenvalues α of the Markov operator are related to the eigenvalues λ of $\mathcal{L}$ by $\alpha = 1 - \lambda$, the claim is equivalent to the following: if α is an eigenvalue of P then $-\alpha$ is also an eigenvalue of P with the same multiplicity (indeed, $\alpha = 1 - \lambda$ implies $-\alpha = 1 - (2 - \lambda)$). Let V^+, V^- be a partition of V such that $x \sim y$ only if x and y belong to same of the subset V^+, V^-. Given an eigenfunction f of P with the eigenvalue α, consider

$$g(x) = \begin{cases} f(x), & x \in V^+ \\ -f(x), & x \in V^- \end{cases} . \tag{4.17}$$

Let us show that g is an eigenfunction of P with the eigenvalue $-\alpha$. For all $x \in V^+$, we have

$$\begin{aligned} Pg(x) &= \sum_{y \in V} P(x,y)\,g(y) = \sum_{y \in V^-} P(x,y)\,g(y) \\ &= -\sum_{y \in V^-} P(x,y)\,f(y) \\ &= -Pf(x) = -\alpha f(x) = -\alpha g(x), \end{aligned}$$

and for $x \in V^-$ we obtain in the same way

$$\begin{aligned} Pg(x) &= \sum_{y \in V^+} P(x,y)\,g(y) \\ &= \sum_{y \in V^+} P(x,y)\,f(y) = Pf(x) = \alpha f(x) = -\alpha g(x). \end{aligned}$$

Hence, $-\alpha$ is an eigenvalue of P with the eigenfunction g.

Let m be the multiplicity of α as an eigenvalue of P, and m' be the multiplicity of $-\alpha$. Let us prove that $m' = m$. There exist m linearly independent eigenfunctions $f_1, ..., f_m$ of the eigenvalue α. Using (4.17), we construct m eigenfunctions $g_1, ..., g_m$ of the eigenvalue $-\alpha$, that are obviously linearly independent, whence we conclude that $m' \geq m$. Since $-(-\alpha) = \alpha$, applying the same argument to the eigenvalue $-\alpha$ instead of α, we obtain the opposite inequality $m \geq m'$, whence $m = m'$.

Finally, since 0 is a simple eigenvalue of $\mathcal{L}$, it follows that 2 is also a simple eigenvalue of $\mathcal{L}$. It follows from the proof that the eigenfunction $g(x)$ with the eigenvalue 2 is as follows: $g(x) = c$ on V^+ and $g(x) = -c$ on V^-, for any non-zero constant c. $\blacksquare$

Theorem 4.1 implies the analogs of Theorem 3.1 and Corollary 3.2 for bipartite graphs, that we state without proof.

Theorem 4.2 *Let (V, μ) be a finite connected weighted graph. Assume that (V, μ) is bipartite, and let V^+, V^- be a bipartition of V. For any function f on V, consider the function $\widetilde{f}$ on V that takes two values as follows:*

$$\widetilde{f}(x) = \frac{2}{\mu(V)} \begin{cases} \sum_{y \in V^+} f(y)\,\mu(y), & x \in V^+, \\ \sum_{y \in V^-} f(y)\,\mu(y), & x \in V^-. \end{cases}$$

Then, for all even n,

$$\left\| P^n f - \widetilde{f} \right\| \leq \rho^n \|f\|$$

where

$$\rho = \max\left(\left|1-\lambda_1\right|, \left|\lambda_{N-2}-1\right|\right).$$

Consequently, for all $x \in V$, we have $P^n f(x) \to \widetilde{f}(x)$ as $n \to \infty$, n is even.

Note that $0 \le \rho < 1$ because the eigenvalues $\lambda_0 = 0$ and $\lambda_{N-1} = 2$ are simple and, hence, $0 < \lambda_1 \le \lambda_{N-2} < 2$.

Corollary 4.3 *Under the hypotheses of* Theorem 4.2, *consider the forward distribution $v_n(x) = \mathbb{P}_{x_0}(X_n = x)$ of the random walk on (V, μ) and the function*

$$\widetilde{v}(x) := \begin{cases} \frac{2\mu(x)}{\mu(V)}, & x \not\sim x_0, \\ 0, & x \sim x_0. \end{cases}$$

Then, for all $x \in V$ and even n,

$$\left|v_n(x) - \widetilde{v}(x)\right| \le \rho^n \sqrt{\frac{\mu(x)}{\mu(x_0)}}.$$

Consequently, for all $x \in V$, we have $v_n(x) \to \widetilde{v}(x)$ as $n \to \infty$, n is even.

It follows from Theorem 4.1 $\lambda_1 \leq 1$ and $\lambda_{N-2} = 2 - \lambda_1$ so that in fact $\rho = 1 - \lambda_1$. It follows that the mixing time (assuming that n is even) is estimated by

$$T = \frac{\ln \frac{1}{\varepsilon}}{\ln \frac{1}{\rho}} \approx \frac{\ln \frac{1}{\varepsilon}}{\lambda_1}$$

assuming $\lambda_1 \approx 0$. Here ε must be chosen so that $\varepsilon << \min_x \frac{\mu(x)}{\mu(V)}$.

5 Eigenvalues of $\mathbb{Z}_m$

Let us compute the eigenvalues of the Markov operator on the cycle graph $\mathbb{Z}_m$ with simple weight. Recall that $\mathbb{Z}_m = \{0, 1, ..., m-1\}$ and the connections are

$$0 \sim 1 \sim 2 \sim ... \sim m-1 \sim 0.$$

The Markov operator is given by

$$Pf(k) = \frac{1}{2}\left(f(k+1) + f(k-1)\right)$$

where k is regarded as a residue $\bmod m$. The eigenfunction equation $Pf = \alpha f$ becomes

$$f(k+1) - 2\alpha f(k) + f(k-1) = 0. \tag{5.18}$$

We know already that $\alpha = 1$ is always a simple eigenvalue of P, and $\alpha = -1$ is a (simple) eigenvalue if and only if $\mathbb{Z}_m$ is bipartite, that is, if m is even. Assume in what follows that $\alpha \in (-1, 1)$.

Consider first the *difference equation* (5.18) on $\mathbb{Z}$, that is, for all $k \in \mathbb{Z}$, and find all solutions f as functions on $\mathbb{Z}$. Observe first that the set of all solutions of (5.18) is a linear space (the sum of two solutions is a solution, and a multiple of a solution is a solution), and the dimension of this space is 2, because function f is uniquely determined by (5.18) and by two initial conditions $f(0) = a$ and $f(1) = b$. Therefore, to find all solutions of (5.18), it suffices to find two linearly independent solutions.

Let us search specific solution of (5.18) in the form $f(k) = r^k$ where the number r is to be found. Substituting into (5.18) and cancelling by r^k, we obtain the equation for r:

$$r^2 - 2\alpha r + 1 = 0.$$

It has two complex roots

$$r = \alpha \pm i\sqrt{1-\alpha^2} = e^{\pm i\theta},$$

where $\theta \in (0, \pi)$ is determined by the condition

$$\cos\theta = \alpha \ \ (\text{and} \ \sin\theta = \sqrt{1-\alpha^2}).$$

Hence, we obtain two independent complex-valued solutions of (5.18)

$$f_1(k) = e^{ik\theta} \ \text{ and } \ f_2(k) = e^{-ik\theta}.$$

Taking their linear combinations and using the Euler formula, we arrive at the following real-valued independent solutions:

$$f_1(k) = \cos k\theta \ \text{ and } \ f_2(k) = \sin k\theta. \tag{5.19}$$

In order to be able to consider a function $f(k)$ on $\mathbb{Z}$ as a function on $\mathbb{Z}_m$, it must be m-periodic, that is,

$$f(k+m) = f(k) \ \text{ for all } k \in \mathbb{Z}.$$

The functions (5.19) are m-periodic provided $m\theta$ is a multiple of 2π, that is,

$$\theta = \frac{2\pi l}{m},$$

for some integer l. The restriction $\theta \in (0, \pi)$ is equivalent to

$$l \in (0, m/2) .$$

Hence, for each l from this range we obtain an eigenvalue $\alpha = \cos\theta$ of multiplicity 2 (with eigenfunctions $\cos k\theta$ and $\sin k\theta$).

Let us summarize this result in the following statement.

Lemma 5.1 *The eigenvalues of the Markov operator P on the graph $\mathbb{Z}_m$ are as follows:*

1. *If m is odd then the eigenvalues are $\alpha = 1$ (simple) and $\alpha = \cos \frac{2\pi l}{m}$ for all $l = 1, ..., \frac{m-1}{2}$ (double);*

2. *if m is even then the eigenvalues are $\alpha = \pm 1$ (simple) and $\alpha = \cos \frac{2\pi l}{m}$ for all $l = 1, ..., \frac{m}{2} - 1$ (double).*

Note that the sum of the multiplicities of all the listed above eigenvalues is m so that we have found indeed all the eigenvalues of P.

For example, in the case $m = 3$ we obtain the Markov eigenvalues $\alpha = 1$ and $\alpha = \cos\frac{2\pi}{3} = -\frac{1}{2}$ (double). The eigenvalues of $\mathcal{L}$ are as follows: $\lambda = 0$ and $\lambda = 3/2$ (double). If $m = 4$ then the Markov eigenvalues are $\alpha = \pm 1$ and $\alpha = \cos\frac{2\pi}{4} = 0$ (double). The eigenvalues of $\mathcal{L}$ are as follows: $\lambda = 0$, $\lambda = 1$ (double), $\lambda = 2$.

Alternatively, one can list all the eigenvalues of P with multiplicities in the following sequence:

$$\left\{\cos\frac{2\pi j}{m}\right\}_{j=0}^{m-1}.$$

Indeed, if m is odd then $j = 0$ gives $\alpha = 1$, and for $j = 1, ..., m-1$ we have

$$\cos\frac{2\pi j}{m} = \cos\frac{2\pi l}{m} \text{ where } l = \begin{cases} j, & 1 \leq j \leq \frac{m-1}{2}, \\ m - j, & \frac{m+1}{2} \leq j \leq m-1, \end{cases}$$

because

$$\cos\frac{2\pi j}{m} = \cos\left(2\pi - \frac{2\pi j}{m}\right) = \cos\frac{2\pi l}{m}.$$

Hence, every value of $\left\{\cos\frac{2\pi l}{m}\right\}_{l=1}^{\frac{m-1}{2}}$ occurs in the sequence $\left\{\cos\frac{2\pi j}{m}\right\}_{j=1}^{m-1}$ exactly twice.

In the same way, if m is even, then $j = 0$ and $j = m/2$ give the values 1 and -1, respectively, while for $j \in [1, m-1] \setminus \{m/2\}$ we have

$$\cos \frac{2\pi j}{m} = \cos \frac{2\pi l}{m} \text{ where } l = \begin{cases} j, & 1 \le j \le \frac{m}{2} - 1, \\ m - j, & \frac{m}{2} + 1 \le j \le m - 1, \end{cases}$$

so that each value of $\left\{\cos \frac{2\pi l}{m}\right\}_{l=1}^{\frac{m}{2}-1}$ is counted exactly twice.

6 Products of weighted graphs

Definition. Let (X, a) and (Y, b) be two finite weighted graphs. Fix two numbers $p, q > 0$ and define the product graph

$$(V, \mu) = (X, a) \,\Box_{p,q}\, (Y, b)$$

as follows: $V = X \times Y$ and the weight μ on V is defined by

$$\begin{aligned} \mu_{(x,y),(x',y)} &= pb\,(y)\, a_{xx'} \\ \mu_{(x,y),(x,y')} &= qa\,(x)\, b_{yy'} \end{aligned}$$

and $\mu_{(x,y),(x',y')} = 0$ otherwise. The numbers p, q are called the *parameters* of the product.

Clearly, the product weight $\mu_{(x,y),(x',y')}$ is symmetric as it should be. If the associated graphs are (X, E_1), (Y, E_2) and (V, E) then comparing the above definition with the definition (1.1) of the product of the graphs, that is shown on the diagram

$$
\begin{array}{cccccc}
y' \bullet & \dots & \overset{(x,y')}{\bullet} & - & \overset{(x',y')}{\bullet} & \dots \\
| & & | & & | & \\
y \bullet & \dots & \overset{(x,y)}{\bullet} & - & \overset{(x',y)}{\bullet} & \dots \\
\vdots & & \vdots & & \vdots & \\
{}^{Y}\!\diagup_{X} & \dots & x \bullet & - & x' \bullet & \dots
\end{array}
$$

we see that $(V, E) = (X, E_1) \square (Y, E_2)$.

The weight on the vertices of V is given by

$$
\begin{aligned}
\mu(x, y) &= \sum_{x', y'} \mu_{(x,y),(x',y')} = p \sum_{x'} a_{xx'} b(y) + q \sum_{y'} a(x) b_{yy'} \\
&= (p + q) a(x) b(y).
\end{aligned}
$$

Claim. *If A and B are the Markov kernels on X and Y, then the Markov kernel P on the product (V, μ) is given by*

$$P\left((x,y),(x',y')\right) = \begin{cases} \frac{p}{p+q} A(x,x'), & \text{if } y = y', \\ \frac{q}{p+q} B(y,y'), & \text{if } x = x', \\ 0, & \text{otherwise.} \end{cases} \tag{6.20}$$

Proof. Indeed, we have in the case $y = y'$

$$\begin{aligned} P\left((x,y),(x',y')\right) &= \frac{\mu_{(x,y),(x',y')}}{\mu(x,y)} = \frac{p a_{xx'} b(y)}{(p+q)\, a(x)\, b(y)} \\ &= \frac{p}{p+q}\frac{a_{xx'}}{a(x)} = \frac{p}{p+q} A(x,x'), \end{aligned}$$

and the case $x = x'$ is treated similarly. ∎

For the random walk on (V, μ), the identity (6.20) means the following: the random walk at (x,y) chooses first between the directions X and Y with probabilities $\frac{p}{p+q}$ and $\frac{q}{p+q}$, respectively, and then chooses a vertex in the chosen direction accordingly to the Markov kernel there.

In particular, if a and b are simple weights, then we obtain

$$\begin{aligned}
\mu_{(x,y),(x',y)} &= p \deg(y) \quad \text{if } x \sim x' \\
\mu_{(x,y),(x,y')} &= q \deg(x) \quad \text{if } \ y \sim y'
\end{aligned}$$

and $\mu_{(x,y),(x',y')} = 0$ otherwise.

If in addition the graphs A and B are regular, that is, $\deg(x) = \text{const} =: \deg(A)$ and $\deg(y) = \text{const} =: \deg(B)$ then the most natural choice of the parameter p and q is as follows

$$p = \frac{1}{\deg(B)} \text{ and } \quad q = \frac{1}{\deg(A)},$$

so that the weight μ is also simple. We obtain the following statement.

Lemma 6.1 *If (X, a) and (Y, b) are regular graphs with simple weights, then the product*

$$(X, a) \square_{\frac{1}{\deg(B)}, \frac{1}{\deg(A)}} (Y, b)$$

is again a regular graph with a simple weight.

Note that the degree of the product graph is $\deg(A) + \deg(B)$.

Example. Consider the graphs $\mathbb{Z}_m^n$ and $\mathbb{Z}_m^k$ with simple weights. Since their degrees are equal to $2n$ and $2k$, respectively, we obtain

$$\mathbb{Z}_m^n \square_{\frac{1}{2k},\frac{1}{2n}} \mathbb{Z}_m^k = \mathbb{Z}_m^{n+k}.$$

Theorem 6.2 *Let (X, a) and (Y, b) be finite weighted graphs without isolated vertices, and let $\{\alpha_k\}_{k=0}^{n-1}$ and $\{\beta_l\}_{l=0}^{m-1}$ be the sequences of the eigenvalues of the Markov operators A and B respectively, counted with multiplicities. Then all the eigenvalues of the Markov operator P on the product $(V, \mu) = (X, a) \square_{p,q} (Y, b)$ are given by the sequence $\left\{\frac{p\alpha_k + q\beta_l}{p+q}\right\}$ where $k = 0, ..., n-1$ and $l = 0, ..., m-1$.*

In other words, the eigenvalues of P are the convex combinations of eigenvalues of A and B, with the coefficients $\frac{p}{p+q}$ and $\frac{q}{p+q}$. Note that the same relation holds for the eigenvalues of the Laplace operators: since those on (X, a) and (Y, b) are $1 - \alpha_k$ and $1 - \beta_l$, respectively, we see that the eigenvalues of the Laplace operator on (V, μ) are given by

$$1 - \frac{p\alpha_k + q\beta_l}{p+q} = \frac{p\,(1-\alpha_k) + q\,(1-\beta_l)}{p+q},$$

that is, the same convex combination of $1 - \alpha_k$ and $1 - \beta_l$.

Proof. Let f be an eigenfunction of A with the eigenvalue α and g be the eigenfunction of B with the eigenvalue β. Let us show that the function $h\,(x, y) = f\,(x)\,g\,(y)$ is the eigenvalue of P with the eigenvalue

$\frac{p\alpha+q\beta}{p+q}$. We have

$$\begin{aligned}
Ph\left(x,y\right) &= \sum_{x',y'} P\left(\left(x,y\right),\left(x',y'\right)\right) h\left(x',y'\right) \\
&= \sum_{x'} P\left(\left(x,y\right),\left(x',y\right)\right) h\left(x',y\right) + \sum_{y'} P\left(\left(x,y\right),\left(x,y'\right)\right) h\left(x,y'\right) \\
&= \frac{p}{p+q}\sum_{x'} A\left(x,x'\right) f\left(x'\right) g\left(y\right) + \frac{q}{p+q}\sum_{y'} B\left(y,y'\right) f\left(x\right) g\left(y'\right) \\
&= \frac{p}{p+q} Af\left(x\right) g\left(y\right) + \frac{q}{p+q} f\left(x\right) Bg\left(y\right) \\
&= \frac{p}{p+q} \alpha f\left(x\right) g\left(y\right) + \frac{q}{p+q} \beta f\left(x\right) g\left(y\right) \\
&= \frac{p\alpha+q\beta}{p+q} h\left(x,y\right),
\end{aligned}$$

which was to be proved.

Let $\{f_k\}$ be a basis in the space of functions on X such that $Af_k = \alpha_k f_k$, and $\{g_l\}$ be a basis in the space of functions on Y, such that $Bg_l = \beta_l g_l$. Then $h_{kl}\left(x,y\right) = f_k\left(x\right) g_l\left(y\right)$ is a linearly independent sequence of functions on $V = X \times Y$. Since the number of such functions is

$nm = \#V$, we see that $\{h_{kl}\}$ is a basis in the space of functions on V. Since h_{kl} is the eigenfunction with the eigenvalue $\frac{p\alpha_k+q\beta_l}{p+q}$, we conclude that the sequence $\frac{p\alpha_k+q\beta_l}{p+q}$ exhausts all the eigenvalues of P. $\blacksquare$

Corollary 6.3 *Let (V,E) be a finite connected regular graph with $N>1$ vertices, and set $(V^n,E_n)=(V,E)^{\square n}$. Let μ be a simple weight on V, and $\{\alpha_k\}_{k=0}^{N-1}$ be the sequence of the eigenvalues of the Markov operator on (V,μ), counted with multiplicity. Let μ_n be a simple weight on V^n. Then the eigenvalues of the Markov operator on (V^n,μ_n) are given by the sequence*

$$\left\{\frac{\alpha_{k_1}+\alpha_{k_2}+...+\alpha_{k_n}}{n}\right\} \tag{6.21}$$

for all $k_i\in\{0,1,...,N-1\}$, where each eigenvalue is counted with multiplicity.

It follows that if $\{\lambda_k\}_{k=0}^{N-1}$ is the sequence of the eigenvalues of the Laplace operator on (V,μ) then the eigenvalues of Laplace operator on (V^n,μ_n) are given by the sequence

$$\left\{\frac{\lambda_{k_1}+\lambda_{k_2}+...+\lambda_{k_n}}{n}\right\}. \tag{6.22}$$

Proof. Induction in n. If $n = 1$ then there is nothing to prove. Let us make the inductive step from n to $n+1$. Let degree of (V, E) be D, then $\deg(V^n) = nD$. Note that.

$$\left(V^{n+1}, E_{n+1}\right) = (V^n, E_n) \square (V, E)$$

It follows from Lemma 6.1 that

$$\left(V^{n+1}, \mu_{n+1}\right) = (V^n, \mu_n) \square_{\frac{1}{D}, \frac{1}{nD}} (V, \mu) .$$

By the inductive hypothesis, the eigenvalues of the Laplacian on (V^n, μ_n) are given by the sequence (6.21). Hence, by Theorem 6.2, the eigenvalues on $\left(V^{n+1}, \mu_{n+1}\right)$ are given by

$$\begin{aligned} & \frac{1/D}{1/D + 1/(nD)} \frac{\alpha_{k_1} + \alpha_{k_2} + \ldots + \alpha_{k_n}}{n} + \frac{1/(nD)}{1/D + 1/(nD)} \alpha_k \\ = \ & \frac{n}{n+1} \frac{\alpha_{k_1} + \alpha_{k_2} + \ldots + \alpha_{k_n}}{n} + \frac{1}{n+1} \alpha_k \\ = \ & \frac{\alpha_{k_1} + \alpha_{k_2} + \ldots + \alpha_{k_n} + \alpha_k}{n+1}, \end{aligned}$$

which was to be proved. ■

7 Eigenvalues and mixing times in $\mathbb{Z}_m^n$

Consider $\mathbb{Z}_m^n$ with an odd m so that the graph $\mathbb{Z}_m^n$ is not bipartite. By Corollary 3.2, the distribution of the random walk on $\mathbb{Z}_m^n$ converges to the equilibrium measure $\frac{\mu(x)}{\mu(V)} = \frac{1}{N}$, where $N = \#V = m^n$, and the rate of convergence is determined by the spectral radius ρ

$$\rho = \max\left(|\alpha_{\min}|, |\alpha_{\max}|\right) \tag{7.23}$$

where $\alpha_{\min}$ is the minimal eigenvalue of P and $\alpha_{\max}$ is the second maximal eigenvalue of P. In the case $n = 1$, all the eigenvalues of P except for 1 are listed in the sequence (without multiplicity):

$$\left\{\cos\frac{2\pi l}{m}\right\},\ 1 \leq l \leq \frac{m-1}{2}.$$

This sequence is obviously decreasing in l, and its maximal and minimal values are

$$\cos\frac{2\pi}{m} \text{ and } \cos\left(\frac{2\pi}{m}\frac{m-1}{2}\right) = -\cos\frac{\pi}{m},$$

respectively. For a general n, by Corollary 6.3, the eigenvalue of P have the form

$$\frac{\alpha_{k_1} + \alpha_{k_2} + ... + \alpha_{k_n}}{n} \tag{7.24}$$

where α_{k_i} are the eigenvalues of P for $n = 1$. The minimal value of (7.24) is equal to the minimal value of α_k, that is

$$\alpha_{\min} = -\cos\frac{\pi}{m}.$$

The maximal value of (7.24) is, of course, 1 when all $\alpha_{k_i} = 1$, and the second maximal value is obtained when one of α_{k_i} is equal to $\cos\frac{2\pi}{m}$ and the rest $n-1$ values are 1. Hence, we have

$$\alpha_{\max} = \frac{n - 1 + \cos\frac{2\pi}{m}}{n} = 1 - \frac{1 - \cos\frac{2\pi}{m}}{n}. \tag{7.25}$$

For example, if $m = 3$ then $\alpha_{\min} = -\cos\frac{\pi}{3} = -\frac{1}{2}$ and

$$\alpha_{\max} = 1 - \frac{1 - \cos\frac{2\pi}{3}}{n} = 1 - \frac{3}{2n}$$

whence

$$\rho = \max\left(\frac{1}{2}, \left|1 - \frac{3}{2n}\right|\right) = \begin{cases} \frac{1}{2}, & n \leq 3, \\ 1 - \frac{3}{2n}, & n \geq 4. \end{cases} \tag{7.26}$$

If m is large then, using the approximation $\cos\theta \approx 1 - \frac{\theta^2}{2}$ for small θ, we obtain

$$\alpha_{\min} \approx -\left(1 - \frac{\pi^2}{2m^2}\right) \quad \text{and} \quad \alpha_{\max} \approx 1 - \frac{2\pi^2}{nm^2}.$$

Using $\ln\frac{1}{1-\theta} \approx \theta$, we obtain

$$\ln\frac{1}{|\alpha_{\min}|} \approx \frac{\pi^2}{2m^2} \quad \text{and} \quad \ln\frac{1}{|\alpha_{\max}|} \approx \frac{2\pi^2}{nm^2}.$$

Finally, by (3.15) and (7.23), we obtain the estimate of the mixing time in $\mathbb{Z}_m^n$ with error ε:

$$T = \frac{\ln\frac{1}{\varepsilon}}{\ln\frac{1}{\rho}} \approx \ln\frac{1}{\varepsilon}\max\left(\frac{2m^2}{\pi^2}, \frac{nm^2}{2\pi^2}\right) = \frac{\ln\frac{1}{\varepsilon}}{2\pi^2}nm^2, \tag{7.27}$$

assuming for simplicity that $n \geq 4$. Choosing $\varepsilon = \frac{1}{N^2}$ (note that ε must be $\ll \frac{1}{N}$) and using $N = m^n$, we obtain

$$T \approx \frac{2\ln m^n}{2\pi^2} nm^2 = \frac{1}{\pi^2} n^2 m^2 \ln m = \frac{1}{\pi^2} \frac{m^2}{\ln m} (\ln N)^2 . \tag{7.28}$$

Example. In $\mathbb{Z}_5^{10}$ with $N = 5^{10} \approx 10^6$, the mixing time is $T \approx 400$, which is relatively short given the fact that the number of vertices in this graph is In fact, the actual mixing time is even smaller than above value of T since the latter is an upper bound for the mixing time.

Example. Take $m = 3$. Using the exact value of ρ given by (7.26), and obtain that, for $\mathbb{Z}_3^n$ with large n and $\varepsilon = \frac{1}{N^2} = 3^{-2n}$:

$$T = \frac{\ln \frac{1}{\varepsilon}}{\ln \left(1 - \frac{3}{2n}\right)} \approx \frac{2\ln N}{3/(2n)} = \frac{4}{3\ln 3} (\ln N)^2 \approx 1.2 (\ln N)^2 .$$

For example, for $\mathbb{Z}_3^{13}$ with $N = 3^{13} > 10^6$ vertices, we obtain a very short mixing time $T \approx 250$.

Example. For $n = 1$ (7.27) slightly changes to become

$$T \approx \ln \frac{1}{\varepsilon} \frac{2m^2}{\pi^2} = \frac{4}{\pi^2} m^2 \ln m,$$

where we have chosen $\varepsilon = \frac{1}{m^2}$. For example, in $\mathbb{Z}_{10^6}$ with $N = m = 10^6$ we have $T \simeq 10^{12}$, which is huge in striking contrast with the previous graph $\mathbb{Z}_5^{10}$ with (almost) the same number of vertices as $\mathbb{Z}_{10^6}$!

Example. Let us estimate the mixing time on the binary cube $\mathbb{Z}_2^n$ that is a bipartite graph. The eigenvalues of the Laplace operator on $\mathbb{Z}_2$ are $\lambda_0 = 0$ and $\lambda_1 = 2$. Then the eigenvalues of the Laplace operator on $\mathbb{Z}_2^n$ are given by (6.22), that is, by

$$\left\{ \frac{\lambda_{k_1} + \lambda_{k_2} + \ldots + \lambda_{k_n}}{n} \right\}$$

where each $k_i = 0$ or 1. Hence, each eigenvalue of $\mathbb{Z}_2^n$ is equal to $\frac{2j}{n}$ where $j \in [0, n]$ is the number of 1's in the sequence $k_1, \ldots, k_n$. The multiplicity of the eigenvalue $\frac{2j}{n}$ is equal to the number of binary sequences $\{k_1, \ldots, k_n\}$ where 1 occurs exactly j times. This number is given by the binomial coefficient $\binom{n}{j}$. Hence, all the eigenvalues of the Laplace operator on $\mathbb{Z}_2^n$ are $\left\{\frac{2j}{n}\right\}$ where $j = 0, 1, \ldots, n$, and the multiplicity of this eigenvalue is $\binom{n}{j}$.

The convergence rate of the random walk on $\mathbb{Z}_2^n$ is determined by $\lambda_1 = \frac{2}{n}$. Assume that n is large and let $N = 2^n$. Taking $\varepsilon = \frac{1}{N^2} = 2^{-2n}$, we obtain the following estimate of the mixing time:

$$T \approx \frac{\ln \frac{1}{\varepsilon}}{\lambda_1} = \frac{2n \ln 2}{2/n} = (\ln 2)\, n^2 = \frac{(\ln N)^2}{\ln 2} \approx 1.4\, (\ln N)^2 \,.$$

For example, for $\{0,1\}^{20}$ with $N = 2^{20} \approx 10^6$ vertices we obtain $T \approx 280$.

8 Expansion rate

Let (V, μ) be a finite connected weighted graph with $N > 1$ vertices. For any two non-empty subsets $X, Y \subset V$, set

$$d(X, Y) = \min_{x \in X, y \in Y} d(x, y)$$

Note that $d(X, Y) \geq 0$ and $d(X, Y) = 0$ if and only if X and Y have non-empty intersection.

We will need also another quantity:

$$l(X, Y) = \frac{1}{2} \ln \frac{\mu(X^c)\,\mu(Y^c)}{\mu(X)\,\mu(Y)},$$

that will be applied for disjoint sets X, Y. Then $X \subset Y^c$ and $Y \subset X^c$ whence it follows that $l(X, Y) \geq 0$. Furthermore, $l(X, Y) = 0$ if and only if $X = Y^c$. To understand better $l(X, Y)$, express it in terms of the set $Z = V \setminus (X \cup Y)$ so that

$$l(X, Y) = \frac{1}{2} \ln \left(1 + \frac{\mu(Z)}{\mu(X)}\right)\left(1 + \frac{\mu(Z)}{\mu(Y)}\right).$$

Hence, the quantity $l(X,Y)$ measures the space between X and Y in terms of the measure of the set Z.

Let the eigenvalues of the Laplace operator $\mathcal{L}$ on (V,μ) be

$$0=\lambda_0<\lambda_1\leq ...\leq\lambda_{N-1}\leq 2.$$

We will use the following notation:

$$\delta=\frac{\lambda_{N-1}-\lambda_1}{\lambda_{N-1}+\lambda_1}. \tag{8.1}$$

Clearly, $\delta\in[0,1)$ and

$$\frac{\lambda_1}{\lambda_{N-1}}=\frac{1-\delta}{1+\delta}.$$

The relation to the spectral radius

$$\rho=\max\left(|1-\lambda_1|,|\lambda_{N-1}-1|\right)$$

is as follows: since $\lambda_1\geq 1-\rho$ and $\lambda_{N-1}\leq 1+\rho$, we obtain that

$$\frac{\lambda_1}{\lambda_{N-1}}\geq\frac{1-\rho}{1+\rho}$$

whence $\delta\leq\rho$.

Theorem 8.1 *For any two disjoint sets $X, Y \subset V$, we have*

$$d(X,Y) \leq 1 + \frac{l(X,Y)}{\ln \frac{1}{\delta}}. \tag{8.2}$$

If $\delta = 0$ then we set by definition $\frac{l(X,Y)}{\ln \frac{1}{\delta}} = 0$. Before the proof, let us discuss consequences and examples.

Example. If $D = d(X,Y) > 1$ then the estimate (8.2) implies that

$$\frac{1}{\delta} \leq \exp\left(\frac{l(X,Y)}{D-1}\right) = \left(\frac{\mu(X^c)\,\mu(Y^c)}{\mu(X)\,\mu(Y)}\right)^{\frac{1}{2(D-1)}} =: A$$

whence $\delta \geq A^{-1}$. In terms of the eigenvalues we obtain

$$\frac{\lambda_1}{\lambda_{N-1}} \leq \frac{A-1}{A+1}.$$

Since $\lambda_{N-1} \leq 2$, this yields an upper bound for λ_1

$$\lambda_1 \leq 2\frac{A-1}{A+1}.$$

Example. Let us show that

$$\operatorname{diam}(V) \leq 1 + \frac{1}{\ln \frac{1}{\delta}} \ln \frac{\mu(V)}{m}, \tag{8.3}$$

where $m = \min_{x \in V} \mu(x)$. Indeed, set in (8.2) $X = \{x\}$, $Y = \{y\}$ where x, y are two distinct vertices. Then

$$l(X, Y) \leq \frac{1}{2} \ln \frac{\mu(V)^2}{\mu(x)\,\mu(y)} \leq \ln \frac{\mu(V)}{m}$$

whence

$$d(x, y) \leq 1 + \frac{1}{\ln \frac{1}{\delta}} \ln \frac{\mu(V)}{m}.$$

Taking in the left hand side the supremum in all $x, y \in V$, we obtain (8.3). In a particular case of a simple weight, we have $m = \min_x \deg(x)$, $\mu(V) = 2\#E$, and

$$\operatorname{diam}(V) \leq 1 + \frac{1}{\ln \frac{1}{\delta}} \ln \frac{2\#E}{m}.$$

For any subset $X \subset V$, denote by $U_r(X)$ the r-neighborhood of X, that is,

$$U_r(X) = \{y \in V : d(y, X) \leq r\}.$$

Corollary 8.2 *For any non-empty set $X \subset V$ and any integer $r \geq 1$, we have*

$$\mu(U_r(X)) \geq \frac{\mu(V)}{1 + \frac{\mu(X^c)}{\mu(X)}\delta^{2r}}. \tag{8.4}$$

Proof. Indeed, take $Y = V \setminus U_r(X)$ so that $U_r(X) = Y^c$. Then $d(X, Y) = r + 1$, and (8.2) yields

$$r \leq \frac{1}{2} \frac{1}{\ln \frac{1}{\delta}} \ln \frac{\mu(X^c)\,\mu(Y^c)}{\mu(X)\,\mu(Y)},$$

whence

$$\frac{\mu(Y^c)}{\mu(Y)} \geq \left(\frac{1}{\delta}\right)^{2r} \frac{\mu(X)}{\mu(X^c)}.$$

Since $\mu(Y) = \mu(V) - \mu(Y^c)$, we obtain

$$\frac{\mu(V) - \mu(Y^c)}{\mu(Y^c)} \leq \delta^{2r} \frac{\mu(X^c)}{\mu(X)}$$

whence (8.4) follows $\blacksquare$

Example. Given a set X, define the *expansion rate* of X to be the minimal positive integer R such that

$$\mu(U_R(X)) \geq \frac{1}{2}\mu(V).$$

Imagine a communication network as a graph where the vertices are the communication centers (like computer servers) and the edges are direct links between the centers. If X is a set of selected centers, then it is reasonable to ask, how many steps from X are required to reach the majority (at least 50%) of all centers? This is exactly the expansion rate of X, and the networks with short expansion rate provide better connectivity.

The inequality (8.4) implies an upper bound for the expansion rate.

Indeed, if

$$\frac{\mu\left(X^{c}\right)}{\mu\left(X\right)}\delta^{2r}\leq 1, \tag{8.5}$$

then (8.4) implies that $\mu\left(U_{r}\left(X\right)\right)\geq\frac{1}{2}\mu\left(V\right)$. The condition (8.5) is equivalent to

$$r\geq\frac{1}{2}\frac{\ln\frac{\mu(X^{c})}{\mu(X)}}{\ln\frac{1}{\delta}}, \tag{8.6}$$

from where we see that the expansion rate R of X satisfies

$$R\leq\frac{1}{2}\frac{\ln\frac{\mu(X^{c})}{\mu(X)}}{\ln\frac{1}{\delta}}.$$

Hence, a good communication network should have the number δ as small as possible (which is similar to the requirement that ρ should be as small as possible for a fast convergence rates to the equilibrium). For many large practical networks, one has the following estimate for the spectral radius: $\rho\simeq\frac{1}{\ln N}$ (recall that λ_{1} and λ_{N-1} are contained in the interval $[1-\rho,1+\rho]$). Since $\delta\leq\rho$, it follows that

$$\frac{1}{\delta}\gtrsim\ln N.$$

Assuming that X consists of a single vertex and that $\frac{\mu(X^c)}{\mu(X)} \approx N$, we obtain the following estimate of the expansion rate of a single vertex:

$$R \lesssim \frac{\ln N}{\ln \ln N}.$$

For example, if $N = 10^8$, which is a typical figure for the internet graph, then $R \simeq 6$ (although neglecting the constant factors). This very fast expansion rate is called "a small world" phenomenon, and it is actually observed in large communication networks.

The same phenomenon occurs in the *coauthor network*: two mathematicians are connected by an edge if they have a joint publication. Although the number of recorded mathematicians is quite high ($\approx 10^5$), a few links are normally enough to get from one mathematician to a substantial portion of the entire network.

Note for comparison that in $\mathbb{Z}_m^n$ we have $1 - \delta \simeq \frac{c}{nm^2}$ whence $R \lesssim \frac{m^2}{\ln m} (\ln N)^2$.

Proof of Theorem 8.1. As before, denote by $\mathcal{F}$ the space of all real-valued functions on V. Let $w_0, w_1, ..., w_{N-1}$ be an orthonormal basis in $\mathcal{F}$ that consists of the eigenfunctions of $\mathcal{L}$, and let their eigenvalues be $\lambda_0 = 0, \lambda_1, ..., \lambda_{N-1}$. Any function $u \in \mathcal{F}$ admits an expansion in the basis $\{w_l\}$ as follows:

$$u = \sum_{l=0}^{N-1} a_l w_l \tag{8.7}$$

with some coefficients a_l. We know already that

$$a_0 w_0 = \overline{u} = \frac{1}{\mu(V)} \sum_{x \in V} u(x) \mu(x)$$

(see the proof of Theorem 3.1). Denote

$$u' = u - \overline{u} = \sum_{l=1}^{N-1} a_l w_l$$

so that $u = \overline{u} + u'$ and $u' \perp \overline{u}$.

Let $\Phi(\lambda)$ be a polynomial with real coefficient. We have

$$\Phi(\mathcal{L}) u = \sum_{l=0}^{N-1} a_l \Phi(\lambda_l) w_l = \Phi(0) \overline{u} + \sum_{l=1}^{N-1} a_l \Phi(\lambda_l) w_l.$$

If v is another function from $\mathcal{F}$ with expansion

$$v = \sum_{l=0}^{N-1} b_l w_l = \overline{v} + \sum_{l=1}^{N-1} b_l w_l = \overline{v} + v',$$

then

$$\begin{aligned} (\Phi(\mathcal{L}) u, v) &= (\Phi(0) \overline{u, v}) + \sum_{l=1}^{N-1} a_l b_l \Phi(\lambda_l) \\ &\geq \Phi(0) \overline{u} \overline{v} \mu(V) - \max_{1 \leq l \leq N-1} |\Phi(\lambda_l)| \sum_{l=1}^{N-1} |a_l| |b_l| \\ &\geq \Phi(0) \overline{u} \overline{v} \mu(V) - \max_{1 \leq l \leq N-1} |\Phi(\lambda_l)| \|u'\| \|v'\|. \end{aligned} \tag{8.8}$$

Assume now that $\operatorname{supp} u \subset X$, $\operatorname{supp} v \subset Y$ and that

$$D = d(X, Y) \geq 2$$

(if $D \leq 1$ then (8.2) is trivially satisfied). Let $\Phi(\lambda)$ be a polynomial of λ of degree $D - 1$. Let us verify that

$$(\Phi(\mathcal{L}) u, v) = 0. \tag{8.9}$$

Indeed, the function $\mathcal{L}^k u$ is supported in the k-neighborhood of $\operatorname{supp} u$, whence it follows that $\Phi(\mathcal{L})\,u$ is supported in the $(D-1)$-neighborhood of X. Since $U_{D-1}(X)$ is disjoint with Y, we obtain (8.9). Comparing (8.9) and (8.8), we obtain

$$\max_{1\leq l\leq N-1}|\Phi(\lambda_l)| \geq \Phi(0)\,\frac{\overline{u}\overline{v}\mu(V)}{\|u'\|\,\|v'\|}. \tag{8.10}$$

Let us take now $u=\mathbf{1}_X$ and $v=\mathbf{1}_Y$. We have

$$\overline{u}=\frac{\mu(X)}{\mu(V)},\qquad \|\overline{u}\|^2=\frac{\mu(X)^2}{\mu(V)},\qquad \|u\|^2=\mu(X),$$

whence

$$\|u'\|=\sqrt{\|u\|^2-\|\overline{u}\|^2}=\sqrt{\mu(X)-\frac{\mu(X)^2}{\mu(V)}}=\sqrt{\frac{\mu(X)\,\mu(X^c)}{\mu(V)}}.$$

Using similar identities for v and substituting into (8.10), we obtain

$$\max_{1\leq l\leq N-1}|\Phi(\lambda_l)| \geq \Phi(0)\sqrt{\frac{\mu(X)\,\mu(Y)}{\mu(X^c)\,\mu(Y^c)}}.$$

Finally, let us specify $\Phi(\lambda)$ as follows:

$$\Phi(\lambda) = \left(\frac{\lambda_1 + \lambda_{N-1}}{2} - \lambda\right)^{D-1}.$$

Since $\max|\Phi(\lambda)|$ on the set $\lambda \in [\lambda_1, \lambda_{N-1}]$ is attained at $\lambda = \lambda_1$ and $\lambda = \lambda_{N-1}$ and

$$\max_{[\lambda_1, \lambda_{N-1}]} |\Phi(\lambda)| = \left(\frac{\lambda_{N-1} - \lambda_1}{2}\right)^{D-1},$$

it follows from (8.18) that

$$\left(\frac{\lambda_{N-1} - \lambda_1}{2}\right)^{D-1} \geq \left(\frac{\lambda_{N-1} + \lambda_1}{2}\right)^{D-1} \sqrt{\frac{\mu(X)\,\mu(Y)}{\mu(X^c)\,\mu(Y^c)}}.$$

Rewriting this inequality in the form

$$\exp(l(X,Y)) \geq \left(\frac{1}{\delta}\right)^{D-1}$$

and taking ln, we obtain (8.2). ■

The next result is a generalization of Theorem 8.1. Set

$$\delta_k = \frac{\lambda_{N-1} - \lambda_k}{\lambda_{N-1} + \lambda_k}.$$

Theorem 8.3 *Let $k \in \{1, 2, ..., N-1\}$ and $X_1, ..., X_{k+1}$ be non-empty disjoint subsets of V, and set*

$$D = \min_{i \neq j} d\left(X_i, X_j\right).$$

Then

$$D \leq 1 + \frac{1}{\ln \frac{1}{\delta_k}} \max_{i \neq j} l\left(X_i, X_j\right). \tag{8.11}$$

Example. Define the k-diameter of the graph by

$$\operatorname{diam}_k\left(V\right) = \max_{\{x_1,...,x_{k+1}\}} \min_{i \neq j} d\left(x_i, x_j\right).$$

In particular, diam_1 is the usual diameter of the graph. Applying Theorem 8.3 to sets $X_i = \{x_i\}$ and maximizing over all x_i, we obtain

$$\mathrm{diam}_k(V) \leq 1 + \frac{\ln \frac{\mu(V)}{m}}{\ln \frac{1}{\delta_k}}, \tag{8.12}$$

where $m = \inf_{x \in V} \mu(x)$.

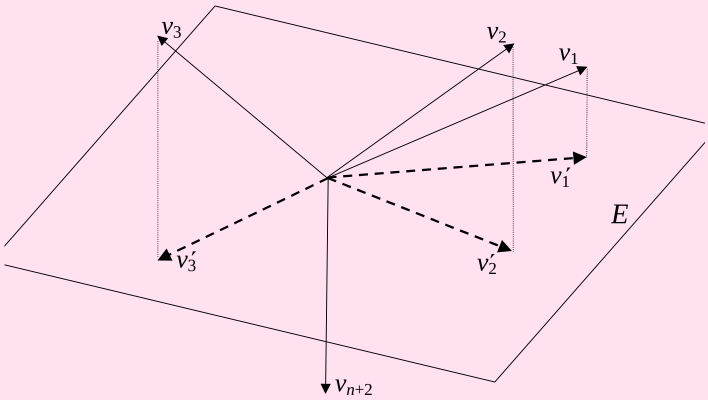
We precede the proof of Theorem 8.3 by a lemma.

Lemma 8.4 *In any sequence of $n+2$ vectors in n-dimensional Euclidean space, there are two vectors with non-negative inner product.*

Note that $n+2$ is the smallest number for which the statement of Lemma 8.4 is true. Indeed, if $e_1, e_2, ..., e_n$ denote an orthonormal basis in the given space, let us set $v := -e_1 - e_2 - ... - e_n$. Then any two of the following $n+1$ vectors

$$e_1 + \varepsilon v,\ e_2 + \varepsilon v,\ ....,\ e_n + \varepsilon v,\ v$$

have a negative inner product, provided $\varepsilon > 0$ is small enough.



Proof. Induction in n. The inductive basis for $n = 1$ is obvious. The inductive step from $n-1$ to n is shown on the diagram. assume that there are $n+2$ vectors v_1, v_2, ..., v_{n+2} such that $(v_i, v_j) < 0$ for all distinct i, j. Denote by E the orthogonal complement of v_{n+2}, and by v_i' the orthogonal projection of v_i onto E.

The difference $v_i' - v_i$ is orthogonal to E and, hence, colinear to v_{n+2} so that

$$v_i = v_i' - \varepsilon_i v_{n+2} \tag{8.13}$$

for some constant $\varepsilon_i > 0$. It follows, for all distinct $i, j = 1, ..., n+1$,

$$(v_i, v_j) = \left(v_i', v_j'\right) + \varepsilon_i \varepsilon_j \left(v_{n+2}, v_{n+2}\right). \tag{8.14}$$

By the inductive hypothesis, in a sequence of $n+1$ vectors $v_1', ..., v_{n+1}'$ in $(n-1)$-dimensional Euclidean space E, there are two vectors with non-negative inner product, say, $\left(v_i', v_j'\right) \geq 0$. It follows from (8.14) that also $(v_i, v_j) \geq 0$, which finishes the proof. ■

Proof of Theorem 8.3. We use the same notation as in the proof of Theorem 8.1. If $D \leq 1$ then (8.11) is trivially satisfied. Assume in the sequel that $D \geq 2$. Let u_i be a function on V with $\operatorname{supp} u_i \subset X_i$. Let $\Phi(\lambda)$ be a polynomial with real coefficients of degree $\leq D-1$, which is non-negative for $\lambda \in \{\lambda_1, ..., \lambda_{k-1}\}$. Let us prove the following estimate

$$\max_{k \leq l \leq N-1} |\Phi(\lambda_l)| \geq \Phi(0) \min_{i \neq j} \frac{\overline{u}_i \overline{u}_j \mu(V)}{\|u_i'\| \|u_j'\|}. \tag{8.15}$$

Expand a function u_i in the basis $\{w_l\}$ as follows:

$$u_i = \sum_{l=0}^{N-1} a_l^{(i)} w_l = \overline{u_i} + \sum_{l=1}^{k-1} a_l^{(i)} w_l + \sum_{l=k}^{N-1} a_l^{(i)} w_l.$$

It follows that

$$
\begin{aligned}
\left(\Phi\left(\mathcal{L}\right)u_i,u_j\right) &= \Phi\left(0\right)\overline{u_iu_j}\mu\left(V\right)+\sum_{l=1}^{k-1}\Phi\left(\lambda_l\right)a_l^{(i)}a_l^{(j)}+\sum_{l=k}^{N-1}\Phi\left(\lambda_l\right)a_l^{(i)}a_l^{(j)} \\
&\geq \Phi\left(0\right)\overline{u_iu_j}\mu\left(V\right)+\sum_{l=1}^{k-1}\Phi\left(\lambda_l\right)a_l^{(i)}a_l^{(j)} \\
&\quad -\max_{k\leq l\leq N-1}\left|\Phi\left(\lambda_l\right)\right|\left\|u_i'\right\|\left\|u_j'\right\|.
\end{aligned}
$$

Since also $\left(\Phi\left(\mathcal{L}\right)u_i,u_j\right)=0$, it follows that

$$
\max_{k\leq l\leq N-1}\left|\Phi\left(\lambda_l\right)\right|\left\|u_i'\right\|\left\|u_j'\right\|\geq\Phi\left(0\right)\overline{u_iu_j}\mu\left(V\right)+\sum_{l=1}^{k-1}\Phi\left(\lambda_l\right)a_l^{(i)}a_l^{(j)}. \tag{8.16}
$$

In order to be able to obtain (8.15), we would like to have

$$
\sum_{l=1}^{k-1}\Phi\left(\lambda_l\right)a_l^{(i)}a_l^{(j)}\geq 0. \tag{8.17}
$$

In general, (8.17) cannot be guaranteed for any couple i,j but we claim that there exists a couple i,j of distinct indices such that (8.17) holds (this is the reason why in (8.15) we have min in all i, j). To prove that, consider the inner product in $\mathbb{R}^{k-1}$ given by[1]

$$(a, b) = \sum_{i=1}^{k-1} \Phi(\lambda_i) a_i b_i,$$

for any two vectors $a = (a_1, ..., a_{k-1})$ and $b = (b_1, ..., b_{k-1})$. Also, consider the vectors $a^{(i)} = \left(a_1^{(i)}, ..., a_{k-1}^{(i)}\right)$ for $i = 1, ..., k+1$. Hence, we have $k+1$ vectors in $(k-1)$-dimensional Euclidean space. By Lemma 8.4, there are two vectors, say $a^{(i)}$ and $a^{(j)}$ such that $\left(a^{(i)}, a^{(j)}\right) \geq 0$, which exactly means (8.17). For these i, j, we obtain from (8.16)

$$\max_{k \leq l \leq N-1} |\Phi(\lambda_l)| \geq \Phi(0) \frac{\overline{u_i u_j} \mu(V)}{\|u_i'\| \left\|u_j'\right\|}$$

whence (8.15) follows.

[1] By hypothesis, we have $\Phi(\lambda_i) \geq 0$. If $\Phi(\lambda_i)$ vanishes for some i then use only those i for which $\Phi(\lambda_i) > 0$ and consider inner product in a space of smaller dimension.

In particular, taking $u_i = \mathbf{1}_{X_i}$ and using that

$$\overline{u}_i = \frac{\mu(X_i)}{\mu(V)}$$

and

$$\|u_i'\| = \sqrt{\frac{\mu(X_i)\,\mu(X_i^c)}{\mu(V)}},$$

we obtain

$$\max_{k\le l\le N-1} |\Phi(\lambda_l)| \ge \Phi(0) \min_{i\ne j} \sqrt{\frac{\mu(X_i)\,\mu(X_j)}{\mu(X_i^c)\,\mu\left(X_j^c\right)}}. \tag{8.18}$$

Consider the following polynomial of degree $D-1$

$$\Phi(\lambda) = \left(\frac{\lambda_k + \lambda_{N-1}}{2} - \lambda\right)^{D-1},$$

which is clearly non-negative for $\lambda \le \lambda_k$. Since $\max|\Phi(\lambda)|$ on the set $\lambda \in [\lambda_k, \lambda_{N-1}]$ is attained at $\lambda = \lambda_k$ and $\lambda = \lambda_{N-1}$ and

$$\max_{\lambda\in[\lambda_k,\lambda_{N-1}]} |\Phi(\lambda)| = \left(\frac{\lambda_{N-1} - \lambda_k}{2}\right)^{D-1},$$

it follows from (8.18) that

$$\left(\frac{\lambda_{N-1}-\lambda_k}{2}\right)^{D-1} \geq \left(\frac{\lambda_{N-1}+\lambda_k}{2}\right)^{D-1} \min_{i\neq j} \sqrt{\frac{\mu\left(X_i\right)\mu\left(X_j\right)}{\mu\left(X_i^c\right)\mu\left(X_j^c\right)}},$$

whence (8.11) follows. ■

9 Cheeger's inequality

Let (V, μ) be a weighted graph with the edges set E. Recall that, for any vertex subset $\Omega \subset V$, its measure $\mu(\Omega)$ is defined by

$$\mu(\Omega) = \sum_{x \in \Omega} \mu(x).$$

Similarly, for any edge subset $S \subset E$, define its measure $\mu(S)$ by

$$\mu(S) = \sum_{\xi \in S} \mu_\xi,$$

where $\mu_\xi := \mu_{xy}$ for any edge $\xi = \overline{xy}$.

For any set $\Omega \subset V$, define its *edge boundary* $\partial\Omega$ by

$$\partial\Omega = \{\overline{xy} \in E : x \in \Omega, y \notin \Omega\}.$$

Definition. Given a finite weighted graph (V, μ), define its *Cheeger constant* by

$$h = h(V, \mu) = \inf_{\substack{\Omega \subset V \\ \mu(\Omega) \leq \frac{1}{2}\mu(V)}} \frac{\mu(\partial\Omega)}{\mu(\Omega)}. \tag{9.19}$$

In other words, h is the largest constant such that the following inequality is true

$$\mu(\partial\Omega) \geq h\mu(\Omega) \tag{9.20}$$

for any subset Ω of V with measure $\mu(\Omega) \leq \frac{1}{2}\mu(V)$.

Lemma 9.1 *We have $\lambda_1 \leq 2h$.*

Proof. Let Ω be a set at which the infimum in (9.19) is attained. Consider the following function

$$f(x) = \begin{cases} 1, & x \in \Omega \\ -a, & x \notin \Omega^c \end{cases}$$

where a is chosen so that $f \perp 1$, that is, $\mu(\Omega) = a\mu(\Omega^c)$ whence

$$a = \frac{\mu(\Omega)}{\mu(\Omega^c)} \leq 1.$$

Then $\lambda_1 \leq \mathcal{R}(f) := \frac{(\mathcal{L}f,f)}{(f,f)}$ so that it suffices to prove that $\mathcal{R}(f) \leq 2h$. We have

$$(f,f) = \sum_{x \in V} f(x)^2 \mu(x) = \mu(\Omega) + a^2\mu(\Omega^c) = (1+a)\mu(\Omega)$$

and

$$\begin{aligned}
(\mathcal{L}f, f) &= \frac{1}{2}\sum_{x,y}(f(x) - f(y))^2 \mu_{xy} \\
&= \sum_{x\in\Omega, y\in\Omega^c}(f(x) - f(y))^2 \mu_{xy} \\
&= (1+a)^2 \sum_{x\in\Omega, y\in\Omega^c} \mu_{xy} \\
&= (1+a)^2 \mu(\partial\Omega).
\end{aligned}$$

Hence,

$$\mathcal{R}(f) \le \frac{(1+a)^2 \mu(\partial\Omega)}{(1+a)\mu(\Omega)} = (1+a)h \le 2h,$$

which was to be proved. ∎

The following lower bound of λ_1 via h is most useful and is frequently used.

Theorem 9.2 (Cheeger's inequality) *We have*

$$\lambda_1 \ge \frac{h^2}{2}. \tag{9.21}$$

Before we prove this theorem, consider some examples.

Example. Consider a weighted path graph (V, μ) where $V = \{0, 1, ...N-1\}$, the edges are

$$0 \sim 1 \sim 2 \sim ... \sim N-1,$$

and the weights $\mu_{k-1,k} = m_k$, where $\{m_k\}_{k=1}^{N-1}$ is a given sequence of positive numbers. Then, for $1 \le k \le N-2$, we have

$$\mu(k) = \mu_{k-1,k} + \mu_{k,k+1} = m_k + m_{k+1},$$

and the same is true also for $k = 0, N-1$ if we define $m_{-1} = m_N = 0$. The Markov kernel is then

$$P(k, k+1) = \frac{\mu_{k,k+1}}{\mu(k)} = \frac{m_{k+1}}{m_k + m_{k+1}}.$$

Claim. *Assume that the sequence $\{m_k\}_{k=1}^{N-1}$ is increasing, that is, $m_k \le m_{k+1}$. Then $h \ge \frac{1}{2N}$.*

Proof. Let Ω be a subset of V with $\mu(\Omega) \le \frac{1}{2}\mu(V)$, and let $\overline{k-1, k}$ be an edge of the boundary $\partial\Omega$ with the largest possible k. We claim that either Ω or Ω^c is contained in $[0, k-1]$. Indeed, if there were vertices from both sets Ω and Ω^c outside $[0, k-1]$, that is, in $[k, N-1]$, then there would have been an edge $\overline{j-1, j} \in \partial\Omega$ with $j > k$, which contradicts the choice of k. It follows that either $\mu(\Omega) \leq \mu([0, k-1])$ or $\mu(\Omega^c) \leq \mu([0, k-1])$. However, since $\mu(\Omega) \leq \mu(\Omega^c)$, we obtain that in the both cases $\mu(\Omega) \leq \mu([0, k-1])$. We have

$$\begin{aligned} \mu([0, k-1]) &= \sum_{j=0}^{k-1} \mu(j) = \sum_{j=0}^{k-1} \left(\mu_{j-1,j} + \mu_{j,j+1}\right) \\ &= \sum_{j=0}^{k-1} (m_j + m_{j+1}) \\ &\leq 2km_k \end{aligned} \tag{9.22}$$

where we have used that $m_j \leq m_{j+1} \leq m_k$. Therefore

$$\mu(\Omega) \leq 2km_k.$$

On the other hand, we have

$$\mu(\partial\Omega) \geq \mu_{k-1,k} = m_k,$$

whence it follows that

$$\frac{\mu(\partial\Omega)}{\mu(\Omega)} \geq \frac{m_k}{2km_k} = \frac{1}{2k} \geq \frac{1}{2N},$$

which proves that $h \geq \frac{1}{2N}$. $\blacksquare$

Consequently, Theorem 9.2 yields

$$\lambda_1 \geq \frac{1}{8N^2}. \tag{9.23}$$

For comparison, in the case of a simple weight, the exact value of λ_1 is

$$\lambda_1 = 1 - \cos\frac{\pi}{N-1}$$

(see Exercises), which for large N is

$$\lambda_1 \approx \frac{\pi^2}{2(N-1)^2} \approx \frac{5}{N^2},$$

which is of the same order in N as the estimate (9.23).

Now assume that the weights m_k satisfy a stronger condition

$$m_{k+1} \geq cm_k,$$

for some constant $c > 1$ and all $k = 0, ..., N-2$. Then $m_k \geq c^{k-j} m_j$ for all $k \geq j$, which allows to improve the estimate (9.22) as follows

$$\begin{aligned} \mu([0, k-1]) &= \sum_{j=0}^{k-1} (m_j + m_{j+1}) \\ &\leq \sum_{j=0}^{k-1} \left(c^{j-k} m_k + c^{j+1-k} m_k\right) \\ &= m_k \left(c^{-k} + c^{1-k}\right) \left(1 + c + ... c^{k-1}\right) \\ &= m_k \left(c^{-k} + c^{1-k}\right) \frac{c^k - 1}{c - 1} \\ &\leq m_k \frac{c+1}{c-1}. \end{aligned}$$

Therefore, we obtain

$$\frac{\mu(\partial\Omega)}{\mu(\Omega)} \geq \frac{c-1}{c+1}$$

whence $h \geq \frac{c-1}{c+1}$ and, by Theorem 9.2,

$$\lambda_1 \geq \frac{1}{2}\left(\frac{c-1}{c+1}\right)^2. \tag{9.24}$$

Let us estimate the mixing time on the above path graph (V, μ). Since it is bipartite, the mixing time is given by

$$T = \frac{\ln \frac{1}{\varepsilon}}{\ln \frac{1}{1-\lambda_1}} \leq \frac{\ln \frac{1}{\varepsilon}}{\lambda_1},$$

where $\varepsilon << \min_k \frac{\mu(k)}{\mu(V)}$. Observe that

$$\mu(V) = \sum_{j=0}^{N-1} (m_j + m_{j+1}) \leq 2 \sum_{j=1}^{N-1} m_j$$

where we put $m_0 = m_N = 0$, whence

$$\min_k \frac{\mu(k)}{\mu(V)} \geq \frac{m_1}{2\sum_{j=1}^{N-1} m_j} = \frac{1}{2M},$$

where

$$M := \frac{\sum_{j=1}^{N-1} m_j}{m_1}.$$

Setting $\varepsilon = \frac{1}{M^2}$ (note that $M \geq N$ and N can be assumed large) we obtain

$$T \leq \frac{2 \ln M}{\lambda_1}.$$

For an arbitrary increasing sequence $\{m_k\}$, we obtain using (9.23) that

$$T \leq 8N^2 \ln M.$$

Consider the weights $m_k = c^k$ where $c > 1$. Then we have

$$M = \sum_{j=1}^{N-1} c^{j-1} = \frac{c^{N-1} - 1}{c - 1}$$

whence $\ln M \approx N \ln c$. Using also (9.24), we obtain

$$T \leq 4N \ln c \left(\frac{c+1}{c-1} \right)^2.$$

Note that T is linear in N!

Consider one more example: $m_k = k^p$ for some $p > 1$. Then

$$\frac{m_{k+1}}{m_k} = \left(1 + \frac{1}{k}\right)^p \geq \left(1 + \frac{1}{N}\right)^p =: c$$

If $N >> p$ then $c \approx 1 + \frac{p}{N}$ whence

$$\lambda_1 \geq \frac{1}{2}\left(\frac{c-1}{c+1}\right)^2 \approx \frac{1}{8}\frac{p^2}{N^2}.$$

In this case,

$$M = \sum_{j=1}^{N-1} j^p \leq N^{p+1},$$

whence $\ln M \leq (p+1)\ln N$ and

$$T \leq \frac{16\,(p+1)}{p^2} N^2 \ln N.$$

We precede the proof Theorem 9.2 by two lemmas. Given a function $f : V \to \mathbb{R}$ and an edge $\xi = \overline{xy}$, let us use the following notation[2]:

$$|\nabla_\xi f| := |\nabla_{xy} f| = |f(y) - f(x)|.$$

Lemma 9.3 (Co-area formula). *Given any real-valued function f on V, set for any $t \in \mathbb{R}$*

$$\Omega_t = \{x \in V : f(x) > t\}.$$

Then the following identity holds:

$$\sum_{\xi \in E} |\nabla_\xi f| \, \mu_\xi = \int_{-\infty}^{\infty} \mu(\partial\Omega_t) \, dt. \tag{9.25}$$

A similar formula holds for differentiable functions on $\mathbb{R}$:

$$\int_a^b |f'(x)| = \int_{-\infty}^{\infty} \# \{x : f(x) = t\} \, dt,$$

[2]Note that $\nabla_\xi f$ is undefined unless the edge ξ is directed, whereas $|\nabla_\xi f|$ makes always sense.

and the common value of the both sides is called the *full variation* of f.

Proof. For any edge $\xi = \overline{xy}$, there corresponds an interval $I_\xi \subset \mathbb{R}$ that is defined as follows:

$$I_\xi = [f(x), f(y))$$

where we assume that $f(x) \leq f(y)$ (otherwise, switch the notations x and y). Denoting by $|I_\xi|$ the Euclidean length of the interval I_ξ, we see that $|\nabla_\xi f| = |I_\xi|$.

Claim. $\xi \in \partial\Omega_t \Leftrightarrow t \in I_\xi$.

Indeed, the boundary $\partial\Omega_t$ consists of edges $\xi = \overline{xy}$ such that $x \in \Omega_t^c$ and $y \in \Omega_t$, that is, $f(x) \leq t$ and $f(y) > t$; which is equivalent to $t \in [f(x), f(y)) = I_\xi$.

Thus, we have

$$\mu(\partial\Omega_t) = \sum_{\xi \in \partial\Omega_t} \mu_\xi = \sum_{\xi \in E : t \in I_\xi} \mu_\xi = \sum_{\xi \in E} \mu_\xi \, \mathbf{1}_{I_\xi}(t),$$

whence

$$\begin{aligned}\int_{-\infty}^{+\infty} \mu(\partial\Omega_t)\,dt &= \int_{-\infty}^{+\infty} \sum_{\xi\in E} \mu_\xi\, \mathbf{1}_{I_\xi}(t)dt \\ &= \sum_{\xi\in E} \int_{-\infty}^{+\infty} \mu_\xi\, \mathbf{1}_{I_\xi}(t)dt \\ &= \sum_{\xi\in E} \mu_\xi\, |I_\xi| = \sum_{\xi\in E} \mu_\xi |\nabla_\xi f|\,,\end{aligned}$$

which finishes the proof. ■

Lemma 9.4 *For any non-negative function f on V, such that*

$$\mu\,\{x\in V : f\,(x) > 0\} \leq \frac{1}{2}\mu\,(V)\,, \tag{9.26}$$

the following is true:

$$\sum_{\xi\in E} |\nabla_\xi f|\,\mu_\xi \geq h \sum_{x\in V} f\,(x)\,\mu\,(x)\,, \tag{9.27}$$

where h is the Cheeger constant of (V,μ).

Note that for the function $f = \mathbf{1}_\Omega$ the condition (9.26) means that $\mu(\Omega) \leq \frac{1}{2}\mu(V)$, and the inequality (9.27) is equivalent to

$$\mu(\partial\Omega) \geq h\mu(\Omega),$$

because

$$\sum_{x\in V} f(x)\mu(x) = \sum_{x\in\Omega}\mu(x) = \mu(\Omega)$$

and

$$\sum_{\xi\in E} |\nabla_\xi f|\,\mu_\xi = \sum_{x\in\Omega,y\in\Omega^c} |f(y)-f(x)|\,\mu_{xy} = \sum_{x\in\Omega,y\in\Omega^c}\mu_{xy} = \mu(\partial\Omega).$$

Hence, the meaning of Lemma 9.4 is that the inequality (9.27) for the indicator functions implies the same inequality for arbitrary functions.

Proof. By the co-area formula, we have

$$\sum_{\xi\in E} |\nabla_\xi f|\,\mu_\xi = \int_{-\infty}^{\infty}\mu(\partial\Omega_t)\,dt \geq \int_0^{\infty}\mu(\partial\Omega_t)\,dt.$$

By (9.26), the set $\Omega_t = \{x \in V : f(x) > t\}$ has measure $\leq \frac{1}{2}\mu(V)$ for any $t \geq 0$. Therefore, by (9.20)

$$\mu(\partial\Omega_t) \geq h\mu(\Omega_t)$$

whence

$$\sum_{\xi \in E} |\nabla_\xi f| \, \mu_\xi \geq h \int_0^\infty \mu(\Omega_t) \, dt$$

On the other hand, noticing that $x \in \Omega_t$ for a non-negative t is equivalent to $t \in [0, f(x))$, we obtain

$$\begin{aligned}
\int_0^\infty \mu(\Omega_t) \, dt &= \int_0^\infty \sum_{x \in \Omega_t} \mu(x) \, dt \\
&= \int_0^\infty \sum_{x \in V} \mu(x) \, \mathbf{1}_{[0, f(x))}(t) \, dt \\
&= \sum_{x \in V} \mu(x) \int_0^\infty \mathbf{1}_{[0, f(x))}(t) \, dt \\
&= \sum_{x \in V} \mu(x) f(x),
\end{aligned}$$

which finishes the proof. ■

Proof of Theorem 9.2. Let f be the eigenfunction of λ_1. Consider two sets

$$V^+ = \{x \in V : f(x) \geq 0\} \quad \text{and } V^- = \{x \in V : f(x) < 0\}.$$

Without loss of generality, we can assume that $\mu(V^+) \leq \mu(V^-)$ (if not then replace f by $-f$). It follows that $\mu(V^+) \leq \frac{1}{2}\mu(V)$. Consider the function

$$g = f_+ := \begin{cases} f, & f \geq 0, \\ 0, & f < 0. \end{cases}$$

Applying the Green formula (2.4)

$$(\mathcal{L}f, g) = \frac{1}{2} \sum_{x,y \in V} (\nabla_{xy} f)(\nabla_{xy} g)\, \mu_{xy}$$

and using so that $\mathcal{L}f = \lambda_1 f$, we obtain

$$\lambda_1 \sum_{x \in V} f(x) g(x) \mu(x) = \frac{1}{2} \sum_{x,y \in V} (\nabla_{xy} f)(\nabla_{xy} g)\, \mu_{xy}.$$

Observing that $fg = g^2$ and

$$(\nabla_{xy} f)(\nabla_{xy} g) = (f(y) - f(x))(g(y) - g(x)) \geq (g(y) - g(x))^2 = |\nabla_{xy} g|^2.$$

we obtain

$$\lambda_1 \geq \frac{\sum_{\xi \in E} |\nabla_\xi g|^2 \mu_\xi}{\sum_{x \in V} g^2(x)\, \mu(x)}.$$

Note that $g \not\equiv 0$ because otherwise $f_+ \equiv 0$ and $(f, 1) = 0$ imply that $f_- \equiv 0$ whereas f is not identical 0.

Hence, to prove (9.21) it suffices to verify that

$$\sum_{\xi \in E} |\nabla_\xi g|^2 \mu_\xi \geq \frac{h^2}{2} \sum_{x \in V} g^2(x) \mu(x). \tag{9.28}$$

Since

$$\mu(x \in V : g(x) > 0) \leq \mu(V^+) \leq \frac{1}{2} \mu(V),$$

we can apply (9.27) to function g^2, which yields

$$\sum_{\xi \in E} \left|\nabla_\xi (g^2)\right| \mu_\xi \geq h \sum_{x \in V} g^2(x) \mu(x). \tag{9.29}$$

Let us estimate from above the left hand side as follows:

$$\begin{aligned}
\sum_{\xi\in E}\left|\nabla_\xi\left(g^2\right)\right|\mu_\xi &= \frac{1}{2}\sum_{x,y\in V}\left|g^2(x)-g^2(y)\right|\mu_{xy} \\
&= \frac{1}{2}\sum_{x,y}|g(x)-g(y)|\mu_{xy}^{1/2}|g(x)+g(y)|\mu_{xy}^{1/2} \\
&\leq \left(\frac{1}{2}(\sum_{x,y}(g(x)-g(y))^2\mu_{xy})\frac{1}{2}(\sum_{x,y}(g(x)+g(y))^2\mu_{xy})\right)^{1/2},
\end{aligned}$$

where we have used the Cauchy-Schwarz inequality

$$\sum_k a_k b_k \leq \left(\sum_k a_k^2\right)^{1/2}\left(\sum_k b_k^2\right)^{1/2}$$

that is true for arbitrary sequences of non-negative reals a_k, b_k. Next, using the inequality $\frac{1}{2}(a+b)^2 \leq a^2+b^2$, we obtain

$$\begin{aligned}
\sum_{\xi\in E}\left|\nabla_\xi\left(g^2\right)\right|\mu_\xi &\leq \left(\sum_{\xi\in E}\left|\nabla_\xi g\right|^2\mu_\xi\sum_{x,y}\left(g^2(x)+g^2(y)\right)\mu_{xy}\right)^{1/2} \\
&= \left(2\sum_{\xi\in E}\left|\nabla_\xi g\right|^2\mu_\xi\sum_{x,y}g^2(x)\mu_{xy}\right)^{1/2} \\
&= \left(2\sum_{\xi\in E}\left|\nabla_\xi g\right|^2\mu_\xi\sum_{x\in V}g^2(x)\mu(x)\right)^{1/2},
\end{aligned}$$

which together with (9.29) yields

$$h\sum_{x\in V}g^2(x)\,\mu(x) \leq \left(2\sum_{\xi\in E}\left|\nabla_\xi g\right|^2\mu_\xi\right)^{1/2}\left(\sum_{x\in V}g^2(x)\mu(x)\right)^{1/2}.$$

Dividing by $\left(\sum_{x\in V}g^2(x)\,\mu(x)\right)^{1/2}$ and taking square, we obtain (9.28).
■