Stanford University — Graph Partitioning and Expanders | Handout 2
Luca Trevisan | April 30, 2013

Lecture 2

In which we prove Cheeger's inequality and we compute the spectrum of Cayley graphs.[1]

1 Expansion and The Second Eigenvalue

Let $G = (V, E)$ be an undirected d-regular graph, A its adjacency matrix, $L = 1 - \frac{1}{d} \cdot A$ its Laplacian matrix, and $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be the eigenvalues of L.

Recall that we defined the *edge expansion* of a cut $(S, V - S)$ of the vertices of G as

$$\phi(S) := \frac{E(S, V - S)}{d \cdot |S|}$$

and that the edge expansion of G is

$$\phi(G) := \min_{S \subseteq V,\ |S| \leq \frac{1}{2}|V|} \phi(S)$$

We also defined the related notion of the *sparsity* of a cut $(S, V - S)$ as

$$\sigma(S) := \frac{E(S, V - S)}{\frac{d}{n} \cdot |S| \cdot |V - S|}$$

and $\sigma(G) := \min_S \sigma(S)$; the *sparsest cut* problem is to find a cut of minimal sparsity.

Recall also that in the last lecture we proved that $\lambda_2 = 0$ if and only if G is disconnected, that is, $\lambda_2 = 0$ if and only if $\phi(G) = 0$. In this lecture we will see that this statement admits an *approximate version* that, qualitatively, says that λ_2 is small if and only if $\phi(G)$ is small. Quantitatively, we have

---

Theorem 1 (Cheeger's Inequalities)

$$\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2 \cdot \lambda_2} \tag{1}$$

1.1 The Easy Direction

In this section we prove

Lemma 2 $\lambda_2 \leq \sigma(G) \leq 2\phi(G)$

From which we have one direction of Cheeger's inequality.

Let us find an equivalent restatement of the sparsest cut problem. If we represent a set $S \subseteq V$ as a bit-vector $\mathbf{x} \in \{0,1\}^V$, then

$$E(S, V - S) = \sum_{\{u,v\} \in E} |x_u - x_v|$$

and

$$|S| \cdot |V - S| = \sum_{\{u,v\}} |x_u - x_v|$$

so that, after some simplifications, we can write

$$\sigma(G) = \min_{\mathbf{x} \in \{0,1\}^V - \{\mathbf{0},\mathbf{1}\}} \frac{\sum_{\{u,v\} \in E} |x_u - x_v|}{\frac{d}{n} \sum_{\{u,v\}} |x_u - x_v|} \tag{2}$$

Note that, when x_u, x_v take boolean values, then so does $|x_u - x_v|$, so that we may also equivalently write

$$\sigma(G) = \min_{\mathbf{x} \in \{0,1\}^V - \{\mathbf{0},\mathbf{1}\}} \frac{\sum_{\{u,v\} \in E} |x_u - x_v|^2}{\frac{d}{n} \sum_{\{u,v\}} |x_u - x_v|^2} \tag{3}$$

In the last lecture, we gave the following characterization of λ_2:

$$\lambda_2 = \min_{\mathbf{x} \in \mathbb{R}^V - \{\mathbf{0}\}, \mathbf{x} \perp \mathbf{1}} \frac{\sum_{\{u,v\} \in E} |x_u - x_v|^2}{d \cdot \sum_v x_v^2}$$

Now we claim that the following characterization is also true

$$\lambda_2 = \min_{\mathbf{x} \in \mathbb{R}^V - \{\mathbf{0},\mathbf{1}\}} \frac{\sum_{\{u,v\} \in E} |x_u - x_v|^2}{\frac{d}{n} \sum_{\{u,v\}} |x_u - x_v|^2} \tag{4}$$

This is because

$$\begin{aligned}
&\sum_{u,v} |x_u - x_v|^2 \\
&= \sum_{u,v} x_i^2 + \sum_{u,v} x_v^2 - 2\sum_{u,v} x_u x_v \\
&= 2n \sum_v x_v^2 - 2\left(\sum_v x_v\right)^2
\end{aligned}$$

so for every $\mathbf{x} \in \mathbb{R}^V - \{\mathbf{0}\}$ such that $\mathbf{x} \perp \mathbf{1}$ we have that

$$\sum_v x_v^2 = \frac{1}{2n}\sum_{u,v} |x_u - x_v|^2 = \frac{1}{n}\sum_{\{u,v\}} |x_u - x_v|^2 \ ,$$

and so

$$\min_{\mathbf{x}\in\mathbb{R}^V-\{\mathbf{0}\},\mathbf{x}\perp\mathbf{1}} \frac{\sum_{\{u,v\}\in E} |x_u - x_v|^2}{d \cdot \sum_v x_v^2} = \min_{\mathbf{x}\in\mathbb{R}^V-\{\mathbf{0}\},\mathbf{x}\perp\mathbf{1}} \frac{\sum_{\{u,v\}\in E} |x_u - x_v|^2}{\frac{d}{n}\sum_{\{u,v\}} |x_u - x_v|^2}$$

To conclude the argument, take an $\mathbf{x}$ that maximizes the right-hand side of (4), and observe that if we shift every coordinate by the same constant then we obtain another optimal solution, because the shift will cancel in all the expressions both in the numerator and the denominator. In particular, we can define $\mathbf{x}'$ such that $x'_v = x_v - \frac{1}{n}\sum_u x_u$ and note that the entries of $\mathbf{x}'$ sum to zero, and so $\mathbf{x}' \perp \mathbf{1}$. This proves that

$$\min_{\mathbf{x}\in\mathbb{R}^V-\{\mathbf{0}\},\mathbf{x}\perp\mathbf{1}} \frac{\sum_{\{u,v\}\in E} |x_u - x_v|^2}{\frac{d}{n}\sum_{u,v} |x_u - x_v|^2} = \min_{\mathbf{x}\in\mathbb{R}^V-\{\mathbf{0},\mathbf{1}\}} \frac{\sum_{\{u,v\}\in E} |x_u - x_v|^2}{\frac{d}{n}\sum_{u,v} |x_u - x_v|^2}$$

and so we have established (4).

Comparing (4) and (3), it is clear that the quantity λ_2 is a *continuous relaxation* of $\sigma(G)$, and hence $\lambda_2 \leq \sigma(G)$.

1.2 Other Relaxations of $\sigma(G)$

Having established that we can view λ_2 as a *relaxation* of $\sigma(G)$, the proof that $\phi(G) \leq \sqrt{2 \cdot \lambda_2}$ can be seen as a *rounding* algorithm, that given a real-valued solution to (4) finds a comparably good solution for (3).

Later in the course we will see two more approximation algorithms for sparsest cut and edge expansion. Both are based on continuous relaxations of σ starting from (2).

The algorithm of Leighton and Rao is based on a relaxation that is defined by observing that every bit-vector $\mathbf{x} \in \{0,1\}^V$ defines the semi-metric $dist(u,v) := |x_u - x_v|$ over the vertices; the Leighton-Rao relaxation is obtained by allowing arbitrary semi-metrics:

$$LR(G) := \min_{\substack{dist : V \times V \to \mathbb{R} \\ dist \text{ semimetric}}} \frac{\sum_{\{u,v\}\in E} dist(u,v)}{\frac{d}{n}\sum_{\{u,v\}} dist(u,v)}$$

It is not difficult to express $LR(G)$ as a linear programming problem.

The algorithm of Arora-Rao-Vazirani is obtained by noting that, for a bit-vector $x \in \{0,1\}^V$, the distances $dist(u,v) := |x_u - x_v|$ define a metric which can also be seen as the Euclidean distance between the x_v, because $|x_u - x_v| = \sqrt{(x_u - x_v)^2}$, and such that $dist^2(u,v)$ is also a semi-metric, trivially so because $dist^2(u,v) = dist(u,v)$. If a distance function $dist(\cdot,\cdot)$ is a semi-metric such that $\sqrt{dist(\cdot,\cdot)}$ is a Euclidean semi-metric, then $dist(\cdot,\cdot)$ is called a *negative type* semi-metric. The Arora-Rao-Vazirani relaxation is

$$ARV(G) := \min_{\substack{dist : V \times V \to \mathbb{R} \\ dist \text{ negative type semimetric}}} \frac{\sum_{\{u,v\}\in E} dist(u,v)}{\frac{d}{n}\sum_{\{u,v\}} dist(u,v)}$$

The Arora-Rao-Vazirani relaxation can be expressed as a semi-definite programming problem.

From this discussion it is clear that the Arora-Rao-Vazirani relaxation is a tightening of the Leigthon-Rao relaxation and that we have

$$\sigma(G) \geq ARV(G) \geq LR(G)$$

It is less obvious in this treatment, and we will see it later, that the Arora-Rao-Vazirani is also a tightening of the relaxation of σ given by λ_2, that is

$$\sigma(G) \geq ARV(G) \geq \lambda_2$$

The relaxations λ_2 and $LR(G)$ are incomparable.

1.3 Spectral partitioning and the proof of the difficult direction

The proof of the more difficult direction of Theorem 1 will be constructive and algorithmic. The proof can be seen as an analysis of the following algorithm.

Algorithm: *SpectralPartitioning*

- Input: graph $G = (V, E)$ and vector $\mathbf{x} \in \mathbb{R}^V$
- Sort the vertices of V in non-decreasing order of values of entries in $\mathbf{x}$, that is let $V = \{v_1, \ldots, v_n\}$ where $x_{v_1} \leq x_{v_2} \leq \ldots x_{v_n}$
- Let $i \in \{1, \ldots, n-1\}$ be such that $\max\{\phi(\{v_1, \ldots, v_i\}), \phi(\{v_{i+1}, \ldots, v_n\})\}$ is minimal
- Output $S = \{v_1, \ldots, v_i\}$ and $\bar{S} = \{v_{i+1}, \ldots, v_n\}$

We note that the algorithm can be implemented to run in time $O(|V| \log |V| + |E|)$, assuming arithmetic operations and comparisons take constant time, because once we have computed $E(\{v_1, \ldots, v_i\}, \{v_{i+1}, \ldots, v_n\})$ it only takes time $O(degree(v_{i+1}))$ to compute $E(\{v_1, \ldots, v_{i+1}\}, \{v_{i+2}, \ldots, v_n\})$.

We have the following analysis of the quality of the solution:

Lemma 3 (Analysis of Spectral Partitioning) *Let $G = (V, E)$ be a d-regular graph, $\mathbf{x} \in \mathbb{R}^V$ be a vector such that $\mathbf{x} \perp \mathbf{1}$, define*

$$R(\mathbf{x}) := \frac{\sum_{\{u,v\} \in E} |x_u - x_v|^2}{d \cdot \sum_v x_v^2}$$

and let S be the output of algorithm SpectralPartitioning on input G and $\mathbf{x}$. Then

$$\phi(S) \leq \sqrt{2R(x)}$$

Remark 4 *If we apply the lemma to the case in which $\mathbf{x}$ is an eigenvector of λ_2, then $R(\mathbf{x}) = \lambda_2$, and so we have*

$$\phi(S) \leq \sqrt{2 \cdot \lambda_2}$$

which is the difficult direction of Cheeger's inequalities.

Remark 5 *If we run the SpectralPartitioning algorithm with the eigenvector $\mathbf{x}$ of the second eigenvalue λ_2, we find a set S whose expansion is*

$$\phi(S) \leq \sqrt{2 \cdot \lambda_2} \leq 2\sqrt{\phi(G)}$$

Even though this doesn't give a constant-factor approximation to the edge expansion, it gives a very efficient, and non-trivial, approximation.

As we will see in a later lecture, there is a nearly linear time algorithm that finds a vector $\mathbf{x}$ *for which the Rayleigh quotient* $R(\mathbf{x})$ *is very close to* λ_2*, so, overall, for any graph* G *we can find a cut of expansion* $O(\sqrt{\phi(G)})$ *in nearly linear time.*

1.4 Proof of Lemma 3

We saw that λ_2 can be seen as a relaxation of $\sigma(G)$, and Lemma 3 provides a *rounding algorithm* for the real vectors which are solutions of the relaxation. In this section we will think of it as a form of *randomized rounding*. Later, when we talk about the Leighton-Rao sparsest cut algorithm, we will revisit this proof and think of it in terms of *metric embeddings*.

To simplify notation, we will assume that $V = \{1, \ldots, n\}$ and that $x_1 \leq x_2 \leq \cdots x_n$. Thus our goal is to prove that there is an i such that $\phi(\{1, \ldots, i\}) \leq \sqrt{2R(\mathbf{x})}$ and $\phi(\{i+1, \ldots, n\}) \leq \sqrt{2R(\mathbf{x})}$

We will derive Lemma 3 by showing that there is a distribution D over sets S of the form $\{1, \ldots, i\}$ such that

$$\frac{\mathbb{E}_{S \sim D} E(S, V-S)}{\mathbb{E}_{S \sim D} d \cdot \min\{|S|, |V-S|\}} \leq \sqrt{2R(\mathbf{x})} \tag{5}$$

We need to be a bit careful in deriving the Lemma from (5). In general, it is not true that a ratio of averages is equal to the average of the ratios, so (5) does not imply that $\mathbb{E}\,\phi(S) \leq \sqrt{2R(\mathbf{x})}$. We can, however, apply linearity of expectation and derive from (5) the inequality

$$\mathop{\mathbb{E}}_{S \sim D} \frac{1}{d} E(S, V-S) - \sqrt{2R(\mathbf{x})} \min\{|S|, |V-S|\} \leq 0$$

So there must exist a set S in the sample space such that

$$\frac{1}{d} E(S, V-S) - \sqrt{2R(\mathbf{x})} \min\{|S|, |V-S|\} \leq 0$$

meaning that, for that for both the set S and its complement, we have $\phi(S) \leq \sqrt{2R(\mathbf{x})}$; at least one of the sets has size at most $n/2$, and so we are done. (Basically we are using the fact that, for random variables X, Y over the same sample space, although it might not be true that $\frac{\mathbb{E} X}{\mathbb{E} Y} = \mathbb{E}\frac{X}{Y}$, we always have $\mathbb{P}[\frac{X}{Y} \leq \frac{\mathbb{E} X}{\mathbb{E} Y}] > 0$, provided that $Y > 0$ over the entire sample space.)

From now on, we will assume that

1. $x_{\lceil n/2 \rceil} = 0$, that is, the median of the entries of $\mathbf{x}$ is zero
2. $x_1^2 + x_n^2 = 1$

which can be done without loss of generality because, if $\mathbf{x} \perp \mathbf{1}$, adding a fixed constant c to all entries of $\mathbf{x}$ can only reduce the Rayleigh quotient:

$$\begin{aligned} R(\mathbf{x} + (c, \dots, c)) &= \frac{\sum_{\{u,v \in E\}} |(x_u + c) - (x_v + x)|^2}{d \sum_v (x_v + c)^2} \\ &= \frac{\sum_{\{u,v \in E\}} |x_u - x_v|^2}{d \sum_v x_v^2 - 2dc \sum_v x_v + nc^2} \\ &= \frac{\sum_{\{u,v \in E\}} |x_u - x_v|^2}{d \sum_v x_v^2 + nc^2} \\ &\leq R(\mathbf{x}) \end{aligned}$$

Multiplying all the entries by a fixed constant does not change the value of $R(\mathbf{x})$, nor does it change the property that $x_1 \leq \cdots \leq x_n$. The reason for these choices is that they allow us to define a distribution D over sets such that

$$\mathop{\mathbb{E}}_{S \sim D} \min\{|S|, |V - S|\} = \sum_i x_i^2 \tag{6}$$

We define the distribution D over sets of the form $\{1, \dots, i\}$, $i \leq n-1$, as the outcome of the following probabilistic process:

- We pick a real value t in the range $[x_1, x_n]$ with probabily density function $f(t) = 2|t|$. That is, for $x_1 \leq a \leq b \leq x_n$, $\mathbb{P}[a \leq t \leq b] = \int_a^b 2|t|dt$.

  Doing the calculation, this means that $\mathbb{P}[a \leq t \leq b] = |a^2 - b^2|$ if a, b have the same sign, and $\mathbb{P}[a \leq t \leq b] = a^2 + b^2$ if they have different signs.
- We let $S := \{i : x_i \leq t\}$

According to this definition, the probability that an element $i \leq n/2$ belongs to the smallest of the sets $S, V - S$ is the same as the probability that it belongs to S, which is the probability that the threshold t is in the range $[x_i, 0]$, and that probability is x_i^2. Similarly, the probability that an element $i > n/2$ belongs to the smallest of $S, V - S$

is the same as the probability that it belongs to $V - S$, which is the probability that t is in the range $[0, x_i]$, which is again x_i^2. So we have established (6).

We will now estimate the expected number of edges between S and $V - S$.

$$\mathbb{E}\,\frac{1}{d}E(S, V - S) = \frac{1}{2}\sum_{i,j} M_{i,j}\,\mathbb{P}[(i,j) \text{ is cut by } (S, V - S)]$$

The event that the edge (i, j) is cut by the partition $(S, V - S)$ happens when the value t falls in the range between x_i and x_j. This means that

- If x_i, x_j have the same sign,

$$\mathbb{P}[(i,j) \text{ is cut by } (S, V - S)] = |x_i^2 - x_j^2|$$

- If x_i, x_j have different sign,

$$\mathbb{P}[(i,j) \text{ is cut by } (S, V - S)] = x_i^2 + x_j^2$$

Some attempts, show that a good expression to upper bound both cases is

$$\mathbb{P}[(i,j) \text{ is cut by } (S, V - S)] \le |x_i - x_j| \cdot (|x_i| + |x_j|)$$

Plugging into our expression for the expected number of cut edges, and applying Cauchy-Schwarz

$$\begin{aligned}\mathbb{E}\,E(S, V - S) &\le \sum_{\{i,j\}\in E} |x_i - x_j| \cdot (|x_i| + |x_j|) \\ &\le \sqrt{\sum_{\{i,j\}\in E} (x_i - x_j)^2} \cdot \sqrt{\sum_{\{i,j\}\in E} (|x_i| + |x_j|)^2}\end{aligned}$$

Finally, it remains to study the expression $\sum_{\{i,j\}\in E}(|x_i| + |x_j|)^2$. By applying the inequality $(a + b)^2 \le 2a^2 + 2b^2$ (which follows by noting that $2a^2 + 2b^2 - (a + b)^2 = (a - b)^2 \ge 0$), we derive

$$\sum_{\{i,j\}\in E} (|x_i| + |x_j|)^2 \le \sum_{\{i,j\}\in E} (2x_i^2 + 2x_j^2) = 2d\sum_i x_i^2$$

Putting all the pieces together we have

$$\frac{\mathbb{E}\,E(S, V - S)}{d\,\mathbb{E}\min\{|S|, |V - S|\}} \le \frac{\sqrt{\sum_{\{i,j\}\in E} |x_i - x_j|^2} \cdot \sqrt{2d\sum_i x_i^2}}{d\sum_i x_i^2} = \sqrt{2R(\mathbf{x})} \qquad (7)$$

which, together with (6) gives (5), which, as we already discussed, implies the Main Lemma 3.

2 Eigenvalues of Cayley Graphs

So far we have proved that

$$\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2 \cdot \lambda_2}$$

and that the SpectralPartitioning algorithm, when given an eigenvector of λ_2, finds a cut $(S, V - S)$ such that $\phi(S) \leq 2\sqrt{\phi(G)}$. In this section we will show that all such results are tight, up to constants, by proving that

- The dimension-d hypercube H_d has $\lambda_2 = \frac{2}{d}$ and $\phi(H_d) = \frac{1}{d}$, giving an infinite family of graphs for which $\frac{\lambda_2}{2} = \phi(G)$, showing that the first Cheeger inequality is exactly tight.
- The n-cycle C_n has $\lambda_2 = O(n^{-2})$, and $\phi(C_n) = \frac{2}{n}$, giving an infinite family of graphs for which $\phi(G) = \Omega(\sqrt{\lambda_2})$, showing that the second Cheeger inequality is tight up to a constant.
- There is an eigenvector of the 2nd eigenvalue of the hypercube H_d, such that the SpectralPartitioning algorithm, given such a vector, outputs a cut $(S, V - S)$ of expansion $\phi(S) = \Omega(1/\sqrt{d})$, showing that the analysis of the SpectralPartitioning algorithm is tight up to a constant.

We will develop some theoretical machinery to find the eigenvalues and eigenvectors of *Cayley graphs of finite Abelian groups*, a class of graphs that includes the cycle and the hypercube, among several other interesting examples. This theory will also be useful later, as a starting point to talk about algebraic constructions of expanders.

For readers familiar with the Fourier analysis of Boolean functions, or the discrete Fourier analysis of functions $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$, or the standard Fourier analysis of periodic real functions, this theory will give a more general, and hopefully interesting, way to look at what they already know.

2.1 Characters

We will use additive notation for groups, so, if Γ is a group, its unit will be denoted by 0, its group operation by +, and the inverse of element a by $-a$. Unless, noted otherwise, however, the definitions and results apply to non-abelian groups as well.

Definition 6 (Character) *Let Γ be a group (we will also use Γ to refer to the set of group elements). A function $f : \Gamma \to \mathbb{C}$ is a character of Γ if*

- *f is a group homomorphism of Γ into the multiplicative group $\mathbb{C} - \{0\}$.*
- *for every $x \in \Gamma$, $|f(x)| = 1$*

Though this definition might seem to not bear the slightest connection to our goals, the reader should hang on because we will see next time that finding the eigenvectors and eigenvalues of the cycle C_n is immediate once we know the characters of the group $\mathbb{Z}/n\mathbb{Z}$, and finding the eigenvectors and eigenvalues of the hypercube H_d is immediate once we know the characters of the group $(\mathbb{Z}/2\mathbb{Z})^d$.

Remark 7 (About the Boundedness Condition) *If Γ is a finite group, and a is any element, then*

$$\underbrace{a + \cdots + a}_{|\Gamma| \text{ times}} = 0$$

and so if $f : \Gamma \to \mathbb{C}$ is a group homomorphism then

$$1 = f(0) = f(\underbrace{a + \cdots + a}_{|\Gamma| \text{ times}}) = f(a)^{|\Gamma|}$$

and so $f(a)$ is a root of unity and, in particular, $|f(a)| = 1$. This means that, for finite groups, the second condition in the definition of character is redundant.

In certain infinite groups, however, the second condition does not follow from the first, for example $f : \mathbb{Z} \to \mathbb{C}$ defined as $f(n) = e^n$ is a group homomorphism of $(\mathbb{Z}, +)$ into $(\mathbb{C} - \{0\}, \cdot)$ but it is not a character.

Just by looking at the definition, it might look like a finite group might have an infinite number of characters; the above remark, however, shows that a character of a finite group Γ must map into $|\Gamma|$-th roots of unity, of which there are only $|\Gamma|$, showing a finite $|\Gamma|^{|\Gamma|}$ upper bound to the number of characters. Indeed, a much stronger upper bound holds, as we will prove next, after some preliminaries.

Lemma 8 *If Γ is finite and χ is a character that is not identically equal to 1, then $\sum_{a \in \Gamma} \chi(a) = 0$*

Proof: Let b be such that $\chi(b) \neq 1$. Note that

$$\chi(b) \cdot \sum_{a \in \Gamma} \chi(a) = \sum_{a \in \Gamma} \chi(b + a) = \sum_{a \in \Gamma} \chi(a)$$

where we used the fact that the mapping $a \to b + a$ is a permutation. (We emphasize that even though we are using additive notation, the argument applies to non-abelian groups.) So we have

$$(\chi(b) - 1) \cdot \sum_{a \in \Gamma} \chi(a) = 0$$

and since we assumed $\chi(b) \neq 1$, it must be $\sum_{a \in \Gamma} \chi(a) = 0$. $\Box$

If Γ is finite, given two functions $f, g : \Gamma \to \mathbb{C}$, define the inner product

$$\langle f, g \rangle := \sum_{a \in \Gamma} f(a) g^*(a)$$

Lemma 9 *If $\chi_1, \chi_2 : \Gamma \to \mathbb{C}$ are two different characters of a finite group Γ, then*

$$\langle \chi_1, \chi_2 \rangle = 0$$

We will prove Lemma 9 shortly, but before doing so we note that, for a finite group Γ, the set of functions $f : \Gamma \to \mathbb{C}$ is a $|\Gamma|$-dimensional vector space, and that Lemma 9 implies that characters are orthogonal with respect to an inner product, and so they are linearly independent. In particular, we have established the following fact:

Corollary 10 *If Γ is a finite group, then it has at most $|\Gamma|$ characters.*

It remains to prove Lemma 9, which follows from the next two statements, whose proof is immediate from the definitions.

Fact 11 *If χ_1, χ_2 are characters of a group Γ, then the mapping $x \to \chi_1(x) \cdot \chi_2(x)$ is also a character.*

Fact 12 *If χ is a character of a group Γ, then the mapping $x \to \chi^*(x)$ is also a character, and, for every x, we have $\chi(x) \cdot \chi^*(x) = 1$.*

To complete the proof of Lemma 9, observe that:

- the function $\chi(x) := \chi_1(x) \cdot \chi_2^*(x)$ is a character;
- the assumption of the lemma is that there is an a such that $\chi_1(a) \neq \chi_2(a)$, and so, for the same element a, $\chi(a) = \chi_1(a) \cdot \chi_2^*(a) \neq \chi_2(a) \cdot \chi_2^*(a) = 1$

- thus χ is a character that is not identically equal to 1, and so

$$0 = \sum_a \chi(a) = \langle \chi_1, \chi_2 \rangle$$

Notice that, along the way, we have also proved the following fact:

Fact 13 *If Γ is a group, then the set of characters of Γ is also a group, with respect to the group operation of pointwise multiplication. The unit of the group is the character mapping every element to 1, and the inverse of a character is the pointwise conjugate of the character.*

The group of characters is called the Pontryagin dual of Γ, and it is denoted by $\hat{\Gamma}$.

We now come to the punchline of this discussion.

Theorem 14 *If Γ is a finite abelian group, then it has exactly $|\Gamma|$ characters.*

PROOF: We give a constructive proof. We know that every finite abelian group is isomorphic to a product of cyclic groups

$$(\mathbb{Z}/n_1\mathbb{Z}) \times (\mathbb{Z}/n_2\mathbb{Z}) \times \cdots \times (\mathbb{Z}/n_k\mathbb{Z})$$

so it will be enough to prove that

1. the cyclic group $\mathbb{Z}/n\mathbb{Z}$ has n characters;
2. if Γ_1 and Γ_2 are finite abelian groups with $|\Gamma_1|$ and $|\Gamma_2|$ characters, respectively, then their product has $|\Gamma_1| \cdot |\Gamma_2|$ characters.

For the first claim, consider, for every $r \in \{0, \ldots, n-1\}$, the function

$$\chi_r(x) := e^{2\pi irx/n}$$

Each such function is clearly a character (0 maps to 1, $\chi_r(-x)$ is the multiplicative inverse of $\chi_r(x)$, and, recalling that $e^{2\pi ik} = 1$ for every integer k, we also have $\chi_r(a + b \bmod n) = e^{2\pi ira/n} \cdot e^{2\pi irb/n}$), and the values of $\chi_r(1)$ are different for different values of r, so we get n distinct characters. This shows that $\mathbb{Z}/n\mathbb{Z}$ has at least n characters, and we already established that it can have at most n characters.

For the second claim, note that if χ_1 is a character of Γ_1 and χ_2 is a character of Γ_2, then it is easy to verify that the mapping $(x, y) \to \chi_1(x) \cdot \chi_2(y)$ is a character of $\Gamma_1 \times \Gamma_2$. Furthermore, if (χ_1, χ_2) and (χ'_1, χ'_2) are two distinct pairs of characters, then the mappings $\chi(x,y) := \chi_1(x)\cdot\chi_2(y)$ and $\chi'(x,y) := \chi_1'(x)\cdot\chi_2'(y)$ are two distinct characters of $\Gamma_1\times\Gamma_2$, because we either have an a such that $\chi_1(a)\neq\chi_1'(a)$, in which case $\chi(a,0)\neq\chi'(a,0)$, or we have a b such that $\chi_2(b)\neq\chi_2'(b)$, in which case $\chi(0,b)\neq\chi'(0,b)$. This shows that $\Gamma_1\times\Gamma_2$ has at least $|\Gamma_1|\cdot|\Gamma_2|$ characters, and we have already established that it can have at most that many $\square$

This means that the characters of a finite abelian group Γ form an orthogonal basis for the set of all functions $f:\Gamma\to\mathbb{C}$, so that any such function can be written as a linear combination

$$f(x) = \sum_\chi \hat{f}(\chi)\cdot\chi(x)$$

For every character χ, $\langle\chi,\chi\rangle = |\Gamma|$, and so the characters are actually a scaled-up orthonormal basis, and the coefficients can be computed as

$$\hat{f}(\chi) = \frac{1}{|\Gamma|}\sum_x f(x)\chi^*(x)$$

Example 15 (The Boolean Cube) *Consider the case* $\Gamma = (\mathbb{Z}/2\mathbb{Z})^n$, *that is the group elements are* $\{0,1\}^n$, *and the operation is bitwise xor. Then there is a character for every bit-vector* $(r_1,\ldots,r_n)$, *which is the function*

$$\chi_{r_1,\ldots,r_n}(x_1,\ldots,x_n) := (-1)^{r_1x_1+\cdots r_nx_n}$$

Every boolean function $f:\{0,1\}^n\to\mathbb{C}$ *can thus be written as*

$$f(x) = \sum_{r\in\{0,1\}^n}\hat{f}(r)\cdot(-1)^{\sum_i r_ix_i}$$

where

$$\hat{f}(r) = \frac{1}{2^n}\sum_{x\in\{0,1\}^n} f(x)\cdot(-1)^{\sum_i r_ix_i}$$

which is the boolean Fourier transform.

Example 16 (The Cyclic Group) *To work out another example, consider the case* $\Gamma = \mathbb{Z}/N\mathbb{Z}$. *Then every function* $f:\{0,\ldots,N-1\}\to\mathbb{C}$ *can be written as*

$$f(x) = \sum_{r\in\{0,\ldots,N-1\}}\hat{f}(r)e^{2\pi irx/n}$$

where

$$\hat{f}(x) = \frac{1}{N}\sum_x f(x)e^{-2\pi i r x/n}$$

which is the discrete Fourier transform.

2.2 A Look Beyond

Why is the term "Fourier transform" used in this context? We will sketch an answer to this question, although what we say from this point on is not needed for our goal of finding the eigenvalues and eigenvectors of the cycle and the hypercube.

The point is that it is possible to set up a definitional framework that unifies both what we did in the previous section with finite Abelian groups, and the Fourier series and Fourier transforms of real and complex functions.

In the discussion of the previous section, we started to restrict ourselves to finite groups Γ when we defined an inner product among functions $f : \Gamma \to \mathbb{C}$.

If Γ is an infinite abelian group, we can still define an inner product among functions $f : \Gamma \to \mathbb{C}$, but we will need to define a measure over Γ and restrict ourselves in the choice of functions. A measure μ over (a sigma-algebra of subsets of) Γ is a Haar measure if, for every measurable subset A and element a we have $\mu(a + A) = \mu(A)$, where $a + A = \{a + b : b \in A\}$. For example, if Γ is finite, $\mu(A) = |A|$ is a Haar measure. If $\Gamma = (\mathbb{Z}, +)$, then $\mu(A) = |A|$ is also a Haar measure (it is ok for a measure to be infinite for some sets), and if $\Gamma = (\mathbb{R}, +)$ then the Lebesgue measure is a Haar measure. When a Haar measure exists, it is more or less unique up to multiplicative scaling. All *locally compact topological* abelian groups have a Haar measure, a very large class of abelian groups, that include all finite ones, $(\mathbb{Z}, +)$, $(\mathbb{R}, +)$, and so on.

Once we have a Haar measure μ over Γ, and we have defined an integral for functions $f : \Gamma \to \mathbb{C}$, we say that a function is an element of $L^2(\Gamma)$ if

$$\int_\Gamma |f(x)|^2 d\mu(x) < \infty$$

For example, if Γ is finite, then all functions $f : \Gamma \to \mathbb{C}$ are in $L^2(\Gamma)$, and a function $f : \mathbb{Z} \to \mathbb{C}$ is in $L^2(\mathbb{Z})$ if the series $\sum_{n\in\mathbb{Z}} |f(n)|^2$ converges.

If $f, g \in L^2(\Gamma)$, we can define their inner product

$$\langle f, g\rangle := \int_\Gamma f(x)g^*(x)d\mu(x)$$

and use Cauchy-Schwarz to see that $|\langle f, g\rangle| < \infty$.

Now we can repeat the proof of Lemma 9 that $\langle \chi_1, \chi_2 \rangle = 0$ for two different characters, and the only step of the proof that we need to verify for infinite groups is an analog of Lemma 8, that is we need to prove that if χ is a character that is not always equal to 1, then

$$\int_\Gamma \chi(x) d\mu(x) = 0$$

and the same proof as in Lemma 8 works, with the key step being that, for every group element a,

$$\int_\Gamma \chi(x+a) d\mu(x) = \int_\Gamma \chi(x) d\mu(x)$$

because of the property of μ being a Haar measure.

We don't have an analogous result to Theorem 14 showing that Γ and $\hat{\Gamma}$ are isomorphic, however it is possible to show that $\hat{\Gamma}$ itself has a Haar measure $\hat{\mu}$, that the dual of $\hat{\Gamma}$ is isomorphic to Γ, and that if $f : \Gamma \to \mathbb{C}$ is continuous, then it can be written as the "linear combination"

$$f(x) = \int_{\hat{\Gamma}} \hat{f}(\chi) \chi(x) d\hat{\mu}(x)$$

where

$$\hat{f}(\chi) = \int_\Gamma f(x) \chi^*(x) d\mu(x)$$

In the finite case, the examples that we developed before correspond to setting $\mu(A) := |A|/|\Gamma|$ and $\hat{\mu}(A) = |A|$.

Example 17 (Fourier Series) *The set of characters of the group $[0,1)$ with the operation of addition modulo 1 is isomorphic to $\mathbb{Z}$, because for every integer n we can define the function $\chi_n : [0,1) \to \mathbb{C}$*

$$\chi_n(x) := e^{2\pi i x n}$$

and it can be shown that there are no other characters. We thus have the Fourier series for continuous functions $f : [0,1) \to \mathbb{C}$,

$$f(x) = \sum_{n \in \mathbb{Z}} \hat{f}(n) e^{2\pi i x n}$$

where

$$\hat{f}(n) = \int_0^1 f(x) e^{-2\pi i x n} dx$$

2.3 Cayley Graphs and Their Spectrum

Let Γ be a finite group. We will use additive notation, although the following definition applies to non-commutative groups as well. A subset $S \subseteq \Gamma$ is *symmetric* if $a \in S \Leftrightarrow -a \in S$.

Definition 18 *For a group Γ and a symmetric subset $S \subseteq \Gamma$, the Cayley graph $Cay(\Gamma, S)$ is the graph whose vertex set is Γ, and such that (a, b) is an edge if and only if $b - a \in S$. Note that the graph is undirected and $|S|$-regular.*

We can also define Cayley *weighted* graphs: if $w : \Gamma \to \mathbb{R}$ is a function such that $w(a) = w(-a)$ for every $a \in \Gamma$, then we can define the weighted graph $Cay(G, w)$ in which the edge (a, b) has weight $w(b - a)$. We will usually work with unweighted graphs.

Example 19 (Cycle) *The n-vertex cycle can be constructed as the Cayley graph $Cay(\mathbb{Z}/n\mathbb{Z}, \{-1, 1\})$.*

Example 20 (Hypercube) *The d-dimensional hypercube can be constructed as the Cayley graph*

$$Cay((\mathbb{Z}/2\mathbb{Z})^d, \{(1, 0, \ldots, 0), (0, 1, \ldots, 0), \ldots, (0, 0, \ldots, 1)\})$$

where the group is the set $\{0, 1\}^d$ with the operation of bit-wise xor, and the set S is the set of bit-vectors with exactly one 1.

If we construct a Cayley graph from a finite abelian group, then the eigenvectors are the characters of the groups, and the eigenvalues have a very simple description.

Lemma 21 *Let Γ be a finite abelian group, $\chi : \Gamma \to \mathbb{C}$ be a character of Γ, $S \subseteq \Gamma$ be a symmetric set. Let A be the adjacency matrix of the Cayley graph $G = Cay(\Gamma, S)$. Consider the vector $\mathbf{x} \in \mathbb{C}^\Gamma$ such that $x_a = \chi(a)$.*

Then $\mathbf{x}$ is an eigenvector of A, with eigenvalue

$$\sum_{s \in S} \chi(s)$$

Proof: Consider the a-th entry of $M\mathbf{x}$:

$$
\begin{aligned}
(A\mathbf{x})_a &= \sum_b A_{a,b} x_b \\
&= \sum_{b : b-a \in S} \chi(b) \\
&= \sum_{s \in S} \chi(a+s) \\
&= x_a \cdot \cdot \sum_{s \in S} \chi(s)
\end{aligned}
$$

And so

$$
A\mathbf{x} = \left(\sum_{s \in S} \chi(s) \right) \cdot \mathbf{x}
$$

□

The eigenvalues of the form $\sum_{s \in S} \chi(s)$, where χ is a character, enumerate all the eigenvalues of the graph, as can be deduced from the following observations:

1. Every character is an eigenvector;
2. The characters are linearly independent (as functions $\chi : \Gamma \to \mathbb{C}$ and, equivalently, as vectors in $\mathbb{C}^\Gamma$);
3. There are as many characters as group elements, and so as many characters as nodes in the corresponding Cayley graphs.

It is remarkable that, for a Cayley graph, a system of eigenvectors can be determined based solely on the underlying group, independently of the set S.

Once we have the eigenvalues of the adjacency matrix, we can derive the eigenvalues of the Laplacian, because to every eigenvalue λ of the adjacency there corresponds an eigenvalue $1 - \frac{1}{|S|}\lambda$ of the Laplacian, with the same eigenvector.

3 The Cycle

The n-cycle is the Cayley graph $Cay(\mathbb{Z}/n\mathbb{Z}, \{-1, +1\})$. Recall that, for every $n \in \{0, \ldots, n-1\}$, the group $\mathbb{Z}/n\mathbb{Z}$ has a character $\chi_r(x) = e^{2\pi i r x/n}$.

This means that for every $r \in \{0, \ldots, n-1\}$ we have the eigenvalue of the adjacency matrix

$$\lambda_r = e^{2\pi i r/n} + e^{-2\pi i r/n} = 2\cos(2\pi r/n)$$

where we used the facts that $e^{ix} = \cos(x) + i\sin(x)$, that $\cos(x) = \cos(-x)$, and $\sin(x) = -\sin(-x)$.

For $r = 0$ we have the eigenvalue 2. For $r = 1$ we have the second largest eigenvalue $2\cos(2\pi/n) = 2 - \Theta(1/n^2)$, meaning the second smallest eigenvalue of the Laplacian is $\Theta(1/n^2)$.

The expansion of the cycle is $\phi(C_n) \geq 2/n$, and so the cycle is an example in which the second Cheeger inequality is tight.

4 The Hypercube

The group $\{0,1\}^d$ with bitwise xor has 2^d characters; for every $r \in \{0,1\}^d$ there is a character $\chi_r : \{0,1\}^d \to \{-1, 1\}$ defined as

$$\chi_r(x) = (-1)^{\sum_i r_i x_i}$$

Let us denote the set S by $\{e^1, \ldots, e^d\}$, where we let $e^j \in \{0,1\}^d$ denote the bit-vector that has a 1 in the j-th position, and zeroes everywhere else. This means that, for every bit-vector $r \in \{0,1\}^d$, the hypercube has the eigenvalue

$$\sum_j \chi_r(e^j) = \sum_j (-1)^{r_j} = (-|r| + d - |r|) = d - 2\frac{|r|}{d}$$

where we denote by $|r|$ the *weight* of r, that is, the number of ones in r.

Corresponding to $r = (0, \ldots, 0)$, we have the eigenvalue d.

For each of the d vectors r with exactly one 1, we have the second largest eigenvalue$d-$ 2, corresponding to an eigenvalue $2/d$ of the Laplacian.

Let us compute the expansion of the hypercube. Consider "dimension cuts" of the form $S_i := \{x \in \{0,1\}^n : x_i = 0\}$. The set S_i contains half of the vertices, and the number of edges that cross the cut $(S_i, V - S_i)$ is also equal to half the number of vertices (because the edges form a perfect matching), so we have $\phi(S_i) = \frac{1}{d}$.

These calculations show that the first Cheeger inequality $\lambda_2/2 \leq \phi(G)$ is tight for the hypercube.

Finally, we consider the tightness of the approximation analysis of the spectral partitioning algorithm.

We have seen that, in the d-dimensional hypercube, the second eigenvalue has multiplicity d, and that its eigenvectors are vectors $\mathbf{x}^j \in \mathbb{R}^{2^d}$ such that $x^j_a = (-1)^{a_j}$. Consider now the vector $\mathbf{x} := \sum_j \mathbf{x}^j$; this is still clearly an eigenvector of the second eigenvalue. The entries of the vector $\mathbf{x}$ are

$$x_a = \sum_j (-1)^{a_j} = d - 2|a|$$

Suppose now that we apply the spectral partitioning algorithm using $\mathbf{x}$ as our vector. This is equivalent to considering all the cuts $(S_t, V - S_t)$ in the hypercube in which we pick a threshold t and define $S_t := \{a \in \{0,1\}^n : |a| \geq t\}$.

Some calculations with binomial coefficients show that the best such "threshold cut" is the "majority cut" in which we pick $t = n/2$, and that the expansion of $S_{n/2}$ is

$$\phi(S_{n/2}) = \Omega\left(\frac{1}{\sqrt{d}}\right)$$

This gives an example of a graph and of a choice of eigenvector for the second eigenvalue that, given as input to the spectral partitioning algorithm, result in the output of a cut $(S, V-S)$ such that $\phi(S) \geq \Omega(\sqrt{\phi(G)})$. Recall that we proved $\phi(S) \leq 2\sqrt{\phi(G)}$, which is thus tight.

Exercises

1. Find the eigenvalues of a clique with n vertices.

2. Consider the bipartite complete graph $K_{n,n}$ with $2n$ vertices. Express it as a Cayley graph and find its eigenvalues.

3. Show that

$$\phi(G) = \min_{\mathbf{x} \in \mathbb{R}^n, \mathrm{med}(\mathbf{x})=0} \frac{\sum_{\{u,v\} \in E} |x_u - x_v|}{d \sum_v |x_v|}$$

   Hint: use an algorithm similar to the proof of the difficult direction of the Cheeger inequality, but pick t uniformly at random.