Graph Spectrum

Yanheng Wang

A graph can be represented algebraically by an adjacency matrix, an incidence matrix, or perhaps many other forms. Given a matrix representation, we may compute efficiently many algebraic quantities such as rank, determinant, eigenvalues and eigenvectors, and so on. But what for? Can we really retrieve combinatorial information from these quantities?

It turns out that the spectrum (i.e. the set of all eigenvalues) of the adjacency matrix reveals plenty of information (but not all, of course). As we will see, it can predict if a graph is bipartite, bound the maximum degree or even the independence number of a graph!

1 Basic Properties of Spectrum

Let $G=([n],E)$ be a graph on n vertices. Its adjacency matrix, denoted $A:=A(G)$, is defined by $A_{ij}:=\mathbb{1}\{(i,j)\in E\}$. In other words, row i encodes which vertices are adjaecnt to vertex i. It is important to remember that the adjacency matrix representation does depend on the labelling: two isomorphic copies of G can have different-looking adjacency matrices (though after row/column swaps they can be made equal.)

Since A is a real symmetric matrix, it has a complete spectrum. That is, it has n independent eigenvectors $\boldsymbol{v}_1,\dots,\boldsymbol{v}_n$ with eigenvalues $\lambda_1\geqslant\cdots\geqslant\lambda_n$. Moreover, the eigenvectors can be assumed orthonormal if we want.

Note that $\operatorname{tr}(A)=0$ since the diagonal of A is all-zero. Hence $\sum_{i=1}^n\lambda_i=0$, meaning that $\lambda_1>0>\lambda_n$ unless the graph is empty.

Let us compute the spectra of two example graphs:

Example. **(complete graph)**

$A(K_n)=J-I$. Since I accepts any vector as its eigenvector, it suffices to search for eigenvectors of J. Before any computation, we note $\operatorname{rank}(J)=1$, so it has only *one* non-zero eigenvalue. Without much effort we figure out $\mathbf{1}=(1,\dots,1)^{\mathrm{T}}$ as a valid choice, with eigenvalue n. Therefore J has spectrum $n,0,\dots,0$; consequently $J-I$ has spectrum $n-1,-1,\dots,-1$.

Example. **(complete bipartite graph)**

$A(K_{m,n})=\begin{pmatrix} O_{m\times m} & J_{m\times n} \\ J_{n\times m} & O_{n\times n}\end{pmatrix}$. Again notice that $\operatorname{rank}(A)=2$, so there are only two non-zero eigenvalues, i.e. $\lambda_1>(0=\lambda_2=\cdots=\lambda_{n-1})>\lambda_n$. Secondly we know $\lambda_1+\lambda_n=0$. So we need only one additional equation to solve them. Observe that $A^2=\begin{pmatrix} n\,J & O \\ O & m\,J\end{pmatrix}$, so $\operatorname{tr}(A^2)=2mn=\lambda_1^2+\lambda_n^2$. Solving the equations we get $\lambda_1=\sqrt{2mn}$ and $\lambda_n=-\sqrt{2mn}$.

Lemma 1.

- $\Delta(G)\geqslant\lambda_1\geqslant|\lambda_i|$ for all $i\in[n]$. Moreover, $\Delta(G)=\lambda_1$ iff G is Δ-regular.
- Assume G is connected. Then every coordinate of $\boldsymbol{v}_1$ has the same sign and is non-zero. Also $\lambda_1>\lambda_2$.
- Assume G is connected. Then G is bipartite $\iff\lambda_1=-\lambda_n\iff\forall i\in[n],\lambda_i=-\lambda_{n-i+1}$.

To establish this sort of results, we usually look at the quadratic form

$$\boldsymbol{x}\mapsto\frac{\langle A\boldsymbol{x},\boldsymbol{x}\rangle}{\langle\boldsymbol{x},\boldsymbol{x}\rangle}=\frac{\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}}{\|\boldsymbol{x}\|^2}.$$

The intuition is explained below. We regard A as a linear transformation, so the eigenvectors are exactly those who do not change direction after the transformation. Among these, $\boldsymbol{v}_1$ is the one that has largest "boost effect" and $\boldsymbol{v}_n$ is the one that has strongest "negative boost effect".

We decompose a general vector $\boldsymbol{x}$ into $\boldsymbol{x} = \sum_{i=1}^n \alpha_i \boldsymbol{v}_i$, then $A\boldsymbol{x}$ is "the average effect" of tranformation A applied on all eigenvectors, weighted by the coefficients α_i's. If $\boldsymbol{x}$ has "more ingredient" from $\boldsymbol{v}_i$ (i.e. α_i is large) then the result is more aligned to $\boldsymbol{v}_i$.

The quadatic form $\frac{\langle A\boldsymbol{x}, \boldsymbol{x}\rangle}{\langle \boldsymbol{x}, \boldsymbol{x}\rangle}$ measures the proximity of $\boldsymbol{x}$ before and after the transformation (the denominator is just a normalisation that removes the metric of $\boldsymbol{x}$). So quite naturally, it will return a large positive value if $\boldsymbol{x} \approx \boldsymbol{v}_1$, and quite the opposite if $\boldsymbol{x} \approx \boldsymbol{v}_n$. A formal proof is also easy. We have $\langle \boldsymbol{x}, \boldsymbol{x}\rangle = \sum_{i=1}^n \alpha_i^2$ since $\boldsymbol{v}_i$'s are orthonormal basis. We can also compute $\langle A\boldsymbol{x}, \boldsymbol{x}\rangle = \sum_{i=1}^n \lambda_i \alpha_i^2$ easily. Hence

$$\frac{\langle A\boldsymbol{v}_n, \boldsymbol{v}_n\rangle}{\langle \boldsymbol{v}_n, \boldsymbol{v}_n\rangle} = \lambda_n \leqslant \frac{\langle A\boldsymbol{x}, \boldsymbol{x}\rangle}{\langle \boldsymbol{x}, \boldsymbol{x}\rangle} \leqslant \lambda_1 = \frac{\langle A\boldsymbol{v}_1, \boldsymbol{v}_1\rangle}{\langle \boldsymbol{v}_1, \boldsymbol{v}_1\rangle}.$$

So the quadratic form is maximised at $\boldsymbol{v}_1$ and minimised at $\boldsymbol{v}_n$, as expected. It also suggests a variational characterization for λ_1 and λ_n:

$$\lambda_1 = \max_{\boldsymbol{x} \in \mathbb{R}^n} \frac{\langle A\boldsymbol{x}, \boldsymbol{x}\rangle}{\langle \boldsymbol{x}, \boldsymbol{x}\rangle}, \qquad \lambda_n = \min_{\boldsymbol{x} \in \mathbb{R}^n} \frac{\langle A\boldsymbol{x}, \boldsymbol{x}\rangle}{\langle \boldsymbol{x}, \boldsymbol{x}\rangle}.$$

Can we express the other eigenvalues in a similar way? The answer is yes, and the form turns out to be neat:

Lemma 2. Suppose a real symmetric matrix A has eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_n$. Then

$$\lambda_i = \max_{\substack{U \subseteq \mathbb{R}^n \\ \dim U = i}} \min_{\boldsymbol{x} \in U} \frac{\langle A\boldsymbol{x}, \boldsymbol{x}\rangle}{\langle \boldsymbol{x}, \boldsymbol{x}\rangle} = \min_{\substack{U \subseteq \mathbb{R}^n \\ \dim U = n-i+1}} \max_{\boldsymbol{x} \in U} \frac{\langle A\boldsymbol{x}, \boldsymbol{x}\rangle}{\langle \boldsymbol{x}, \boldsymbol{x}\rangle}.$$

Proof. We prove the first equation only; the second one is similar.

($\geqslant$) We intend to show: for all $U \subseteq \mathbb{R}^n : \dim U = i$, there exists $\boldsymbol{x} \in U$ such that $\lambda_i \geqslant \frac{\langle A\boldsymbol{x}, \boldsymbol{x}\rangle}{\langle \boldsymbol{x}, \boldsymbol{x}\rangle}$. Let us pause for a minute and consider how to construct such an $\boldsymbol{x}$. If it is made of ingredients $\boldsymbol{v}_i, \ldots, \boldsymbol{v}_n$ only, then we can guarantee that the quadratic form has smaller value. So we naturally construct a apace $V := \operatorname{span}(v_i, \ldots, v_n)$. Note that $\dim V = n - i + 1$, hence $\dim(U \cap V) \geqslant 1$. So there indeed exists $\boldsymbol{x} \in U \cap V$, which by definition satisfies

$$\frac{\langle A\boldsymbol{x}, \boldsymbol{x}\rangle}{\langle \boldsymbol{x}, \boldsymbol{x}\rangle} = \frac{\sum_{j=i}^n \lambda_j \alpha_j^2}{\sum_{j=i}^n \alpha_j^2} \leqslant \lambda_i.$$

($\leqslant$) We intend to show: there exists $U \subseteq \mathbb{R}^n : \dim U = i$ such that $\lambda_i \leqslant \frac{\langle A\boldsymbol{x}, \boldsymbol{x}\rangle}{\langle \boldsymbol{x}, \boldsymbol{x}\rangle}$ for all $\boldsymbol{x} \in U$. Now it should be clear that $U := \operatorname{span}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_i)$ does the job. □

After grabbing the intuition, we return to prove Lemma 1.

Proof of Lemma 1.

- Suppose $\boldsymbol{v}_1 = (v_{11}, \ldots, v_{1n})^{\mathrm{T}}$ where v_{1i} is the largest coordinate. Assume without loss of generality that $v_{1i} > 0$. Then

$$\lambda_1 v_{1i} = (A\boldsymbol{v}_1)_i = \sum_{j=1}^n a_{ij} v_{1j} \leqslant v_{1i} \sum_{j=1}^n a_{ij} \leqslant v_{1i} \Delta,$$

that is $\lambda_1 \leqslant \Delta$. It becomes an equality iff $v_{11} = \cdots = v_{1n}$. This happens iff the graph is Δ-regular. (Exercise)

To prove the other inequality, we use the quadratic form and triangle inequality:

$$|\lambda_i| = |\langle A\boldsymbol{v}_i, \boldsymbol{v}_i\rangle| \leqslant \langle A|\boldsymbol{v}_i|, |\boldsymbol{v}_i|\rangle \leqslant \lambda_1.$$

- Define a unit vector $\boldsymbol{\varphi} := |\boldsymbol{v}_1|$ and compute $\langle A\boldsymbol{\varphi}, \boldsymbol{\varphi}\rangle \geqslant |\langle A\boldsymbol{v}_1, \boldsymbol{v}_1\rangle| = \lambda_1$. If $\boldsymbol{v}_1$ does contain contradicting signs, then the $\geqslant$ will become $>$, which is impossible.

  Now we could safely assume $\boldsymbol{v}_1 \geqslant \boldsymbol{0}$. We shall show $v_{1i} > 0$ for all $i \in [n]$. Suppose to the contrary that $v_{1i} = 0$ for some $i \in [n]$. Then

$$0 = (A\boldsymbol{v}_1)_i = \sum_{j=1}^{n} a_{ij} v_{1j} = \sum_{j \sim i} v_{1j}.$$

  Since $v_{1j} \geqslant 0$, we have no choice but $\forall j \sim i, v_{1j} = 0$. This argument then propogates to *all* vertices in the graph due to connectivity. Hence $\boldsymbol{v}_i = \boldsymbol{0}$, a contradiction.

  Finally, we prove $\lambda_1 > \lambda_2$ by contradiction. Suppose $\lambda_1 = \lambda_2$, then any linear combination of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ is also an eigenvector of λ_1. So we could take for instance $\boldsymbol{v}_1 - \frac{v_{1i}}{v_{2i}}\boldsymbol{v}_2$ where $v_{2i} \neq 0$. This would zero the coordinate i, contradicting the previous claim.

- ($\Rightarrow$) Let us observe the pattern of the adjacency matrix A of a bipartite graph. Since relabelling the vertices shall not change the spectrum, we may assume without loss of generality that

$$A = \begin{pmatrix} O_{m\times m} & B_{m\times n} \\ B^{\mathrm{T}}_{n\times m} & O_{n\times n} \end{pmatrix}.$$

  We claim the following: If $\boldsymbol{x} = (\boldsymbol{x}^+, \boldsymbol{x}^-)$ is an eigenvector of A with eigenvalue λ, then $\boldsymbol{y} = (\boldsymbol{y}^+, \boldsymbol{y}^-) := (\boldsymbol{x}^+, -\boldsymbol{x}^-)$ is an eigenvector of A with eigenvalue $-\lambda$.

  The proof is just by definition. Since $A\boldsymbol{x} = \lambda\boldsymbol{x}$, we know $B\boldsymbol{x}^- = \lambda\boldsymbol{x}^+$ and $B^{\mathrm{T}}\boldsymbol{x}^+ = \lambda\boldsymbol{x}^-$. Hence

$$A\boldsymbol{y} = \begin{pmatrix} B\boldsymbol{y}^- \\ B^{\mathrm{T}}\boldsymbol{y}^+ \end{pmatrix} = \begin{pmatrix} -B\boldsymbol{x}^- \\ B^{\mathrm{T}}\boldsymbol{x}^+ \end{pmatrix} = \begin{pmatrix} -\lambda\boldsymbol{x}^+ \\ \lambda\boldsymbol{x}^- \end{pmatrix} = -\lambda\boldsymbol{y}.$$

  This claim implies that the spectrum of a bipartite graph is symmetric with respect to 0.

  ($\Leftarrow$) Consider the matrix A^k for odd number k. The diagonal element $(A^k)_{i,i}$ counts the number of walks of length k that both starts and ends at i. Then $\mathrm{tr}(A^k) > 0$ iff there is a cycle C_k in the graph.

  Since the spectrum of A is symmetric, the spectrum of A^k is also symmetric for any odd number k. Therefore, $\mathrm{tr}(A^k) = \sum_{i=1}^{n} \lambda_i(A^k) = 0$. So the graph doesn't contain odd cycle, hence bipartite. □

Lemma 3. $\mathrm{diam}(G) := \max_{u \rightsquigarrow v} d(u, v)$ is less than the number of distinct eigenvalues of $A(G)$.

Proof. Let $\mu_1, \ldots, \mu_d$ be distinct eigenvalues of $A(G)$. To facilitate our proof, we define a polynomial $m(x) := \prod_{i=1}^{d}(x - \mu_i)$. It is essentially the characteristic polynomial of $A(G)$ without multiplicities. We claim that $m(A) = O$.

To see the claim, we diagonalise $A = Q^{\mathrm{T}}\Lambda Q$ where Q is orthonormal and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ with $\{\lambda_1, \ldots, \lambda_n\} = \{\mu_1, \ldots, \mu_d\}$. Observe that $m(\Lambda) = \prod_{i=1}^{d}(\Lambda - \mu_i I) = O$. (The product of diagonal matrices is the same as elementwise product!) Therefore $m(A) = m(Q^{\mathrm{T}}\Lambda Q) = Q^{\mathrm{T}} m(\Lambda) Q = O$.

By moving terms around, the claim implies $A^d = \sum_{i=1}^{d-1} c_i A^i$ for some coefficients c_i's. Hence, for any $t \geqslant d$, we may prove by induction that $A^t = \sum_{i=1}^{d-1} d_i A^i$ for some coefficients d_i's.

Now we assume, for the sake of contradiction, that $\operatorname{diam}(G) \geqslant d$. Then there exist connected vertices u, v with $d(u,v) = \operatorname{diam}(G) \geqslant d$. So $A^t(u,v) = 0$ for all $t \leqslant d-1$, which implies that $A^t(u,v) = 0$ for all $t \in \mathbb{N}$. This contradicts with connectivity between u and v. □

2 Relating Spectrums of Two Graphs

Although efficient algorithms can compute the spectrum of any *specific* adjacency matrix, it is often the case that we want to study a certain *class* of graphs. For this purpose, hand-calculation is the only option, and the tools in this section may be helpful.

Lemma 4. If G is a d-regular graph with spectrum $\lambda_1, \ldots, \lambda_n$, then its complement graph $\overline{G}$ has spectrum $n-1-\lambda_1, -(1+\lambda_2), \ldots, -(1+\lambda_n)$.

Proof. Note that $A(\overline{G}) = J - I - A(G)$. Assume $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are the eigenvectors of $A(G)$. Since G is d-regular, we know $\boldsymbol{v}_1 = \mathbf{1}$ and all other eigenvectors are orthogonal to it. Therefore

$$\begin{aligned} A(\overline{G})\,\boldsymbol{v}_1 &= n\,\boldsymbol{v}_1 - \boldsymbol{v}_1 - \lambda_1\,\boldsymbol{v}_1 = (n-1-\lambda_1)\,\boldsymbol{v}_1 \\ A(\overline{G})\,\boldsymbol{v}_i &= \mathbf{0} - \boldsymbol{v}_i - \lambda_1\,\boldsymbol{v}_i = -(1+\lambda_i)\,\boldsymbol{v}_i. \end{aligned}$$

So these eigenvectors form an eigenbasis for $A(\overline{G})$, with corresponding eigenvalues $n-1-\lambda_1$, $-(1+\lambda_2), \ldots, -(1+\lambda_n)$. □

Before getting into the next tool, we recall some definitions and facts:

- The *incidence matrix* $B := B(G)$ is defined by $B_{ij} := \mathbb{1}\{i \in e_j\}$. (We assumed an order of the edges.) Note that column j encodes the vertices that e_j contains; row i encodes the edges that contains i.
- The line graph of G is defined by $L(G) := \left(E, \left\{\{e,f\} \in \binom{E}{2} : e \cap f \neq \emptyset\right\}\right)$.
- The matrices $C^{\mathrm{T}}C$ and CC^{T} share the same set of non-zero eigenvalues. The proof is simple: Let $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_t$ be independent eigenvectors of $C^{\mathrm{T}}C$ corresponding to non-zero eigenvalues $\lambda_1, \ldots, \lambda_t$. Then $CC^{\mathrm{T}}(C\boldsymbol{\gamma}_i) = C(C^{\mathrm{T}}C\boldsymbol{\gamma}_i) = C\lambda_i\boldsymbol{\gamma}_i = \lambda_i(C\boldsymbol{\gamma}_i)$, so $C\boldsymbol{\gamma}_i$ is an eigenvector of CC^{T}. Moreover, the vectors $C\boldsymbol{\gamma}_i$'s are linearly independent for $i \in [t]$ (by checking definition of linear independence).

Lemma 5. If $G = (V, E)$ is a d-regular graph with spectrum $\lambda_1, \ldots, \lambda_t, -d, \ldots, -d$, then its line graph $L(G)$ has spectrum $\lambda_1 + d - 2, \ldots, \lambda_t + d - 2, -2, \ldots, -2$.

Proof. Let us first see what the matrices $B^{\mathrm{T}}B$ and BB^{T} mean.

The matrix $B^{\mathrm{T}}B$ is formed by taking inner product between columns of B. Namely $(B^{\mathrm{T}}B)_{ij}$ is the inner product of columns i and j of B. With our experience in set systems, this means

$$(B^{\mathrm{T}}B)_{ij} = |e_i \cap e_j| = \begin{cases} 2 & i = j \\ 1 & \{e_i, e_j\} \in L(G) \\ 0 & \text{otherwise.} \end{cases}$$

Therefore $A(L(G)) = B^{\mathrm{T}}B - 2I$.

The matrix BB^{T} is formed by taking inner product between rows of B. Namely $(BB^{\mathrm{T}})_{ij}$ is the inner product of columns i and j of B. Hence

$$(BB^{\mathrm{T}})_{ij} = \#\text{edges containing both } i \text{ and } j = \begin{cases} d & i = j \\ 1 & \{i,j\} \in E \\ 0 & \text{otherwise.} \end{cases}$$

Therefore $A(G) = BB^{\mathrm{T}} - dI$.

Now we combine the two parts and see that $A(L(G)) + 2I$ and $A(G) + dI$ have identical non-zero eigenvalues. The lemma then follows by distinguishing zeros and non-zeros. □

Our final tool relates the spectra of a graph and its induced subgraphs.

Theorem 6. (Cauchy interlacing theorem)

Let $A \in \mathbb{R}^{n\times n}$ be a real symmetric matrix with eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_n$. Let $B \in \mathbb{R}^{m\times m}$ be a principal submatrix (i.e. selecting same columns and rows in A) with eigenvalues $\mu_1 \geqslant \cdots \geqslant \mu_m$. Then we have $\lambda_i \geqslant \mu_i \geqslant \lambda_{i+n-m}$.

Proof. This is an easy consequence of the variational characterization of eigenvalues. Assume without loss of generality that B resides in the top-left corner of A, namely

$$A = \begin{pmatrix} B & C \\ C^{\mathrm{T}} & D \end{pmatrix}.$$

For any vector $\boldsymbol{x} \in \mathbb{R}^m$, we denote its extension $\boldsymbol{x}' := \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{0} \end{pmatrix} \in \mathbb{R}^n$. Note that $\langle A\boldsymbol{x}', \boldsymbol{x}'\rangle = \langle B\boldsymbol{x}, \boldsymbol{x}\rangle$. From Lemma 2,

$$\mu_i = \max_{\substack{U\subseteq\mathbb{R}^m \\ \dim U = i}} \min_{\boldsymbol{x}\in U} \frac{\langle A\boldsymbol{x}', \boldsymbol{x}'\rangle}{\langle \boldsymbol{x}', \boldsymbol{x}'\rangle} = \max_{\substack{V\subseteq\mathbb{R}^m\times\{0\}^{n-m} \\ \dim V = i}} \min_{\boldsymbol{y}\in V} \frac{\langle A\boldsymbol{y}, \boldsymbol{y}\rangle}{\langle \boldsymbol{y}, \boldsymbol{y}\rangle} \leqslant \lambda_i$$

where the inequality follows because μ_i optimises over a smaller region than λ_i does. Similarly,

$$\mu_i = \min_{\substack{U\subseteq\mathbb{R}^m \\ \dim U = m-i+1}} \max_{\boldsymbol{x}\in U} \frac{\langle A\boldsymbol{x}', \boldsymbol{x}'\rangle}{\langle \boldsymbol{x}', \boldsymbol{x}'\rangle} = \max_{\substack{V\subseteq\mathbb{R}^m\times\{0\}^{n-m} \\ \dim V = m-i+1}} \min_{\boldsymbol{y}\in V} \frac{\langle A\boldsymbol{y}, \boldsymbol{y}\rangle}{\langle \boldsymbol{y}, \boldsymbol{y}\rangle} \geqslant \lambda_{i+n-m}.$$ □

Corollary 7. Let G be a graph on n vertices and $H \subseteq G$ be an induced subgraph on m vertices. Then $\lambda_i(G) \geqslant \lambda_i(H) \geqslant \lambda_{i+n-m}(G)$ for all $i \in [m]$.

3 Applications

3.1 Graph decomposition

The Peterson graph is an important toy graph on 10 vertices. It is 3-regular and thus has 15 edges. The complete graph K_{10} is 9-regular and happens to have 45 edges – exactly threefold the Peterson graph. So the question is: can we decompose K_{10} into three disjoint copies of Peterson graph?

This seemingly harmless question entails a very large search space actually: A brute-force approach needs at least $\binom{45}{15}$ enumerations – far too much for even a decent personal computer!

But with simple algebraic method the question is falsified by Schwenk. First, observe that the decomposition is modelled by $J - I = P + Q + R$ where P, Q, R are adjacency matrices of copies of the Peterson graph. P, Q, R have the same spectrum, but the eigenspaces could be different.

Next we note that Peterson graph is exactly $\overline{L(K_5)}$. We may then compute by hand, using tools in previous section, its spectrum $3, 1, 1, 1, 1, 1, -2, -2, -2, -2$, where the eigenvalue 3 always corresponds to eigenvector $\mathbf{1}$.

Let U be the eigenspace of eigenvalue 1 in P. Similarly let V be the eigenspace of eigenvalue 1 in Q. Then $\dim(U) = \dim(V) = 5$. In addition, $(U+V) \perp \{\mathbf{1}\}$, so $\dim(U+V) \leqslant 10 - 1 = 9$. This implies $\dim(U \cap V) \geqslant 1$, so there does exist an eigenvector $\boldsymbol{v} \in U \cap V$ *shared* by P and Q. Multiplying the decomposition equation by $\boldsymbol{v}$ gives

$$\mathbf{0} - \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{v} + R\,\boldsymbol{v}$$

meaning that $R\,\boldsymbol{v} = 3\boldsymbol{v}$. So 3 is an eigenvalue of R, contradicting the spectrum we computed.

Remark. It is a general technique to express "graph decomposition" as an adjacency matrix equation. Consider for instance a different graph decomposition problem. We want to cover the complete graph K_n by some (not necessarily disjoint) complete bipartite subgraphs. Each edge should be covered an odd number of times. How many subgraphs do we need?

We can again model the problem by $J - I = \sum_{i=1}^{m} M_i$ where M_i's are the adjacency matrices of complete bipartite subgraphs. Here we work in the finite field $\mathbb{Z}_2$. As we know $\operatorname{rank}(M_i) = 2$, so

$$n - 1 = \operatorname{rank}(J - I) = \operatorname{rank}\left(\sum_{i=1}^{m} M_i\right) \leqslant \sum_{i=1}^{m} \operatorname{rank}(M_i) = 2m$$

and $m \geqslant \frac{n-1}{2}$. We remind the reader that $\operatorname{rank}(J - I) = n - 1$ instead of n since we work in $\mathbb{Z}_2$.

3.2 Independent sets, cuts, and expanders

In many contexts, we want to associate each vertex v with a weight x_v and measure smoothness by

$$\sum_{u \sim v} (x_u - x_v)^2 = d \sum_{v \in V} x_v^2 - 2 \sum_{u \sim v} x_u\, x_v.$$

We will see three important applications of this kind later.

Observe that it is a quadratic form $\langle B\boldsymbol{x}, \boldsymbol{x} \rangle$ with matrix $B = I - A$. Hence it equals

$$d\langle \boldsymbol{x}, \boldsymbol{x} \rangle - \langle A\boldsymbol{x}, \boldsymbol{x} \rangle$$

and the eigenvalues naturally come into play. We already know that $\lambda_n \leqslant \frac{\langle A\boldsymbol{x}, \boldsymbol{x} \rangle}{\langle \boldsymbol{x}, \boldsymbol{x} \rangle} \leqslant \lambda_1$, so we immediately have bounds

$$(d - \lambda_1)\, \|\boldsymbol{x}\|^2 \leqslant \sum_{u \sim v} (x_u - x_v)^2 \leqslant (d - \lambda_n)\, \|\boldsymbol{x}\|^2.$$

The RHS looks useful, but no so for the LHS. We know $\lambda_1 = d$, so we have proven the big fact that 0 is smaller than a positive number.

But we can fix it by replacing λ_1 with λ_2, provided $\boldsymbol{x} \perp \mathbf{1} = \boldsymbol{v}_1$. Under this condition $\frac{\langle A\boldsymbol{x}, \boldsymbol{x} \rangle}{\langle \boldsymbol{x}, \boldsymbol{x} \rangle} \leqslant \lambda_2$ because the $\boldsymbol{v}_1$ component of $\boldsymbol{x}$ is literally zero.

Maximum cut

The foregoing bounds can directly applied to bound the size of cuts. Let (S,T) be a cut of the graph, whose size $e(S,T)$ is the number of edges between S and T. To account for this, we had better specify uniform weights on S and on T respectively, so the internal edges contribute nothing:

$$x_v := \begin{cases} n-s & v \in S \\ -s & v \in T \end{cases}.$$

where $s := |S|$ and the weights are chosen to get optimal result. Then we plug this into our bound and deduce

$$n^2 \cdot e(S,T) \leqslant (d-\lambda_n)\,(s\,(n-s)^2 + (n-s)\,s^2).$$

After simplification we get $e(S,T) \leqslant (d-\lambda_n)\,s\,(n-s)\,/\,n \leqslant (d-\lambda_n)\,n\,/\,4$.

For bipartite graphs $\lambda_n = -d$, hence the upper bound is $nd/2$, which is the number of edges. Indeed we cannot hope for better since all edges in a bipartite graph can involve in a single cut.

Independence number

Let $S \subseteq V$ be an independent set and let $s := |S|$. We use essentially the same weight function:

$$x_v := \begin{cases} n-s & v \in S \\ -s & v \notin S \end{cases}$$

and find $e(S,\overline{S}) \leqslant (d-\lambda_n)\,s\,(n-s)\,/\,n$. But by definition S does not contain internal edges, so all edges related to S must go between S and $\overline{S}$. Hence $d \cdot s = e(S,\overline{S})$. Putting back into the inequality, we yield $s \leqslant \frac{-\lambda_n}{d-\lambda_n}\,n$.

Needless to say, this bound is valuable since it is hard to compute or even reasonably approximate the independence number, while on the other hand computing λ_n is a piece of cake.

Expanders

An *expander* is a graph with very strong connectivity. One possible definition: If for all $S \subseteq V$ with $|S| \leqslant n/2$ we have $e(S,\overline{S}) \geqslant \delta\,|S|$, then we call the graph a δ-expander. You should immediately note its practical significance: expanders are very stable since it takes much effort to "cut off a large piece".

Given a constant $\delta > 0$, constructing an infinite family of δ-expanders is highly-nontrivial. Literally all explicit constructions known today make use of the spectral inequality we saw just now.

Once more, we use the foregoing weight function. Notice that $\boldsymbol{x} \perp \mathbf{1}$ indeed, so we derive $e(S,\overline{S}) \geqslant (d-\lambda_2)\,s\,(n-s)\,/\,n$ where $s := |S| \geqslant \frac{n}{2}$. Since $n - s \geqslant n/2$, we deduce $e(S,\overline{S}) \geqslant \frac{d-\lambda_2}{2}\,s$. Hence, the problem reduces to constructing a family of graphs for which the spectral gap $d - \lambda_2 \geqslant 2\,\delta$. (For a single graph the task is trivial, but for a family of graphs with $n \to \infty$, this is quite tough still.) We will not introduce any detailed construction here.

3.3 Sensitivity conjecture

As one of the most exciting progresses of theoretical computer science in the past decade, the long-standing "sensitivity conjecture" is resolved by a fantastically simple proof via graph spectrum.

Theorem 8. (Huang) Any induced subgraph G of the Hamming cube $H_n := \{0,1\}^n$ containing more than 2^{n-1} vertices satisfies $\Delta(G) \geqslant \sqrt{n}$.

Note that if G contains $\leqslant 2^{n-1}$ vertices then $\Delta(G)$ could be as small as 0, since H_n has equal bipartition. With only one additional vertex, the maximum degree jumps from 0 to at least $\sqrt{n}$.

Proof. Let $G \subseteq H_n$ be any induced subgraph containing $m \geqslant 2^{n-1}+1$ vertices. Recall that $\Delta(G) \geqslant \lambda_1(G)$. On the other hand, Corollary 7 asserts $\lambda_1(G) \geqslant \lambda_{1+2^n-m}(H_n) \geqslant \lambda_{2^{n-1}}(H_n)$. So if we could show $\lambda_{2^{n-1}}(H_n) \geqslant \sqrt{n}$ then we are done.

Unfortunately the spectrum of H_n is spread over the interval $[n, -n]$ and the very middle eigenvalue $\lambda_{2^{n-1}}(H_n)$ is quite close to 0. So our preliminary approach does not work.

But here comes the ingenious *signing trick*: Instead of working directly on the adjacency matrices, we put some minus signs on their entries. Then we apply the arguments above to the signed version and see everything works.

Now we describe the signing. Inductively define

$$B_1 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad B_{i+1} := \begin{pmatrix} B_i & I \\ I & -B_i \end{pmatrix}.$$

Note that the matrix without signs, $|B_n|$, is exactly the adjacency matrix of H_n. Also observe

$$B_{i+1}^2 = \begin{pmatrix} B_i^2 + I & O \\ O & B_i^2 + I \end{pmatrix},$$

so by easy induction we see $B_n^2 = nI$, and all its eigenvalues are n. Hence the eigenvalues of matrix B_n can only be $\pm\sqrt{n}$. Since $\operatorname{tr}(B_n) = 0$, the two possibilities are 50-50, thus $\lambda_{2^{n-1}}(B_n) = \sqrt{n}$.

Now we translate our old argument. Let C be the principal submatrix of B_n that corresponds to G; note again $|C|$ is the adjacency matrix of G. We have $\lambda_1(C) \geqslant \lambda_{2^{n-1}}(B_n) \geqslant \sqrt{n}$. So it remains to show $\Delta(G) \geqslant \lambda_1(C)$. To see this, we modify the proof of Lemma 1 to remove the signing. Let $\boldsymbol{v}_1 = (v_{11}, \ldots, v_{1n})^{\mathrm{T}}$ be the eigenvector corresponding to $\lambda_1(C)$, with largest coordinate $v_{1i} > 0$. Then

$$|\lambda_1(C)\, v_{1i}| = |C\boldsymbol{v}_1|_i = \left| \sum_{j=1}^{n} c_{ij}\, v_{1j} \right| \leqslant \sum_{j=1}^{n} |c_{ij}|\, v_{1j} \leqslant v_{1i} \sum_{j=1}^{n} |c_{ij}| \leqslant v_{1i}\, \Delta(G),$$

hence $|\lambda_1(C)| \leqslant \Delta(G)$ and in particular $\lambda_1(C) \leqslant \Delta(G)$. □

4 Not Only Graphs

The use of eigenvalues/eigenvectors finds its way in a broader context than graph theory. In the theory of Markov chains, for example, there exists a beautiful connection between the spectral gap $1 - \lambda_2$ of a transition matrix and the mixing time of the corresponding chain.

In what follows we give an application of eigenvalues (and more broadly, linear algebra) in coding theory. A binary code $C \subseteq \{-1, 1\}^k$ is called ε-balanced if the Hamming distance between any two codewords is between $\left[\frac{1-\varepsilon}{2}k, \frac{1+\varepsilon}{2}k\right]$; equivalently, if $|\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \leqslant \varepsilon k$ for all distinct $\boldsymbol{u}, \boldsymbol{v} \in C$.

The main goal of coding theory is to come up with a code C such that

- the Hamming distance between any two codewords is large;
- the number of codewords is relatively large, that is $|C|/2^k$ should be non-negligible.

Clearly an ε-balanced code has large separation between codewords, which is desirable. But on the other hand, its rather restrictive definition might harm the ratio $|C|/2^k$. Via algebraic methods, we could probe its limitations before we ever try to construct an ε-balanced code.

Theorem 9. For any $\varepsilon \in \left[\frac{1}{\sqrt{k}}, \frac{1}{2}\right]$ and ε-balanced code $C \subseteq \{-1,1\}^k$, it holds that $n := |C| \leqslant 2^{6\epsilon\epsilon^2 \log(1/\epsilon)k}$.

By definition of ε-balanced codes, the Gram matrix $A := C^{\mathrm{T}}C$ has a clear-cut structure: The diagonal entries are all k, while the remaining entries have absolute values at most εk. So the following lemma immediately tells us that $\mathrm{rank}(A)$ is large.

Lemma 10. For any square matrix A we have $\mathrm{rank}(A) \geqslant \frac{\mathrm{tr}^2(A)}{\mathrm{tr}(A^2)} = \frac{\mathrm{tr}^2(A)}{\sum_i \sum_j a_{ij}^2}$. In simple words, a matrix has large rank if the mass is concentrated at its diagonal.

Proof. Denote $r := \mathrm{rank}(A)$ and suppose $\lambda_1, \ldots, \lambda_r$ are the non-zero eigenvalues of A. Then by Cauchy-Schwarz inequality we have

$$\mathrm{tr}(A^2) = \sum_{i=1}^{r} \lambda_i^2 = \frac{1}{r}\left(\sum_{i=1}^{r} \lambda_i^2\right)\left(\sum_{i=1}^{r} 1^2\right) \geq \frac{1}{r}\left(\sum_{i=1}^{r} \lambda_i\right)^2 = \frac{\mathrm{tr}^2(A)}{r}. \qquad \square$$

But on the other hand $k \geqslant \mathrm{rank}(C) = \mathrm{rank}(A)$. Linking with the lower bound of $\mathrm{rank}(A)$ gives us some reasonable upper bound for m in terms of k.

To refine the result, we study $A^{(\ell)}$ instead of A to "boost" the separation between diagonal and non-diagonal, and then repeat the same argument. Here $A^{(\ell)}$ denotes the coordinate-wise ℓ-th power of matrix A.

Lemma 11. For any square matrix A and $\ell \in \mathbb{N}$, we have $\mathrm{rank}(A^{(\ell)}) \leqslant \binom{\ell + \mathrm{rank}(A) - 1}{k}$.

Proof. For vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ of the same dimension, we denote their coordinate-wise product as $\boldsymbol{x} \odot \boldsymbol{y}$. Naturally, we denote $\boldsymbol{x}^{\ell} := \boldsymbol{x} \odot \cdots \odot \boldsymbol{x}$ (k many times).

Denote $r := \mathrm{rank}(A)$. Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ be the column vectors of matrix A, where we ordered them appropriately so that $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$ constitute a basis of the column space. Hence for any $i \in [n]$ we could find coefficients $\beta_{i1}, \ldots, \beta_{ir}$ such that $\boldsymbol{v}_i = \sum_{j=1}^{r} \beta_{ij} \boldsymbol{v}_j$. Lifted to ℓ-th power:

$$\boldsymbol{v}_i^{\ell} = \left(\sum_{j=1}^{r} \beta_{ij} \boldsymbol{v}_j\right)^{\ell}.$$

Consider the set of vectors $\mathcal{S} := \{\boldsymbol{v}_1^{\ell_1} \odot \cdots \odot \boldsymbol{v}_r^{\ell_r} \colon \ell_1, \ldots, \ell_r \in \mathbb{N}_0, \sum_{j=1}^{r} \ell_j = \ell\}$. Clearly the vectors $\boldsymbol{v}_1^{\ell}, \ldots, \boldsymbol{v}_n^{\ell}$ are spanned by $\mathcal{S}$ due to our derivation. Therefore

$$\mathrm{rank}(A^{(\ell)}) = \mathrm{rank}(\boldsymbol{v}_1^{\ell}, \ldots, \boldsymbol{v}_n^{\ell}) \leqslant |\mathcal{S}| = \binom{\ell + r - 1}{r - 1} = \binom{\ell + r - 1}{\ell}. \qquad \square$$

Now we can assemble our final argument.

Proof of Theorem 9. Using the two lemmas, we have

$$\frac{n}{1 + (n-1)\,\epsilon^{2\ell}} \leqslant \frac{(\sum_i a_{ii}^{\ell})^2}{\sum_i \sum_j a_{ij}^{2\ell}} \leqslant \mathrm{rank}(A^{(\ell)}) \leqslant \binom{\ell + \mathrm{rank}(A) - 1}{\ell} \leqslant \binom{\ell + k - 1}{\ell}.$$

We choose $\ell := \frac{\log n}{2\log(1/\epsilon)}$ so that $\frac{n}{2} \leqslant \frac{n}{1+(n-1)\,\epsilon^{2\ell}}$. Hence

$$n \leqslant 2\binom{\ell+k-1}{\ell} \leqslant 2\left(\frac{e\,(\ell+k-1)}{\ell}\right)^{\ell} < 2\left(e+\frac{e\,k}{\ell}\right)^{\ell}.$$

Taking logarithm and dividing both sides by $\log n$ gives

$$1 \leqslant \frac{1}{\log n} + \frac{\log(e+e\,k/\ell)}{2\log(1/\epsilon)}.$$

When n is sufficiently large, we may neglect the $\frac{1}{\log n}$ term (this could be made formal if you like.) Then $2\log(1/\epsilon) \leq \log(e+e\,k/\ell)$, or simply $\frac{k}{\ell} \geq \frac{1}{e\,\epsilon^2} - 1 \geq \frac{3}{e\,\epsilon^2}$ when $\epsilon \leq 1/2$. Therefore, $\ell \leq 3\,e\,\epsilon^2\,k$ and consequently $n \leq 2^{6e\epsilon^2\log(1/\epsilon)k}$ as desired. □