Introduction to spectral graph theory

Uriel Feige

May 29, 2022

1 Background

With every graph (or directed graph) one can associate several different matrices. Here we shall concentrate mainly on the **adjacency matrix** of (undirected) graphs, and also discuss briefly the Laplacian. We shall show that spectral properties (the eigenvalues and eigenvectors) of these matrices provide useful information about the structure of the graph. It turns out that for regular graphs, the information one can deduce from one matrix representation (e.g., the adjacency matrix) is similar to the information one can deduce from other representations (such as the Laplacian). We remark that for nonregular graphs, this is not the case, and the choice of matrix representation may make a significant difference. Our main concern here will be either with regular or nearly regular graphs.

The adjacency matrix of a connected undirected graph is nonnegative, symmetric and irreducible (namely, it cannot be decomposed into two diagonal blocks and two off-diagonal blocks, one of which is all-0). As such, standard results in linear algebra, including the **Perron-Frobenius** theorem, imply that:

1. All its eigenvalues are real. Let us denote them by $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_n$. (Equality between eigenvalues corresponds to eigenvalue multiplicity.)
2. Eigenvectors that correspond to different eigenvalues are orthogonal to each other.
3. The eigenvector that corresponds to λ_1 is all positive.
4. $\lambda_1 > \lambda_2$ and $\lambda_1 \geq |\lambda_n|$.

An eigenvector with eigenvalue λ can be interpreted as associating values (the coordinate entries of the eigenvalue) with the vertices of the graph, such that each vertex has a value that is the sum of its neighbors, scaled by $1/\lambda$.

Observe that the trace of the adjacency matrix is 0 and hence $\sum \lambda_i = 0$, implying in particular that $\lambda_n < 0$.

There is useful characterization of eigenvalues by **Raleigh quotients**. Let $v_1, \ldots, v_n$ be an orthonormal basis of eigenvectors. For a nonzero (column) vector x, let $a_i = x^t \cdot v_i$ (inner product, where superscript t denotes transpose) and hence $x = \sum a_i v_i$. Observe that:

$$\frac{x^t A x}{x^t x} = \frac{\sum \lambda_i (a_i)^2}{\sum (a_i)^2}$$

This implies that $\lambda_n \leq \frac{x^t Ax}{x^t x} \leq \lambda_1$. Moreover, if x is orthogonal to v_1 then $\frac{x^t Ax}{x^t x} \leq \lambda_2$.

It is useful to keep in mind that $x^t Ay = \sum_{i,j} x_i A_{i,j} y_j = \sum_{i,j} x_i y_j A_{ij} = \sum_{i,j} Z_{i,j} A_{i,j}$, where Z is the matrix $Z = xy^t$.

For a d-regular graph, $\lambda_1 = d$, and the corresponding eigenvector is the all 1 vector, which we denote by $\bar{1}$. If the graph is disconnected, then its adjacency matrix decomposes into adjacency matrices of its connected components. In this case (and again, we assume here d-regularity), the multiplicity of the eigenvalue d is equal to the number of connected components. This indicates that a small gap between λ_1 and λ_2 corresponds to the graph having small cuts. There is also a converse, saying that a large gap between λ_1 and λ_2 implies that the graph has no small cuts (small compared to the number of verices cut). Namely, it is an **expander**.

1.1 Mixing time of random walks

To understand intuitively the relation between graph expansion and eigenvalue gap (in fact, we shall consider here the gap between λ_1 and $\max[\lambda_2, -\lambda_n]$), consider a random walk on the d-regular graph G, starting at an arbitrary vertex v. We ask how long it takes the random walk to *mix*, meaning that it is roughly equally likely to be at any vertex. Clearly, at least $\Omega(\log_d(n))$ steps are required, as only at most $d(d-1)^k$ vertices are reachable in k steps. For constant d, this lower bound is $\Omega(\log n)$. However, $O(\log n)$ might not suffice if the graph has sets of vertices with relatively few outgoing edges (referred to as *nonexpanding sets*), as the walk might get stuck at a nonexpanding set for many steps. Hence if the walk mixes in $O(\log n)$ steps (regardless of the starting vertex), this means that all subsets of vertices are expanding, and the graph is an expander. (We keep the discussion here at an intuitive level, and omit formal definitions of the expansion parameters involved.)

We now explain how to analyse the mixing time of the random walk. The initial starting vertex can be viewed as an initial distribution $x^{(0)}$ over vertices, where this distribution is supported only on one vertex. Hence the vector $x^{(0)}$ has one 1 entry, and its remaining entries are 0. The uniform distribution over vertices is $\frac{1}{n}\bar{1}$ (recall that $\bar{1}$ denotes the all 1 vector). After one step, the distribution is $x^{(1)} = \frac{1}{d}Ax$, and after k steps it is $x^{(k)} = \frac{1}{d^k}A^k x$. Represent $x^{(0)}$ in the orthonormal basis of eigenvectors as $x^0 = \sum_{i=1}^n a_i v_i$. Note that $v_1 = \frac{1}{\sqrt{n}}\bar{1}$. Observe that $a_1 = \frac{1}{\sqrt{n}}$ (because $a_1 = (x^{(0)})^t \cdot v_1$, and the vector $x^{(0)}$ has one 1 entry), and $a_i < 1$ for every i. We get that $x^{(k)} = \frac{1}{d^k}\sum_{i=1}^n a_i(\lambda_i)^k v_i = \frac{1}{n}\bar{1} + \frac{1}{d^k}\sum_{i=2}^n a_i(\lambda_i)^k v_i$. If $\max[\lambda_2, -\lambda_n] < (1-\epsilon)d$, then after $k \geq \frac{3}{\epsilon}\log n$ steps, the vector $\frac{1}{d^k}\sum_{i=2}^n a_i(\lambda_i)^k v_i$ is negligible (in every coordinate) compared to the vector $\frac{1}{n}\bar{1}$, implying that the distribution of vertices reached is very close to uniform. Hence the walk mixes in $O(\frac{1}{\epsilon}\log n)$ steps, where ϵ is the eigenvalue gap $\frac{\lambda_1 - \max[\lambda_2, -\lambda_n]}{\lambda_1}$. This is one way of seeing that if the eigenvalue gap ϵ is large, all subset of vertices are expanding, and the graph is an expander.

1.2 Some additional relations between eigenvalues and graph properties

If the graph is not regular, the largest eigenvalue is at least as large as the average degree (by taking Raleigh quotient for all 1 vector) and likewise, at least as large as the average degree of any of its subgraphs.

If the graph is bipartite, nonzero eigenvalues come in pairs (flipping the entries of vertices

of one side in the eigenvectors). Hence $\lambda_n = -\lambda_1$. If $|\lambda_n|$ is much smaller than λ_1, this implies that the graph does not have very large cuts (that cut almost all the edges). The converse is not true, in the sense that $|\lambda_n|$ might be close to λ_1, and still the graph has no large cuts. This may happen for example if a d-regular graph G contains a small subgraph that is nearly a complete bipartite graph on $2d$ vertices. This small subgraph affects λ_n (as can be seen by an appropriate Rayleigh quotient) but is insignificant in terms of the existence of large cuts in G. Extensions of the eigenvalue based technique to semidefinite programming (SDP will be addressed in other lectures) give an approximation algorithm for max-cut with a very good approximation ratio (roughly 0.87856).

The following theorem is due to **Hoffman**, and illustrates the connection between eigenvalues and graph properties.

Theorem 1 *Let G be a d-regular n-vertex graph and let λ_n be the most negative eigenvalue of its adjacency matrix A. Then the size of its largest independent set is at most*

$$\alpha(G) \leq -\frac{n\lambda_n}{d-\lambda_n} = \frac{n|\lambda_n|}{d+|\lambda_n|}$$

Proof: Let S be an independent set in G. Consider the vector x with value $n-|S|$ on S and value $-|S|$ on the remaining vertices. Then we use the Rayleigh quotient to bound λ_n from above.

$$\lambda_n \leq \frac{x^t A x}{x^t x} = \frac{|S|^2(nd-2|S|d)-2d|S|(n-|S|)|S|}{|S|(n-|S|)^2+(n-|S|)|S|^2} = \frac{-n|S|^2 d}{n|S|(n-|S|)} = \frac{-|S|d}{n-|S|}$$

The proof follows by rearranging. □

1.3 Some notes on random graphs

The eigenvalues of A^2 are $(\lambda_i)^2$. The trace of A^2 implies that $\sum(\lambda_i)^2 = \sum d_i$. For d-regular graphs, this implies that the average absolute value of eigenvalues is at most $\sqrt{d}$. For random regular graphs, or random graphs that are nearly regular, it is in fact the case that with high probability over the choice of graph, all but the largest eigenvalue have absolute value $O(\sqrt{d})$. This can be proved by the trace method (considering higher even powers of A). An alternative proof can be based on considering the form $x^t A x$ for all possible unit vectors x orthogonal to v_1. (There are infinitely many such vectors so a certain discretization needs to be used.)

Much of what we said about the spectrum of regular graphs applies to graphs that are nearly regular. In establishing that a random graph in nearly regular, it is useful to have some bounds of the probability that a degree of a vertex deviates from its expectation. These bounds are quantitative versions of the law of large numbers and of the central limit theorem (that says that sums of independent random variables converge to the normal distribution), and we present one representative such bound.

Theorem 2 *Let $p_1, \ldots, p_n \in [0,1]$ and set $p = \frac{p_1+\ldots+p_n}{n}$. Assume that $X_1, \ldots X_n$ are mutually independent random variables with $Pr[X_i = 1-p_i] = p_i$ and $Pr[X_i = -p_i] = 1-p_i$. Let $X = X_1 + \ldots + X_n$. (Observe that $E[X_i] = 0$, and $E[X] = 0$.) Then*

$$Pr[X > a] < e^{-a^2/2pn + a^3/2(pn)^2}$$

and

$$Pr[X < -a] < e^{-a^2/2pn}$$

The theorem above is typically used when $p \leq 1/2$. Note that the bounds for $X < -a$ seem stronger than those for $X > a$. They can be used for $X > a$ after replacing p by $1-p$. Namely, $\Pr[X > a] < e^{-a^2/2(1-p)n}$. Note also that the bounds can be tightened in certain cases. See [2] for more details.

For a random graph G chosen according to the $G_{n,\frac{1}{2}}$ model, with high probability the following statements hold. The average degree is roughly $\frac{n}{2}$, the graph is nearly regular, and $|\lambda_n| \leq O(\sqrt{n})$. Theorem 1 can then be used to certify in polynomial time that $\alpha(G) \leq O(\sqrt{n})$. Likewise, by considering the edge complement of G, one can certify that $\omega(G) \leq O(\sqrt{n})$, where $\omega(G)$ is the size of the largest clique in G. If one plants in G an independent set of size larger that $c\sqrt{n}$ (for a sufficiently large constant c), then necessarily $|\lambda_n|$ will increase to at least $\frac{c}{2}\sqrt{n}$, a fact that we can detect in polynomial time. The proof of Theorem 1 suggests that the eigenvector that corresponds to λ_n should give us some information regarding which vertices belong to the planted independent set. Indeed, such an algorithm is designed in [1].

1.4 The Laplacian of a graph

If the graph is irregular, the **Laplacian** $L = D - A$ (where D is a diagonal matrix with the degrees of vertices along its diagonal) may be more informative then the adjacency matrix A. Note that for regular graphs, the spectra of these two matrices are easily related. For general graphs, the Laplacian gives the quadratic form $x^T Lx = \sum_{(i,j)\in E}(x_i - x_j)^2$. Hence L is **positive semidefinite**, with the all 1 vector as an eigenvector of eigenvalue 0. (More generally, the multiplicity of 0 as an eigenvalue equals the number of connected components.) The next eigenvalue is known as the Fiedler value and the associated eigenvector as the **Fiedler vector**. It gives a useful heuristic for spectral graph partitioning. The Fiedler vector and the next eigenvector form a convenient coordinate system for graph drawing (see [4]). An alternative (equivalent) definition of the Laplacian of a graph is as follows. Direct edges of the graph arbitrarily. Consider the incidence matrix M of the directed graph with vertices as rows, edges as columns, and $+1$ and -1 entries for incoming and outgoing edges. Then one can verify term by term that $L = MM^T$.

Sometimes it is useful to consider the normalized Laplacian, whose diagonal is 1, and for edge (i, j), the entry (i, j) has value $\frac{1}{\sqrt{d_i d_j}}$.

2 Refuting random 4SAT

We illustrate here an application (due to Goerdt and Krivelevich [3]) of spectral techniques to a seemingly unrelated problem.

A 4CNF formula contains clauses, where each clause contains four literals, and each literal is either a Boolean variable or its negation. An assignment satisfies the formula if

in each clause at least one literal is set to true. Checking whether a given 4CNF formula is satisfiable is NP-complete. Namely, it is in NP (if the answer is yes there is a satisfying assignment certifying this), and it is NP-hard. Checking whether a given 4CNF formula is not satisfiable is coNP-complete. Hence there are no witnesses for non-satisfiability unless coNP=NP. This applies for worst case instances. Here we study the average case (for some natural distribution).

Consider the distribution $F_{n,m}$ of random 4CNF formulas with n variables and m clauses. The formula is random in the sense that every literal in the formula is chosen independently at random. (We may further require that clauses are distinct and that variables in a clause are distinct, but this has negligible effect on the results that we present.) Observe that given any assignment to the variables, a random clause is satisfied by this assignment with probability exactly 15/16 (every literal has probability 1/2 of getting polarity opposite to the assignment). Hence the probability that it satisfies the random formula is $(15/16)^m$, and the expected number of satisfying assignments is exactly $2^n(15/16)^m$. Hence when m is sufficiently large (say, $m = 16n$) it is very likely that the formula is not satisfiable. We shall be interested in developing refutation heuristics that run in polynomial time and combine the following two properties:

1. *Perfect Soundness.* If the heuristic outputs *not satisfiable* then indeed the 4CNF formula is not satisfiable.

2. *Average completeness.* For most non-satisfiable formulas from a given distribution of formulas (this distribution will be specified shortly), the heuristic answers *not satisfiable.*

The distribution that we shall consider is $F_{n,m}$ with m very large compared to n. For m above the threshold for satisfiability, the heuristic refutation problem becomes easier as m grows (because refuting a subformula implies refuting the whole formula). Hence we shall consider first very large values of m, and then address smaller values of m.

When $m = \Omega(n^4)$ refutation becomes very easy. With high probability, there is a set of four variables that provide the support to a super constant number of clauses. With high probability, all possible polarity combinations (there are 16 of them) appear, and hence any assignment to these four variables leaves at least one clause unsatisfied. It is easy to detect this subformula of 16 clauses in polynomial time.

When $m = n^3$, we have the following refutation algorithm. Pick two variables (say x_1 and x_2). Try all possible assignments to these two variables (four combinations altogether). For each such assignment, simplify the resulting 4CNF formula by removing satisfied clauses, and removing unsatisfied literals. The resulting formula is expected to have roughly $3n$ 2CNF clauses. (Each 4CNF clause has 6 pairs of variables, there are $\binom{n}{2}$ distinct pairs, and there are 4 assignments.) The probability that it is satisfiable is at most $2^n(3/4)^{3n} << 1/8$ (for sufficiently large n). (In fact, it is known that a random 2CNF formula with $(1+\epsilon)n$ clauses is already unlikely to be satisfiable, but we are not interested in optimizing the parameters for the approach described here, so we will not use this fact.) It follows that with high probability, for each of the four assignments simultaneously, the corresponding 2CNF formula is not satisfiable. As satisfiability for 2CNF can be tested in polynomial time, we get a refutation heuristic for 4SAT (when $m = n^3$).

Now let us consider the case $m = cn^2$ for large enough c. To simplify the presentation (though with tighter analysis this simplification is not needed), let us give ourselves more slackness and view c not as a constant, but rather as n^ϵ (where $\epsilon > 0$ and n is sufficiently large).

Given the 4CNF formula ϕ, partition it into three subformulas. ϕ^+ contains only those clauses in which all literals are positive, ϕ^- contains only those clauses in which all literals are negative, and ϕ' contains the remaining clauses. We completely ignore ϕ', and construct two graphs, G^+ based on ϕ^+, and G^- based on ϕ^-. We describe the construction of G^+, and the construction of G^- is similar.

The vertex set V contains $\binom{n}{2}$ vertices, where each vertex is labeled by a distinct pair of distinct variables. For every clause in ϕ^+ (that we assume contains 4 distinct variables), put an edge in G^+ between the vertex labeled by the first two variables in the clause and the vertex labeled by the last two variables in the clause.

Lemma 3 *If ϕ is satisfiable, then at least one of the two graphs G^+ and G^- has an independent set of size at least $\binom{n/2}{2} \simeq |V|/4$.*

Proof: Consider an arbitrary satisfying assignment for ϕ, let S^+ be the set of variables assigned to true, and let S^- be the set of variables assigned to false. Consider the set of $\binom{|S^-|}{2}$ vertices in G^+ labeled by pairs of vertices from S^-. They must form an independent set because ϕ cannot have a clause containing only variables from S^- in which all literals are positive. Likewise, G^- has an independent set of size at least $\binom{|S^+|}{2}$. As $\min[|S^+|, |S^-|] \geq n/2$, the proof follows. $\square$

Observe that for $F_{n,m}$ with $m = cn^2$, both G^+ and G^- are random graphs with roughly $m/16 \simeq c|V|/8$ edges (recall that literals within a clause are chosen independently at random, and remove parallel edges and self loops), and hence average degree $c/4$. For large enough c, they are nearly regular (the standard deviation of degrees is $O(\sqrt{c})$ which becomes negligible as c grows). Hence it will be the case that for their adjacency matrices, the eigenvalues satisfy $\lambda_1 \simeq c/4$ and $|\lambda_{|V|}| = O(\sqrt{c})$. Using Theorem 1 (adapted to nearly regular graphs), this can be used in order to certify that neither G^+ nor G^- have independent sets larger than $O(|V|/\sqrt{c})$. For sufficiently large c, this establishes that ϕ cannot have a satisfying assignment.

It is a major open question to design refutation heuristics for $F_{n,m}$ when $m = n^{2-\epsilon}$.

References

[1] Noga Alon, Michael Krivelevich, Benny Sudakov: Finding a large hidden clique in a random graph. Random Struct. Algorithms 13(3-4): 457–466 (1998).

[2] Noga Alon and Joel Spencer. The Probabilistic Method. Wiley-Interscience Series in Discrete Mathematics and Optimization.

[3] Andreas Goerdt, Michael Krivelevich: Efficient Recognition of Random Unsatisfiable k-SAT Instances by Spectral Methods. STACS 2001: 294-304

[4] Dan Spielman. Lecture notes in http://cs-www.cs.yale.edu/homes/spielman/.