Algebraic Methods in Computer Science	December 19, 2014
Lecture: 4	
Lecturer: *Amir Shpilka*	Scribe: *Shir Peleg*

Eigenvalues of Adjacency Matrices

In this and the next lecture we will use spectral methods to obtain information about graphs. Specifically, we will look at eigenvalues of the adjacency matrix of a graph and use it to neat combinatorial theorems. We will then discuss expander graphs where eigenvalues play an important role both by giving an algebraic way of defining expanders and by providing an efficient certificate for expansion.

Let $A \in M_n(\mathbb{F})$ ($n \times n$ matrices over the field $\mathbb{F}$), we call λ an eigenvalue of A if there is $v \in \mathbb{F}^n$, $v \neq 0$ such that $A \cdot v = \lambda v$. Such v is called an eigenvector of A with eigenvalue λ.

Fact 1. *Let A be a symmetric $n \times n$ matrix over $\mathbb{R}$, then A has n orthogonal eigenvectors (not necessarily with different eigenvalues).*

Let $G = (E, V)$, $|V| = n$ we denote by A_G its adjacency matrix: $(A_G)_{(i,j)} = 1 \iff (i, j) \in E$. A_G is symmetric and due to Fact 1, there are $\lambda_1 \geq ... \geq \lambda_n \in \mathbb{R}$ eigenvalues of A_G.

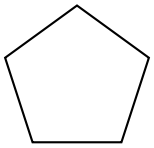
Hoffman - Singleton Theorem

The first example for the use of eigenvalues that we give is the Hoffman-Singelton theorem. Recall that the girth of a graph G is the length of the shortest cycle in G. We will use the following notation. For $v \in V$ we denote with $\Gamma(v)$ the neighborhood of $v$in G. That is, $\Gamma(v) = \{u \in V \ s.t (u, v) \in E\}$. We also denote the $d$egree of v to be $d(v) = |\Gamma(v)|$. A graph G is called d-regular if all the vertices have degree d, i.e., $\forall v \in V$, $d(v) = d$.

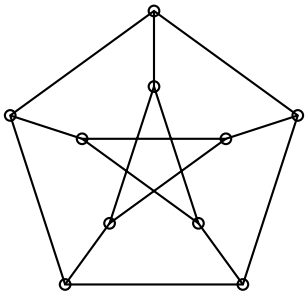
The Hoffman-Singleton theorem is concerned with the following question: "What is the minimal number of vertices that a d-regular graph with girth g can have?" For g = 1 or 2 the question in not interesting, because it demands graphs with self loops or double edges. The case $g = 3$ is also quite simple. The smallest d-regular graph must have $d + 1$ vertices and K_{d+1} is such a graph with girth 3. The case $d = 4$ is also quite simple. The minimal d-regular graph with girth 4 is $K_{d,d}$. Indeed, let $G = (E, V)$ be a d-regular graph on n vertices with girth of 4. Let v be an arbitrary vertex of G. Clearly v has d different neighbors, call them $x_1, ..., x_d$. Since the girth of G is 4, we don't have any edges between those x_i's (as otherwise we will get a triangle). Since x_1 has $d - 1$ neighbors besides v, there must be additional $d - 1$ vertices in the graph. Thus, $n \geq 2d$. Thus, the first non-trivial case is $g = 5$, and this is the focus of the Hoffman-Singleton theorem.

We first try to understand what is the obvious lower bound on n. Let $G = (E, V)$, $|V| = n$ be a d-regular graph with girth 5, and let v be a vertex in G. As before, v has d different neighbors $x_1, ..., x_d$, and we do not have any edge among the x_i's. Denote the additional $d-1$

neighbors of x_i with $\Gamma(x_i) = \{y_{i,1}, ..., y_{i,d-1}\}$. Since the girth is larger than 4 we must have that $\Gamma(x_i) \cap \Gamma(x_k) = \emptyset$ for $i \neq k$. Indeed, if $k \neq i,\ y_{i,j} = y_{k,l}$ then we get a cycle of length 4, namely, $v, x_i, y_{i,j} = y_{k,l}, x_k$. As a consequence we get that $n \geq 1 + d + d(d-1) = d^2 + 1$. Now, a natural question to ask is whether such a graph with n vertices exist. The first case to study is $d = 2$. Here, $n = d^2 + 1 = 5$ and indeed, the cycle of length 5 is such a graph.



We next go to $d =$ and $n = 10$. Here, Peterson's graph is what we are looking for.



the picture now gets more complicates as there are no such extremal graphs (girth 5, d-regular and $n = d^2 + 1$ vertices) for $d = 4, 5, 6$. However, for $d = 7$ Hoffman and Singleton provided a construction of such a graph. Thus, an interesting question is for which values of d such graphs exist? The Hoffman-Singleton theorem provides a surprising answer to this question.

Theorem 1 (Hoffman-Singleton Theorem)**.** *Let $G = (E, V)$ be a d-regular graph with girth 5 over $|V| = n = d^2 + 1$ vertices. Then $d \in \{2, 3, 7, 57\}$.*

Thus, the theorem says that besides the examples that we mentioned so far there may be another one for $d = 57$. It is still an intriguing question whether such a graph exists.

The proof of the theorem is a bit magical. We will first show that such graphs are *s*trongly regular. That is, the number of common neighbors of two vertices v and u only depends on whether v and u are neighbors or not. We will then argue that such graphs have exactly three eigenvalues, one of them must be d. By looking into the multiplicities of each eigenvalue we will get a polynomial equation of degree 4. The four solutions to this equation correspond to the possible values of d.

Proof. Let $A = A_G$ be the adjacency matrix of G and $\lambda_1 \geq ... \geq \lambda_n \in \mathbb{R}$ its eigenvalues. Consider A^2, notice that $(A^2)_{(i,i)} = d$ and for $i \neq j$, $(A^2)_{(i,j)} = |\{v \in V \mid (i, v) \wedge (v, j) \in E\}|$, i.e., it is the number of common neighbors of i and j. Let $u, v \in V$. It follows that:

- If $(u, v) \in E$ then $(A^2)_{(u,v)} = 0$ as otherwise they have a common neighbor, creating a cycle of length 3.

- If $(u,v) \notin E$ then $(A^2)_{(u,v)} = 1$. Indeed, if $(A^2)_{(u,v)} > 1$, then they have at least 2 common neighbors creating a cycle of length 4. On the other hand, if $(A^2)_{(u,v)} = 0$ then the graph is not extremal, as our counting showed that when $n = d^2 + 1$ every two vertices are either connected or at distance 2 from each other.

Therefore,

$$A^2 = d \cdot I + A_{\overline{G}} = d \cdot I + J - A - I$$

where I is the identity matrix, and J is the matrix with all ones. We thus have

$$A^2 = J - A + (d-1)I. \tag{1}$$

Recall that $A \in M_n(\mathbb{R})$ is symmetric, and as we saw (e.g. in HW 1), d is an eigenvalue of A, with an eigenvector $u = (1, ..., 1)$. Furthermore, d is the maximal eigenvalue of A in absolute value[1]. We also know (again, we saw it in HW 1) that since G is connected, the multiplicity of d is 1. We can therefore deduce that if v is an eigenvector of A with eigenvalue different than d, then $\langle v, u \rangle = 0$, and so $J \cdot v = 0$ (as all the rows in J equal u). Thus, if v is such an eigenvector, with eigenvalue s, then we have (using Equation (1))

$$(A^2 + A - J) \cdot v = (d-1)I \cdot v,$$

hence,

$$x^2 \cdot v + x \cdot v + 0 \cdot v = (d-1)v.$$

Thus,

$$x^2 + x = d - 1$$

and the possible values for x are

$$\lambda_1 = \frac{-1 + \sqrt{4d-3}}{2},$$
$$\lambda_2 = \frac{-1 - \sqrt{4d-3}}{2}.$$

These may or may not be eigenvalues, but there can be no other. Denote with m_i the multiplicity of λ_i. Since $d$has multiplicity 1, it follows that

$$1 + m_1 + m_2 = n,$$

and since $n = d^2 + 1$ we get

$$m_1 + m_2 = d^2. \tag{2}$$

Since the trace of a symmetric matrix equals the sum of its eigenvalues (counted with multiplicities) we have that

$$0 = \mathrm{Tr}(A) = d + \lambda_1 \cdot m_1 + \lambda_2 \cdot m_2.$$

[1] All eigenvalue are bounded by the value of the maximal L_1 norm of a row of A.

Substituting λ_1, λ_2 with their value, and using a simple manipulation, we obtain

$$\begin{aligned} 0 &= d \cdot 1 + \frac{-1+\sqrt{4d-3}}{2} \cdot m_1 + \frac{-1-\sqrt{4d-3}}{2} \cdot m_2 \\ d - \frac{1}{2}(m_1+m_2) &= (m_2-m_1)\frac{\sqrt{4d-3}}{2} \\ 2d - (m_1+m_2) &= (m_2-m_1)\sqrt{4d-3}. \end{aligned}$$

Combining with Equation 2 we get

$$2d - d^2 = (m_1 - m_2)\sqrt{4d-3}.$$

If $m_2 - m_1 = 0$ then $2d - d^2 = 0$ and hence $d = 2$, which is one of the cases we allow. Assuming $m_2 - m_1 \neq 0$ we get

$$\frac{2d-d^2}{m_1-m_2} = \sqrt{4d-3}. \tag{3}$$

Thus, $4d-3$ is an integer whose square root is a rational number. It therefore must be the case that $\sqrt{4d-3}$ is an integer. Denote $k = \sqrt{4d-3} \in \mathbb{N}$. Thus,

$$k^2 = 4d - 3 \Rightarrow d = \frac{k^2+3}{4}.$$

Rewriting Equation 3 as a polynomial in k we get

$$\frac{k^2+3}{2} - \frac{k^4+6k^2+9}{16} = (m_1-m_2)k$$

and after some simple manipulations we obtain

$$15 = (k^3 - 2k^2 + 16(m_1-m_2))k.$$

In particular k divides 15. Thus, $k \in \{1,3,5,15\}$ and therefore, $d = \frac{k^2+3}{4} \in \{1,3,7,57\}$. Earlier we also obtained $d = 2$ as a possible solution. Since we don't care about graphs with $d = 1$ it follows that $d \in \{2,3,7,57\}$, as claimed. □

Definition 1 (Strongly regular graphs)**.** *$G = (E,V)$ is called strongly regular if $\forall u, v \in V$ $|\Gamma(v) \cap \Gamma(u)|$ depends only on whether $(u,v) \in E$.*

Note that in the previous proof we only used the fact that G is strongly regular.

Friendship Theorem

The second example concerns friends and politicians. If in a room with n people, every two have exactly one common neighbor (different than both of them), then there is a politician in the room (a person who is friend of everyone).

Theorem 2. *Let $G = (V,E)$ be such that any two different vertices $u \neq v \in V$ have exactly one common neighbor, i.e., $|\Gamma(v) \cap \Gamma(u)| = 1$. Then, there exists $w \in V$ that is connected to all other vertices: $\forall w \neq u \in V$, $(u,w) \in E$.*

The proof proceeds as follows. We assume for a contradiction that this is not the case. We first prove that G must be a regular graph. Together with the assumption on G this implies that G is strongly regular. We then analyze the spectrum of G, similarly to what we did in the proof of Theorem 1, and using properties of integer numbers conclude that G must be a triangle. This is a contradiction as a triangle has the required property.

Proof. Assume towards a contradiction that there is no such w.

Claim 1. *G is regular.*

Proof. Since no vertex is connected to all others, there must be a pair of non-neighbors, u and v. First we show that $\deg(u) = \deg(v)$. Denote $\Gamma(v) = \{w_1, ..., w_k\}$. By the assumption on G, u and v have a common neighbor, w.l.o.g. this neighbor is w_2. As they share exactly one common neighbor, u is not connected to any other w_i. We also know that v and w_2 have a common neighbor, which w.l.o.g. will assume it is w_1. Now, for every $2 \leq i \leq k$, u and w_i have a common neighbor, which we denote with z_i. Observe that $z_i \neq z_j$ because v is the unique neighbor of w_i and w_j. Thus, u has $w_2, z_2, ..., z_k$ as neighbors. In particular this implies that $\deg(u) \geq \deg(v)$, which by symmetry implies $\deg(u) = \deg(v)$.

We now show that every other w satisfies $\deg(w) = \deg(u) = \deg(v)$. Indeed, for every $w \neq w_2 \in V$, we know that either $(u, w) \notin E$ or $(v, w) \notin E$ because w_2 is the unique common neighbor of u and v. Thus, we can use the same proof again and conclude that $\deg(w) = \deg(v) = \deg(u)$. In particular, $\forall w \neq w_2 \in V$, $\deg(w) = k$. As we know w_2 is not connected to all other vertices (we assumed that no vertex is connected to all others), there is $t \in V$ such that $(w_2, t) \notin E$ and, as before, we can deduce $\deg(w_2) = k$. This proves that G is regular. □

Back to the proof of the theorem, we now know that G is k-regular. As before, we have that k is an eigenvalue of G, with eigenvector $u = (1, ..., 1)$, and it has multiplicity one because G is connected. As before we denote $A = A_G$ and consider A^2. We have that

$$(A^2)_{(i,j)} = \begin{cases} 1 & i \neq j \\ k & i = j \end{cases}.$$

Hence, $A^2 = (k-1)I + J$. If x is an eigenvector of A with eigenvalue $\lambda \neq d$ then $x \perp u$ and therefore,

$$A^2 \cdot x = ((k-1)I + J) \cdot x$$
$$\lambda^2 \cdot x = (k-1)x.$$

Thus, λ must satisfy $\lambda^2 = k - 1$ and the two possible solutions are $\lambda_1 = \sqrt{k-1}$ and $\lambda_2 = -\sqrt{k-1}$. Letting m_i denote the multiplicity of λ_i we get

$$0 = \mathrm{Tr}(A) = k + \lambda_1 \cdot m_1 + \lambda_2 \cdot m_2,$$

hence,

$$k = (m_2 - m_1)\sqrt{k-1}$$
$$k^2 = (m_2 - m_1)^2 \cdot (k-1).$$

In particular $k-1$ divides k^2. This is only possible if $k=2$ (since $k-1$ and k do not share any prime factor). However, the only 2 regular graph that satisfies the requirements of the theorem is K_3. And K_3 does have a vertex that is a neighbor of all other vertices, in contradiction to our assumption. We thus conclude the proof of the theorem. □

Expander Graphs

In the next couple of lectures we shall look at a certain family of graphs that roughly "have the characteristics of random graphs." For example, consider a random d-regular graph, and any subset A consisting of k vertices, with k not too large. One would expect that the number of neighbors of A will be about $k \cdot d$, in other words, not too large sets expand very well in random graphs. Expander graphs are graphs in which every not too large set expands, i.e., has many neighbors.

Before describing expander graphs we will start by talking of two unrelated problems that can be solved using expander graphs. We will then define a class of "magical" graphs and show how they can be used for solving those problems.

Two motivating problems

Amplification of randomized algorithms Let $A(x,r)$ be a randomized algorithm. That is, A is an algorithm that has to decide whether an input x belongs to some language $\mathcal{L}$ and for doing so A may use a string r of random bits (independent of the input). We say that A has *one-sided error* if the following holds: if $x \in \mathcal{L}$ then for every r, $A(x,r)=1$. On the other hand, if $x \notin \mathcal{L}$ then $\Pr_r[A(x,r)=1] \leq 1/20$. That is, A never makes a mistake when x is in the language, but it may err when x is not in $\mathcal{L}$. Notice that if A returns 0 for some setting of r then it must be the case that $x \notin \mathcal{L}$.

The problem that we are interested in is how to reduce the error probability. It is clear that if we run A with k randomly chosen random strings and take the AND of the outputs then the probability of error goes down to $(1/20)^k$. That is, $\Pr_{r_1,\ldots,r_k}[A(x,r_1) \wedge A(x,r_2) \wedge \cdots \wedge A(x,r_k)=1] \leq (1/20)^k$. However, this amplification requires $k \cdot |r|$ random bits and if we think of randomness as a resource we wish to save then this is too costly. Indeed, the "randomness" that computers provide is not truly uniformly distributed so we would like to save on the amount of randomness we obtain from the computer. Thus, it is an intriguing question how to reduce the probability of error without spending more random bits.

Constructing error correcting codes In real life scenarios when we transmit data over a communication channel some disruptions may occur which will change some of the bits we have meant to send. The other side who received the message will have trouble interpreting the message or even worse, will get a totally different message. We wish to find a way to encode our messages such that if few bits have changed the other side will both understand that the content message have been corrupted and be able to deduce what the original content was. Such as encoding is called an Error Correcting Code.

We next define what a good linear error correcting code is. Basically, every linear subspace of $\mathbb{F}_2^n$ is an error correcting code, albeit with not so good parameters. To understand which parameters make a code good we first define a notion of distance between two n-bit strings.

Definition 2 (Hamming distance)**.** *Let $x, x' \in \mathbb{F}_2^n$. Their Hamming distance is defined as $dist(x, x') = |\{1 \leq i \leq n \ s.t \ x_i \neq x'_i\}|$. That is, it is the number of different bits in the two words.*

Notice that if the distance between x and x' is d and we modify x in less than $d/2$ locations to get a word y, then we will have that $\text{dist}(y, x) < \text{dist}(y, x')$. Thus, we will be able to recover x from y, which can be tough of as a corrupted version of x. This gives rise to the following definition.

Definition 3 (Distance of Code)**.** *Let $C \subseteq \mathbb{F}_2^n$. The distance of C is defined as $dist(C) = \min_{x,x' \in C} dist(x, x')$.*

Thus, if we view C as the set of allowed codewords to transmit, then even if we meant to transmit a message $x \in C$ and the codeword was corrupted in less than $\text{dist}(C)/2$ many locations, then x will still be the closest codeword to the received corrupted word. Thus, decoding is possible (although in a non-efficient way). In particular, a code with a large distance can tolerate many errors. Clearly if C only contains the all 0 word and the all 1 word then it has a fantastic distance. The problem is that such C has only two codewords and thus allow us to send only two different messages. The next definition gives another important parameter of the code, which basically measures how many bits does any codeword "stores". Given a $C \subset \mathbb{F}_2^n$ of size 2^k, we can think of every message of C as storing k bits of information, by mapping C to $\mathbb{F}_2^k$. The rate of C is a measure of how wasteful this encoding is.

Definition 4 (Rate of Code)**.** *Let $C \subseteq \mathbb{F}_2^n$ be a code such that $|C| = 2^k$. I.e., any single message of C holds k bits of information. The* rate *of C, denoted $R(C)$, is the ratio between the information transferred and the amount of data transmitted. That is, $R(c) = \frac{k}{n}$.*

Thus, our goal is to construct codes that will have high rate (so that each message will contain a lot of information) and large distance (so we can tolerate many errors). This two parameters clearly affect each other. For example, it is clear that we cannot have $R(c) = 1$ and $\text{dist}(X) = n$, while each of them can be achieved separately.

The efficiency of a code depends on the complexity of encoding messages (the mapping between $\mathbb{F}_2^k$ and C), on the decoding complexity (finding the transmitted codeword from its corrupted version), on the rate and on the amount of errors that can be handled (which is related to the distance of the code). It is an important question of coming up with families of codes, of lengths going to infinity, that can be efficiently encoded and coded, that have constant rate, and that can decode from a constant fraction of errors. A family of codes that satisfy this requirements is called a "good family of codes".

One can easily note that picking the code C at random, for $|C| = 2^{Rn}$ gives a code of rate R and distance, roughly, $(1 - H^{-1}(1 - R))n$, where $H(\cdot)$ is the entropy function, $H(x) = -x'logx - (1 - x)\log(1 - x)$ for $x \in (0, 1)$. The same also holds when C is a linear subspace of dimension Rn.

We will focus on the question of constructing a good family of linear codes. It is clear that if C is a linear subspace then the mapping between $\mathbb{F}_2^k$ can be computed efficiently, as it is just multiplication of a vector by a matrix. However, decoding is still a difficult task.

Magical graphs

We shall now define a certain class of "magical' graphs that will allow us to solve the two problems raised above. The main property of this family is that "small" sets have many neighbors.

Definition 5 (Magical Graphs)**.** *Let $G = (V, E)$, be a d-regular bipartite graph graph with $V = L \sqcup R$, where $|L| = n$ and $|R| = m$. We say that G is "magical" if*

1. *$\forall S \subseteq L$ if $|S| \leq \frac{n}{10d}$ then $|\Gamma(S)| \geq \frac{5}{8}d|S|$ and*
2. *$\forall S \subseteq L$ if $|S| \leq \frac{n}{2}$ then $|\Gamma(S)| \geq |S|$.*

The main property is that if the size of S is not too large than nearly all edges leaving S go to distinct vertices in R.

Later, when we define expander graphs, we will see that this definitions captures our notion of an expander graph with expansion $5d/8$.

Applications

Error amplification: We will now see how one can use such "magical" graphs in order to deterministically reduce error in randomized algorithms. That is, to reduce the error without using additional random bits.

Let $A(x, r)$ be a one-sided randomized algorithm as before, and assume that the probability of error is smaller than $1/20$. We would like to reduce the error that A makes, without investing more random bits. Our first application will show how to trade randomness for running time. That is, we will show that in order to reduce the error to $1/T$ we only need to run the algorithm, $O(T)$ many times, without using more than $|r|$ bits. Later in the course we will see how to reduce the error to $\exp(-t)$ while using only $|r| + O(t)$ many random bits (compare this to running the algorithm t times with fresh randomness that requires $|r| \cdot t$ many random bits).

Our approach is, given a random string of length r we will generate T different $|r|$-bit strings from it, and run A on those random strings. Notice, that since this process is deterministic those random strings highly depend on each other. Nevertheless, we will prove that it is still the case that A can err on all runs with small probability.

Construction 3 (Amplified A)**.** *Let G be a "magical" graph as in Definition 5, with parameters $n = m = 2^r$. We identify the elements $[n]$ with the vertices of the Boolean cube $\{0,1\}^r$. Given a random string $v \in \{0,1\}^r$, think of v as a vertex in L and let $u_1, \ldots, u_d$ be its neighbors in R. Return the value $A(x, u_1) \wedge A(x, u_2) \wedge \cdots \wedge A(x, u_d)$.*

In other words, given a random string of length $|r|$ we first compute its neighbors in G and then run the algorithm on each of its neighbors.

Let us analyze this algorithm. If $x \in \mathcal{L}$ then A never errs on x and therefore will always return 1. So assume that $x \notin \mathcal{L}$ and let

$$\Pr_r[A(x, r) = 1] = \delta \leq 1/20.$$

Let $B_x \in \{0,1\}^r$ the set of random strings that cause A to make a mistake on x. Note that our amplified algorithm makes a mistake on x only if we picked a random string v such that $\Gamma(v) \subseteq B_x$. Thus, to estimate the probability of error we have to understand the probability of picking such a v. Let $S = \{v \in \{0,1\}^r \mid \Gamma(v) \subseteq B_x\}$. Since $|B_x| \leq n/20$ we get from Definition 5 that

$$|S| \leq \frac{|B_x|}{\frac{5d}{8}} \leq \frac{8}{5d}\delta \cdot 2^r \leq 2^r/10d.$$

In particular, the probability of picking a bad v is at most $1/10d$.

Thus, if we take such a graph G with left-degree $T/10$, then the overall error probability will be at most $1/T$, and we only need to run A, our algorithm, $T/10$ many times.

Error correcting codes: We will give this application in the next class.

Notes

Our treatment of the Hoffman-Singleton theorem follows that of Babai and Frankl [BF92]. The proof of the Friendship theorem follows the one given in [AZ04]. Discussion of magical graphs is taken from [HLW06].

References

[AZ04] Martin Aigner and Günter M. Ziegler, *Proofs from THE BOOK (3. ed.)*, Springer, 2004. 4-9

[BF92] László Babai and Péter Frankl, *Linera algebra methods in combinatorics (with applications to geometry and computer science)*, Manuscript, 1992. 4-9

[HLW06] S. Hoory, N. Linial, and A. Wigderson, *Expander graphs and their applications*, Bulleting of the American Mathematical Society **43** (2006), 439–561. 4-9