Introduction to expander graphs

Michael A. Nielsen[1,*]

[1]School of Physical Sciences, The University of Queensland, Brisbane, Queensland 4072, Australia

(Dated: June 22, 2005)

I. INTRODUCTION TO EXPANDERS

Expander graphs are one of the deepest tools of theoretical computer science and discrete mathematics, popping up in all sorts of contexts since their introduction in the 1970s. Here's a list of some of the things that expander graphs can be used to do. Don't worry if not all the items on the list make sense: the main thing to take away is the sheer *range* of areas in which expanders can be applied.

- *Reduce the need for randomness:* That is, expanders can be used to reduce the number of random bits needed to make a probabilistic algorithm work with some desired probability.
- *Find good error-correcting codes:* Expanders can be used to construct error-correcting codes for protecting information against noise. Most astonishingly for information theorists, expanders can be used to find error-correcting codes which are efficiently encodable and decodable, with a non-zero rate of transmission. This is astonishing because finding codes with these properties was one of the holy grails of coding theory for *decades* after Shannon's pioneering work on coding and information theory back in the 1940s.
- *A new proof of PCP:* One of the deepest results in computer science is the PCP theorem, which tells us that for all languages L in **NP** there is a randomized polyonomial-time proof verifier which need only check a *constant* number of bits in a purported proof that $x \in L$ or $x \notin L$, in order to determine (with high probability of success) whether the proof is correct or not. This result, originally established in the earlier 1990s, has recently been given a new proof based on expanders.

What's remarkable is that none of the topics on this list appear to be related, *a priori*, to any of the other topics, nor do they appear to be related to graph theory. Expander graphs are one of these powerful unifying tools, surprisingly common in science, that can be used to gain insight into an an astonishing range of apparently disparate phenomena.

I'm not an expert on expanders. I'm writing these notes to help myself (and hopefully others) to understand a little bit about expanders and how they can be applied. I'm not learning about expanders with any specific intended application in mind, but rather because they seem to behind some of the deepest insights we've had in recent years into information and computation.

What is an expander graph? Informally, it's a graph $G = (V, E)$ in which every subset S of vertices *expands* quickly, in the sense that it is connected to many vertices in the set $\overline{S}$ of complementary vertices. Making this definition precise is the main goal of the remainder of this section.

Suppose $G = (V, E)$ has n vertices. For a subset S of V we define the *edge boundary* of S, ∂S, to be the set of edges connecting S to its complement, $\overline{S}$. That is, ∂S consists of all those edges (v, w) such that $v \in S$ and $w \notin S$. The *expansion parameter for* G is defined by

$$h(G) \equiv \min_{S:|S|\leq n/2} \frac{|\partial S|}{|S|}, \tag{1}$$

where $|X|$ denotes the size of a set X.

One standard condition to impose on expander graphs is that they be d-regular graphs, for some constant d, i.e., they are graphs in which every vertex has the same degree, d. I must admit that I'm not entirely sure why this d-regularity condition is imposed. One possible reason is that doing this simplifies a remarkable result which we'll discuss later, relating the expansion parameter $h(G)$ to the *eigenvalues* of the adjacency matrix of G. (If you don't know what the adjacency matrix is, we'll give a definition later.)

Example: Suppose G is the complete graph on n vertices, i.e., the graph in which every vertex is connected to every other vertex. Then for any vertex in S, each vertex in S is connected to *all* the vertices in $\overline{S}$, and thus $|\partial S| = |S| \times |\overline{S}| = |S|(n - |S|)$. It follows that the expansion parameter is given by

$$h(G) = \min_{S:|S|\leq n/2} n - |S| = \left\lceil \frac{n}{2} \right\rceil. \tag{2}$$

For reasons I don't entirely understand, computer scientists are most interested in the case when the degree, d, is a small constant, like $d = 2, 3$ or 4, not $d = n - 1$, as is the case for the complete graph. Here's an example with constant degree.

Example: Suppose G is an $n \times n$ square lattice in 2 dimensions, with periodic boundary conditions (so as to make the graph 4-regular). Then if we consider a large connected subset of the vertices, S, it ought to be plausible that that the edge boundary set ∂S contains roughly one edge for each vertex on the perimeter of the region S. We expect there to be roughly $\sqrt{|S|}$ such vertices, since

*nielsen@physics.uq.edu.au and www.qinfo.org/people/nielsen

we are in two dimensions, and so $|\partial S|/|S| \approx 1/\sqrt{|S|}$. Since the graph can contain regions S with up to $O(n^2)$ vertices, we expect

$$h(G) = O\left(\frac{1}{n}\right) \tag{3}$$

for this graph. I do not know the exact result, but am confident that this expression is correct, up to constant factors and higher-order corrections. It'd be a good exercise to figure out exactly what $h(G)$ is. Note that as the lattice size is increased, the expansion parameter decreases, tending toward 0 as $n \to \infty$.

Example: Consider a random d-regular graph, in which each of n vertices is connected to d other vertices, chosen at random. Let S be a subset of at most $n/2$ vertices. Then a typical vertex in S will be connected to roughly $d \times |\overline{S}|/n$ vertices in $\overline{S}$, and thus we expect $|\partial S| \approx d \times |S||\overline{S}|/n$, and so

$$\frac{|\partial S|}{|S|} \approx d\frac{|\overline{S}|}{n}. \tag{4}$$

Since $|\overline{S}|$ has its minimum at approximately $n/2$ it follows that $h(G) \approx d/2$, independent of the size n.

Exercise: Show that a disconnected graph always has expansion parameter 0.

In each of our examples, we haven't constructed just a single graph, but rather an entire family of graphs, indexed by some parameter n, with the property that as n gets larger, so too does the number of vertices in the graph. Having access to an entire family in this way turns out to be much more useful than having just a single graph, a fact which motivates the definition of expander graphs, which we now give.

Suppose we have a family $G_j = (V_j, E_j)$ of d-regular graphs, indexed by j, and such that $|V_j| = n_j$ for some increasing function n_j. Then we say that the family $\{G_j\}$ is a *family of expander graphs* if the expansion parameter is bounded strictly away from 0, i.e., there is some small constant c such that $h(G_j) \geq c > 0$ for all G_j in the family. We'll often abuse nomenclature slightly, and just refer to the expander $\{G_j\}$, or even just G, omitting explicit mention of the entire family of graphs.

II. EXPLICIT EXAMPLES OF EXPANDERS

We've seen previously that a family of d-regular random graphs on n vertices defines an expander. For applications it is often more useful to have more explicit constructions for expanders. In particular, for applications to algorithms it is often useful to construct expanders on $O(2^n)$ vertices, where n is some parameter describing problem size. Just to store a description of a random graph on so many vertices requires exponentially much time and space, and so is not feasible. Fortunately, more parsimonious constructions are possible, which we now describe.

Example: In this example the family of graphs is indexed by a prime number, p. The set of vertices for the graph G_p is just the set of points in Z_p, the field of integers modulo p. We construct a 3-regular graph by connecting each vertex $x \neq 0$ to $x-1, x+1$ and x^{-1}. The vertex $x = 0$ is connected to $p-1, 0$ and 1. According to the lecture notes by Linial and Wigderson, this was proved to be a family of expanders by Lubotsky, Phillips and Sarnak in 1988, but I don't know a lower bound on the expansion parameter. Note that even for $p = O(2^n)$ we can do basic operations with this graph (e.g., random walking along its vertices), using computational resources that are only polynomial in time and space. This makes this graph potentially far more useful in applications than the random graphs considered earlier.

Example: A similar but slightly more complex example is as follows. The vertex set is $Z_m \times Z_m$, where m is some positive integer, and Z_m is the additive group of integers modulo m. The degree is 4, and the vertex (x, y) has edges to $(x \pm y, y)$, and $(x, x \pm y)$, where all addition is done modulo m. Something which concerns me a little about this definition, but which I haven't resolved, is what happens when m is even and we choose $y = m/2$ so that, e.g., the vertices $(x + y, y)$ and $(x - y, y)$ coincide with one another. We would expect this duplication to have some effect on the expansion parameter, but I haven't thought through exactly what.

III. GRAPHS AND THEIR ADJACENCY MATRICES

How can we prove that a family of graphs is an expander? Stated another way, how does the expansion parameter $h(G)$ vary as the graph G is varied over all graphs in the family?

One way of tackling the problem of computing $h(G)$ is to do a brute force calculation of the ratio $|\partial S|/|S|$ for every subset S of vertices containing no more than half the vertices in the graph. Doing this is a time-consuming task, since if there are n vertices in the graph, then there are exponentially many such subsets S.

Problem: In general, how hard is it to find the subset S minimizing $|\partial S|/|S|$? Can we construct an **NP-Complete** variant of this problem? I don't know the answer to this question, and I don't know if anyone else does, either.

Fortunately, there is an extraordinarily beautiful approach to the problem of determining $h(G)$ which is far less computationally intensive. It involves the *adjacency matrix* $A(G)$ of the graph G. By definition, the rows and columns of the adjacency matrix are labelled by the vertices of V. For vertices v and w the entry $A(G)_{vw}$ is defined to be 1 if (v, w) is an edge, and 0 if it is not an edge.

It is a marvellous fact that properties of the *eigenvalue spectrum* of the adjacency matrix $A(G)$ can be used to

understand properties of the graph G. This occurs so frequently that we refer to the spectrum of $A(G)$ as *the spectrum of the graph* G. It is useful because the eigenvalue spectrum can be computed quickly, and certain properties, such as the largest and smallest eigenvalue, the determinant and trace, can be computed extremely quickly.

More generally, by recasting graphs in terms of adjacency matrices, we open up the possibility of using tools from linear algebra to study the properties of graphs. Although we're most interested in studying expanders, for the rest of this sectionI'm going to digress from the study of expanders, studying how the linear algebraic point of view can help us understand graphs, without worrying about how this connects to expanders. This digression is partially motivated by the fact that this is beautiful stuff (at least in my opinion), and is partially because our later discussion of expanders will be based on this linear algebraic point of view, and so it's good to get comfortable with this point of view.

The following exercise provides a good example of how graph properties can be related to the eigenvalues of the graph.

Exercise: Prove that if two graphs are isomorphic, then they have the same spectrum.

This result is often useful in proving that two graphs are not isomorphic: simply compute their eigenvalues, and show that they are different. A useful extension of the exercise is to find an example of two graphs which have the same spectra, but which are not isomorphic.

Note that the adjacency matrix may be considered as a matrix over any field, and the result of the exercise is true over any field. (I've often wondered if the converse is true, but don't know the answer.) Nonetheless, by and large, we'll consider the adjacency matrix as a matrix over the field R of real numbers. Assuming that G is an undirected graph, we see that $A(G)$ is a real symmetric matrix, and thus can be diagonalized. We will find it convenient to write the eigenvalues of a graph G in non-increasing order, as $\lambda_1(G) \geq \lambda_2(G) \geq \ldots \geq \lambda_n(G)$.

A fact we'll make a lot of use of is that when G is d-regular the largest eigenvalue of G is just d. To see this, note that the vector $\vec{1} \equiv (1, 1, \ldots, 1)$ is an eigenvector of G with eigenvalue d. To prove that d is the largest eigenvalue seems to be a little bit harder. We'll just sketch a proof. To prove this it is sufficient to show that $v^T A(G) v \leq d$ for all normalized vectors v. From the d-regularity of G it follows that $A(G)/d$ is a doubly stochastic matrix, i.e., has non-negative entries, and all rows and columns sum to one. A theorem of Birkhoff ensures that $A(G)/d$ can be written as a convex combination of permutation matrices, so $A(G) = d \sum_j p_j P_j$, where p_j are probabilities, and the P_j are permutation matrices. This gives $v^T A(G) v = d \sum_j p_j v^T P_j v$. But $v^T P_j v \leq 1$ for any permutation P_j, which gives the desired result.

The following proposition gives another example of the relationships one can find between a graph and its spectrum.

Proposition: A d-regular graph G is connected if and only if $\lambda_1(G) > \lambda_2(G)$.

Proof: The easy direction is the reverse implication, for which we prove the contrapositive, namely, that a d-regular disconnected graph has $\lambda_1(G) = \lambda_2(G)$. This follows by breaking G up into disconnected components G_1 and G_2, and observing that $A(G) = A(G_1) \oplus A(G_2)$, where $\oplus$ is the matrix direct sum. Since both G_1 and G_2 are d-regular it follows that they both have maximal eigenvalue d, and so d appears at least twice in the spectrum of $A(G)$.

At the moment, I don't see an easy way of proving the forward implication. One not very satisfying proof is to observe that $A(G)/d$ is the Markov transition matrix for a random walk on the graph, and that since the graph is connected, the random walk must converge to a unique distribution, which implies that in the limit of large n there can only be one vector v such that $(G^n/d^n)v = v$. This means that G^n's largest eigenvalue is non-degenerate, from which it follows that G's largest eigenvalue is non-degenerate. This is a sketch, but it can all be established rigorously with a little work and the aid of well-known theorems on Markov chains.

The proof sketched in the previous paragraph is not really satisfactory, since it involves an appeal to theorems which are in some sense less elementary than the result under discussion. Another possibility which I've explored but haven't made work with complete rigour is to investigate G^n/d^n more explicitly. With a little thought one can prove that the entry G^n_{vw} is just the number of paths between v and w of length n. Since G is connected, we'd expect in the limit of large n this number would be dominated by a term which does not depend on w, and would just scale like the total number of paths of length n starting at v (which is d^n), divided by the total number of possible destinations w, which is n, giving $G^n_{vw}/d^n \rightarrow 1/n$. (This would only be true if G has self-loops (v, v).) Of course, the matrix whose entries are all $1/n$ has a single eigenvalue 1, with all the rest 0, which would suffice to establish the theorem.

QED

Problem: How should we interpret the determinant of a graph? What about the trace?

Problem: If we consider $A(G)$ as a matrix over the field $Z_2 = \{0, 1\}$, then it is possible to define a matrix sum $G_1 + G_2$, whose adjacency matrix is just $A(G_1) + A(G_2)$, and a matrix product $G_1 \times G_2$ whose adjacency matrix is just $A(G_1)A(G_2)$. Many questions naturally suggest themseves: (1) when is there an edge between v and w in $G_1 + G_2$; (2) when is there an edge between v and w in $G_1 \times G_2$ (these first two questions are easy to answer); (3) for which graphs is $A(G)$ invertible, and thus a natural inverse graph G^{-1} exists; (4) how can we interpret the inverse graph; (5) when do two graphs commute?

Problem: Along similar lines to the previous problem, it's possible to define a tensor product of graphs. What

are the properties of the graph tensor product?

The ideas I've described in this sectionare examples of the important general principle that once you've defined a mathematical object, you should seek out alternate representations (or even just partial representations) of that object in terms of mathematical objects that you already understand. By recasting graphs as matrices, we open up the possibility of using all the tools of linear algebra to answer questions about graphs. This can work in one of two ways: we can ask a question about graphs, and try to see if it's possible to give a linear algebraic answer, or we can ask what implication known results of linear algebra have for graphs — what does the Gaussian elimination procedure correspond to, or the spectral decomposition, or two matrices commuting, or the wedge product, or whatever. Exploring such connections has the potential to greatly enrich both subjects.

IV. EXPANSION AND THE EIGENVALUE GAP

Let's return our attention to expander graphs, and see what the eigenvalues of a graph have to do with its expansion parameter. We define the *gap for the graph* G to be the difference $\Delta(G) \equiv \lambda_1(G) - \lambda_2(G)$ between the largest and second-largest eigenvalues. The expansion parameter and the gap are connected by the following theorem:

Theorem: The expansion parameter $h(G)$ for a d-regular graph G is related to the gap $\Delta(G)$ by:

$$\frac{\Delta(G)}{2} \leq h(G) \leq \sqrt{2d\Delta(G)}. \tag{5}$$

Thus, properties of the eigenvalue gap can be used to deduce properties of the expansion parameter. For example, if the eigenvalue gap for a family of d-regular graphs is bounded below by a positive constant, then the expansion parameter must also be bounded below by a positive constant, and so the family is an expander.

One reason for finding the connection between the gap and the expansion parameter interesting is that it is far easier to estimate the gap of an n by n matrix than it is to enumerate the exponentially many subsets S of the vertex set V, and compute $|\partial S|/|S|$ for each one.

Proof discussion: We already understand that $\lambda_1(G) = d$ for this graph, with eigenvector $\vec{1} = (1, 1, \ldots, 1)$. So we'll concentrate on trying to understand the behaviour of the second largest eigenvalue, $\lambda_2(G)$. The theorem tells us that the difference between d and $\lambda_2(G)$ is controlled both above and below by the expansion parameter $h(G)$.

How can we get control over the second largest eigenvalue of G? One way is to observe that $\lambda_2(G)$ is just the maximum of the expression $v^T Av/v^T v$, where A is the adjacency matrix of G, and we maximize over all vectors v orthogonal to the eigenvector $\vec{1}$. An encouraging fact is that this expression is quite easy to deal with, because the condition that v be orthogonal to $\vec{1}$ is actually equivalent to the sum of v's entries being equal to 0, so we have

$$\lambda_2(G) = \max_{v:\mathrm{tr}(v)=0} \frac{v^T Av}{v^T v}, \tag{6}$$

where $\mathrm{tr}(v)$ is just the sum of the entries of the vector v.

We're going to provide a *lower bound* on $\lambda_2(G)$ by simply guessing a good choice of v satisfying $\mathrm{tr}(v) = 0$, and using the fact that

$$\lambda_2(G) \geq \frac{v^T Av}{v^T v}. \tag{7}$$

To make a good guess, it helps to have a way of thinking about expressions like $v^T Av$, where $\mathrm{tr}(v) = 0$. A convenient way of thinking is to rewrite v as the difference of two disjoint probability distributions, p and q, i.e., $v = p - q$, where p and q are non-negative vectors each summing to 1, and with disjoint support. This results in terms like $p^T Aq$, which we can think of in terms of transition probabilities between q and p. This will allow us to apply the expansion properties of the graph.

Let's make these ideas a little more concrete. The key is to define $\vec{1}_S$ to be the vector whose entries are 1 on S, and 0 elsewhere, and to observe that

$$\vec{1}_S^T A \vec{1}_T = |E(S,T)|, \tag{8}$$

where $|E(S,T)|$ is the number of edges between the vertex sets S and T. This suggests that we should choose p and q in terms of vectors like $\vec{1}_S$, since it will enable us to relate expressions like $v^T Av$ to the sizes of various edge sets, which, in turn, can be related to the expansion parameter.

Suppose in particular that we choose

$$v = \frac{\vec{1}_S}{|S|} - \frac{\vec{1}_{\overline{S}}}{|\overline{S}|}. \tag{9}$$

This satisfies the condition $\mathrm{tr}(v) = 0$, and gives

$$v^T v = \frac{1}{|S|} + \frac{1}{|\overline{S}|} \tag{10}$$

and

$$v^T Av = \frac{1}{|S|^2} E(S,S) + \frac{1}{|\overline{S}|^2} E(\overline{S},\overline{S}) - \frac{2}{|S||\overline{S}|} E(S,\overline{S}) \tag{11}$$

The definition of an expander graph gives us control over $E(S,\overline{S})$, so it is convenient to rewrite $E(S,S)$ and $E(\overline{S},\overline{S})$ in terms of $E(S,\overline{S})$, using the d-regularity of the graph:

$$E(S,S) + E(S,\overline{S}) = d|S|; \;\; E(\overline{S},\overline{S}) + E(S,\overline{S}) = d|\overline{S}| \tag{12}$$

A straightforward substitution and a little algebra gives:

$$v^T Av = d\left(\frac{1}{|S|} + \frac{1}{|\overline{S}|}\right) - \left(\frac{1}{|S|} + \frac{1}{|\overline{S}|}\right)^2 E(S,\overline{S}). \tag{13}$$

Comparing with the earlier expression for the denominator $v^T v$, we obtain

$$\lambda_2(G) \geq d - \left(\frac{1}{|S|} + \frac{1}{|\overline{S}|}\right) E(S, \overline{S}). \quad (14)$$

Now choose S so that $E(S, \overline{S}) = h(G)|S|$, and $|S| \leq n/2$, giving after a little algebra

$$\lambda_2(G) \geq d - 2h(G), \quad (15)$$

and thus

$$\frac{\Delta(G)}{2} \leq h(G), \quad (16)$$

which was the first of the two desired inequalities in the theorem.

The proof of the second inequality is a little more complicated. Unfortunately, I haven't managed to boil the proof down to a form that I'm really happy with, and for this reason I won't describe the details. If you're interested, you should try to prove it yourself, or refer to the notes of Linial and Wigderson.

QED

Problem: Can we generalize this result so that it applies to a general undirected graph G, not just to d-regular graphs? Can we prove an analogous statement for directed graphs, perhaps in terms of singular values? Can we define a a generalized notion of "expansion" which can be applied to *any* symmetric matrix A with non-negative entries, and connect that notion of expansion to the eigenvalue gap? Can we generalize even further? What happens if we change the field over which the matrix is considered?

V. RANDOM WALKS ON EXPANDERS

Many applications of expanders involve doing a random walk on the expander. We start at some chosen vertex, and then repeatedly move to any one of the d neighbours, each time choosing a neighbour uniformly at random, and independently of prior choices.

To describe this random walk, suppose at some given time we have a probability distribution p describing the probability of being at any given vertex in the graph G. We then apply one step of the random walk procedure described above, i.e., selecting a neighbour of the current vertex uniformly at random. The updated probability distribution is easily verified to be:

$$p' = \frac{A(G)}{d} p. \quad (17)$$

That is, the Markov transition matrix describing this random walk is just $\hat{A}(G) \equiv A(G)/d$, i.e., up to a constant of proportionality the transition matrix is just the adjacency matrix. This relationship between the adjacency matrix and random walks opens up a whole new world of connections between graphs and Markov chains.

One of the most important connections is between the eigenvalues of Markov transition matrices and the rate at which the Markov chain converges to its stationary distribution. In particular, the following beautiful theorem tells us that when the uniform distribution is a stationary distribution for the chain, then the Markov chain converges to the uniform distribution exponentially quickly, at a rate determined by the second largest eigenvalue of M.

Exercise: Show that if M is a normal transition matrix for a Markov chain then $1 = \lambda_1(M) \geq \lambda_2(M) \geq \ldots$.

Theorem: Suppose M is a normal transition matrix for a Markov chain on n states, with the uniform distribution $u = \vec{1}/n$ as a stationary point, $Mu = u$. Then for any starting distribution p,

$$\|M^t p - u\|_1 \leq \sqrt{n} \lambda_2(M)^t, \quad (18)$$

where $\| \cdot \|_1$ denotes the l_1 norm.

The normality condition in this theorem may appear a little surprising. The reason it's there is to ensure that M can be diagonalized. The theorem can be made to work for general M, with the second largest eigenvalue replaced by the second largest singular value. However, in our situation M is symmetric, and thus automatically normal, and we prefer the statement in terms of eigenvalues, since it allows us to make a connection to the expansion parameter of a graph. In particular, when $M = \hat{A}(G)$ we obtain:

$$\|\hat{A}(G)^t p - u\|_1 \leq \sqrt{n} \left(\frac{\lambda_2(G)}{d}\right)^t. \quad (19)$$

Combining this with our earlier results connecting the gap to the expansion parameter, we deduce that

$$\|\hat{A}(G)^t p - u\|_1 \leq \sqrt{n} \left(1 - \frac{h(G)^2}{2d^2}\right)^t. \quad (20)$$

Thus, for a family of expander graphs, the rate of convergence of the Markov chain is exponentially fast in the number of time steps t.

Exercise: Suppose M is a transition matrix for a Markov chain. Show that the uniform distribution u is a stationary point point for the chain, i.e., $Mu = u$, if and only if M is doubly stochastic, i.e., has non-zero entries, and all rows and columns of the matrix sum to 1.

Proof: We start by working with the l_2 norm $\| \cdot \|_2$. Since $Mu = u$ we have $M^t u = u$, and so:

$$\|M^t p - u\|_2 = \|M^t (p - u)\|_2. \quad (21)$$

A computation shows that $p - u$ is orthogonal to u. But u is an eigenvector of M with the maximum eigenvalue, 1, and thus $p - u$ must lie in the span of the eigenspaces with eigenvalues $\lambda_2(M), \lambda_3(M), \ldots$. It follows that

$$\|M^t (p - u)\|_2 \leq \lambda_2(M)^t \|p - u\|_2 \leq \lambda_2(M)^t, \quad (22)$$

where we used the fact that $\|p - u\|_2 \leq 1$, easily established by observing that $\|p - u\|_2$ is convex in p, and thus

must be maximized at an extreme point in the space of probability distributions; the symmetry of u ensures that without loss of generality we may take $p = (1, 0, \ldots, 0)$. To convert this into a result about the l_1 norm, we use the fact that in n dimensions $\|v\|_1 \leq \sqrt{n}\|v\|_2$, and thus we obtain

$$\|M^t(p-u)\|_1 \leq \sqrt{n}\lambda_2(M)^t, \tag{23}$$

which was the desired result. **QED**

What other properties do random walks on expanders have? We now prove another beautiful theorem which tells us that they "move around quickly", in the sense that they are exponentially unlikely to stay for long within a given subset of vertices, B, unless B is very large.

More precisely, suppose B is a subset of vertices, and we choose some vertex X_0 uniformly at random from the graph. Suppose we use X_0 as the starting point for a random walk, $X_0, \ldots, X_t$, where X_t is the vertex after the tth step. Let $B(t)$ be the event that $X_j \in B$ for *all* j in the range $0, \ldots, t$. Then we will prove that:

$$\Pr(B(t)) \leq \left(\frac{\lambda_2(G)}{d} + \frac{|B|}{n}\right)^t \tag{24}$$

Provided $\lambda_2(G)/d + |B|/n < 1$, we get the desired exponential decrease in probability. For a family of expander graphs it follows that there is some constant $\epsilon > 0$ such that we get an exponential decrease for any B such that $|B|/n < \epsilon$. These results are special cases of the following more general theorem about Markov chains.

Theorem: Let X_0 be uniformly distributed on n states, and let $X_0, \ldots, X_t$ be a time-homogeneous Markov chain with transition matrix M. Suppose the uniform distribution u is a stationary point of M, i.e., $Mu = u$. Let B be a subset of the states, and let $B(t)$ be the event that $X_j \in B$ for all $j \in 0, \ldots, t$. Then

$$\Pr(B(t)) \leq \left(\lambda_2(M) + \frac{|B|}{n}\right)^t. \tag{25}$$

Proof: The first step in the proof is to observe that:

$$\Pr(B(t)) = \|(PMP)^t Pu\|_1, \tag{26}$$

where the operator P projects onto the vector space spanned by those basis vectors corresponding to elements of B. This equation is not entirely obvious, and proving it is a good exercise for the reader.

The next step is to prove that $\|PMP\| \leq \lambda_2(M) + |B|/n$, where the norm here is the operator norm. We will do this below, but note first that once this is done, the result follows, for we have

$$\Pr(B(t)) = \|(PMP)^t Pu\|_1 \leq \sqrt{n}\|(PMP)^t Pu\|_2 \tag{27}$$

by the standard inequality relating l_1 and l_2 norms, and thus

$$\Pr(B(t)) \leq \sqrt{n}\|PMP\|^t \|Pu\|_2, \tag{28}$$

by definition of the operator norm, and finally

$$\Pr(B(t)) \leq \left(\lambda_2(M) + \frac{|B|}{n}\right)^t, \tag{29}$$

where we used the assumed inequality for the operator norm, and the observation that $\|Pu\|_2 = \sqrt{|B|}/n \leq 1/\sqrt{n}$.

To prove the desired operator norm inequality, $\|PMP\| \leq \lambda_2(M) + |B|/n$, suppose v is a normalized state such that $\|PMP\| = |v^T PMPv|$. Decompose $Pv = \alpha u + \beta u_\perp$, where $u_\perp$ is a normalized state orthogonal to u. Since $\|Pv\|_2 \leq \|v\|_2 = 1$ we must have $|\beta| \leq 1$. Furthermore, multiplying $Pv = \alpha u + \beta u_\perp$ on the left by nu^T shows that $\alpha = nu^T Pv$. It follows that $|\alpha|$ is maximized by choosing v to be uniformly distributed over B, from which it follows that $|\alpha| \leq \sqrt{|B|}$. A little algebra shows that

$$v^T PMPv = \alpha^2 u^T Mu + \beta^2 u_\perp^T M u_\perp. \tag{30}$$

Applying $|\alpha| \leq \sqrt{|B|}$, $u^T Mu = u^T u = 1/n$, $|\beta| \leq 1$, and $u_\perp^T M u_\perp \leq \lambda_2(M)$ gives

$$v^T PMPv \leq \frac{|B|}{n} + \lambda_2(M), \tag{31}$$

which completes the proof. **QED**

VI. REDUCING THE NUMBER OF RANDOM BITS REQUIRED BY AN ALGORITHM

One surprising application of expanders is that they can be used to reduce the number of random bits needed by a randomized algorithm in order to achieve a desired success probability.

Suppose, for example, that we are trying to compute a function $f(x)$ that can take the values $f(x) = 0$ or $f(x) = 1$. Suppose we have a randomized algorithm $A(x, Y)$ which takes as input x and an m-bit uniformly distributed random variable Y, and outputs either 0 or 1. We assume that:

- $f(x) = 0$ implies $A(x, Y) = 0$ with certainty.
- $f(x) = 1$ implies $A(x, Y) = 1$ with probability at least $1 - p_f$.

That is, p_f is the maximal probability that the algorithm fails, in the case when $f(x) = 1$, but $A(x, Y) = 0$ is output by the algorithm.

An algorithm of this type is called a *one-sided* randomized algorithm, since it can only fail when $f(x) = 1$, not when $f(x) = 0$. I won't give any concrete examples of one-sided randomized algorithms here, but the reader unfamiliar with them should rest assured that they are useful and important — see, e.g., the book of Motwani and Raghavan (Cambridge University Press, 1995) for examples.

As an aside, the discussion of one-sided algorithms in this sectioncan be extended to the case of randomized algorithms which can fail when either $f(x) = 0$ or $f(x) = 1$. The details are a little more complicated, but the basic ideas are the same. This is described in Linial and Wigderson's lecture notes. Alternately, extending the discussion to this case is a good problem.

How can we descrease the probability of failure for a one-sided randomized algoerithm? One obvious way of decreasing the failure probability is to run the algorithm k times, computing $A(x, Y_0), A(x, Y_1), \ldots, A(x, Y_{k-1})$. If we get $A(x, Y_j) = 0$ for all j then we output 0, while if $A(x, Y_j) = 1$ for at least one value of J, then we output $f(x) = 1$. This algorithm makes use of km bits, and reduces the failure probability to at most p_f^k.

Expanders can be used to substantially decrease the number of random bits required to achieve such a reduction in the failure probability. We define a new algorithm A' as follows. It requires a d-regular expander graph G whose vertex set V contains 2^m vertices, each of which can represent a possible m-bit input y to $A(x, y)$. The modified algorithm A' works as follows:

- Input x.
- Sample uniformly at random from V to generate Y_0.
- Now do a $k-1$ step random walk on the expander, generating random variables $Y_1, \ldots, Y_{k-1}$.
- Compute $A(x, Y_0), \ldots, A(x, Y_{k-1})$. If any of these are 1, output 1, otherwise output 0.

We see that the basic idea of the algorithm is similar to the earlier proposal for running $A(x, Y)$ repeatedly, but the sequence of independent and uniformly distributed samples $Y_0, \ldots, Y_{k-1}$ is replaced by a random walk on the expander. The advantage of doing this is that only $m + k\log(d)$ random bits are required — m to sample from the initial uniform distribution, and then $\log(d)$ for each step in the random walk. When d is a small constant this is far fewer than the km bits used when we simply repeatedly run the algorithm $A(x, Y_j)$ with uniform and independently generated random bits Y_j.

With what probability does this algorithm fail? Define B_x to be the set of values of y such that $A(x, y) = 0$, yet $f(x) = 1$. This is the "bad" set, which we hope our algorithm will avoid. The algorithm will fail only if the steps in the random walk $Y_0, Y_1, \ldots, Y_{k-1}$ all fall within B_x. From our earlier theorem we see that this occurs with probability at most:

$$\left(\frac{|B_x|}{2^m} + \frac{\lambda_2(G)}{d}\right)^{k-1}. \tag{32}$$

But we know that $|B_x|/2^m \leq p_f$, and so the failure probability is at most

$$\left(p_f + \frac{\lambda_2(G)}{d}\right)^{k-1}. \tag{33}$$

Thus, provided $p_f + \lambda_2(G)/d < 1$, we again get an exponential decrease in the failure probability as the number of repetitions k is increased.

VII. CONCLUSION

These notes have given a pretty basic introduction to expanders, and there's much we haven't covered. More detail and more applications can be found in the online notes of Linial and Wigderson, or in the research literature. Still, I hope that these notes have given some idea of why these families of graphs are useful, and of some of the powerful connections between graph theory, linear algebra, and random walks.