EXPANDERS, EIGENVALUES, AND ALL THAT

NATI LINIAL

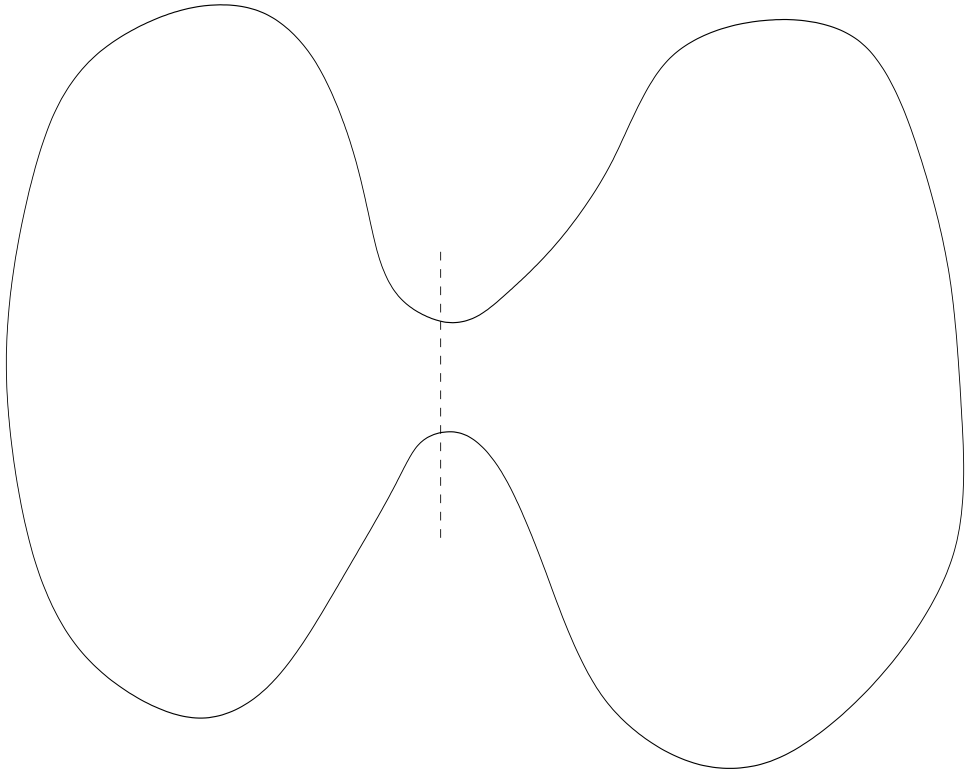
- Hebrew University
- Jerusalem, Israel
- http://www.cs.huji.ac.il/~nati/
- Visiting Microsoft Research.

- See also my lecture notes from a joint class with Avi Wigderson.

Expansion - A relative notion of connectivity

- In your discrete math and algorithms classes you must have encountered concepts of graph connectivity.
- Namely, how many vertices/edges must be removed to make the graph disconnected. (How small can S be so that $G \setminus S$ is disconnected).
- However, in graph connectivity you do not care whether the connected components of the graph $G \setminus S$ are small or large.
- In contrast, G may still have large expansion even though there is a small set S for which $G \setminus S$ is disconnected. Namely, if $G \setminus S$ consists of one very large component and several small ones.

- We only forbid bottlenecks. Namely, a small set S for which $G \setminus S$ has at least two large connected components.



The main questions

- What are expanders good for?
- Do they exist in abundance, or are they rare?
- How can you tell if a given graph is an expander?
- Can you explicitly construct them?
- How expanding can you get? (and does it matter)?

The formal definition

A graph $G = (V, E)$ is said to be ϵ-edge-expanding if for every partition of the vertex set V into X and $X^c = V \setminus X$, where X contains at most a half of the vertices, the number of cross edges

$$e(X, X^c) \geq \epsilon |X|.$$

In words: in every cut in G, the number of cut edges is at least proportionate to the size of the smaller side.

Do expanders actually exist?

It should be quite clear that for every finite connected graph G there is an $\epsilon > 0$, such that G is ϵ-edge-expanding. What is not obvious is whether we can stop ϵ from tending to zero when G gets large.

This, too, is not hard to do, provided that G has a lot of edges. So, the essence of the problem is whether we can keep ϵ bounded away from zero, while the vertex degrees remain bounded from above by a constant. (e.g. every vertex has exactly d neighbors, in which case we say that G is d-regular.)

Do expanders actually exist? (cont'd.)

Here is our first concrete nontrivial question about expanders: Do there exist an integer d and $\epsilon > 0$, and infinitely many graphs $G_n = (V_n, E_n)$ so that:

- G_n get arbitrarily large ($|V_n| \to \infty$).
- All graphs G_n are d-regular (every vertex has exactly d neighbors).
- All the G_n are ϵ-edge-expanding.

ANSWER: Yes. This can be shown using the probabilistic method.

Do expanders actually exist? (conclusion)

They do. In fact, they exist in abundance. They are the rule rather than the exception.
The probabilistic argument not only shows that a regular graph can be an expander. Rather, that almost every regular graph is an expander.
If so, why don't we see them all over the place?
Because we are too shortsighted and because our computational resources are too limited.

The computational hardness of expansion

Theorem 1. *The following computational problem is co-NP hard*
Input: A *graph* G, *and* $\epsilon > 0$.
Output: *Is* G *an* ϵ*-edge-expander?*

The probabilistic method as an observational aid

We are incapable of getting a good view of large graphs. The probabilistic method is our major tool for overcoming this deficiency.
It works particularly well when (as is often the case) a property that we seek is, in fact, very common.
The probabilistic method can help us detect such situations.
More refined applications of the method (e.g. The Lovász Local Lemma) can even help you find needles in haystacks.

But this seems rather bad....

In the present case, we are able to show that expanding graphs are all around us, but we are unable to even recognize such a graph (even if it falls on our head....)

... and besides: why do we really care?
Let's address this first.

What are expander graphs good for? I

They make excellent communication networks. When you need to design a communication network, it is very beneficial to avoid bottlenecks. Relatively simple heuristics provably work in this case and allow you to efficiently route messages among processors, reconfigure quickly when the communication requests change etc.

What are expander graphs good for? II

They yield good error-correcting codes. A bipartite graph $H = (A, B, E)$ that's a good expander, yields a good error-correcting code that comes with efficient encoding and decoding algorithms. Roughly, if A has n vertices, there is a $1:1$ correspondence between subsets of A and n-bit messages. The vertices in B correspond to parity check bits. It turns out that if H is a good expander, then not only is the resulting (linear) code good, it can be provably corrected by a simple belief-propagation scheme applied to H.

What are expander graphs good for? III

They offer good deterministic emulation of random behavior. Or, as we often say, they exhibit certain desired pseudo random properties. More on this - later.

What are expander graphs good for? IV

They exhibit many extremal (graph-theoretic and other) properties:

For example: It is known (Bourgain, Johnson-Lindenstrauss) that every n-point metric space can be embedded in $O(\log n)$ dimensions with metric distortion only $O(\log n)$. (The metric of) a constant-degree expander graph shows that this is tight. "The metric of an expander is as non-Euclidean as it gets".

More on the role of extremal problems in discrete math - later.

What are expander graphs good for? V

A bridge between computer science and mathematics: The explicit construction of expander graphs helps recruit a lot of deep mathematics into the area. Expander graphs were recently used in resolving several old problems in mathematics....

How to save on random bits with expanders

Suppose that you have a randomized algorithm that uses k random bits and succeeds with probability $\frac{3}{4}$. If you can tolerate only error rates $< \delta$. What can you do to reach such a low probability of error?

The natural answer would be to repeat the experiment enough times to bring the probability of failure below δ. To carry out each single run of the test you must generate k random bits for a total of $k \log \frac{1}{\delta}$ random bits. So in this straightforward solution, you will generate k random bits each time you want to reduce the error probability by a factor of $\frac{1}{4}$.

Seems unavoidable, right?

Well, expander graphs can often be used to do things that seem "too good to be true", and this is one of them.

Randomized decision algorithms - A quick refresher

A language $\mathcal{L}$ is a collection of strings. In the decision problem for $\mathcal{L}$, we are given a string x and we have to determine whether x belongs to $\mathcal{L}$ or not.

What is a randomized algorithm that uses k random bits and has success probability $\geq \frac{3}{4}$? This is a (computationally feasible) boolean function $f(\cdot, \cdot)$. The first variable is x. The second one is r, a k-bit string. When $f(x, r) = 1$ we know with certainty that x is in $\mathcal{L}$. However, $f(x, r) = 0$ only means that x is apparently not in $\mathcal{L}$.

Randomized decision algorithms (cont'd.)

So, if we ever find an r for which $f(x,r) = 1$, our search is over, and we know that x is in $\mathcal{L}$. However, if we repeatedly sample r's with $f(x,r) = 0$, we are collecting more and more statistical evidence, that apparently x is not in $\mathcal{L}$.
In other words, if x is not in $\mathcal{L}$, then $f(x,r) = 0$ for every binary string r. However, if x is in $\mathcal{L}$, then at least $\frac{3}{4}$ of all strings r are witness to that. Namely, if $x \in \mathcal{L}$, then there is a subset W comprising at least $\frac{3}{4}$ of the 2^k binary strings r of length k such that $f(x,r) = 1$.

The trouble is, we only know that W is large, but have no idea how to search for its members.

A classical example of a randomized decision algorithms Primality testing.

Here $\mathcal{L}$ is the set of all composite (=non-prime) integers.
Fermat's Little Theorem says that if n is prime and $1 \leq a < n$, then $a^{n-1} - 1$ is divisible by n.
So if n is your input and if you found an integer a for which $a^{n-1} - 1$ is indivisible by n, then a is a witness to the fact that n is composite.

(Warning: Here I am oversimplifying things a bit....)
Usually, if n is composite, then at least a half of the integers $1 \leq a < n$ are witnesses to that.

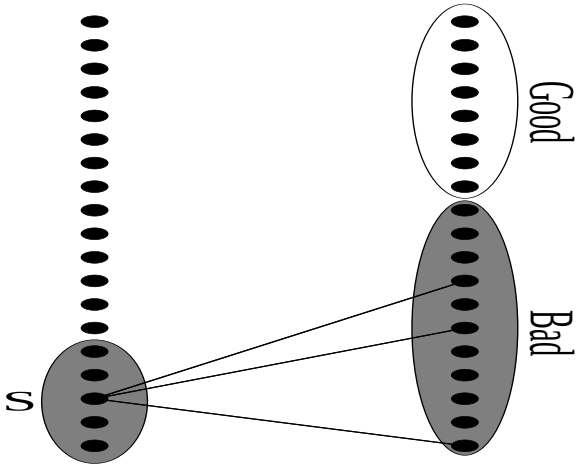
Good. There are lots of witnesses, but where are they and how will we find even one?

How to hit a hiding elephant?

In some huge universe (the collection of all 2^k k-bit strings; the range $1 \leq a < n$) at least $\frac{3}{4}$ of the elements are good. We need to find at least one good element, but we must succeed with overwhelming probability. If we repeatedly sample the space, say m times, then our probability of failure will indeed be very low, only $(\frac{1}{4})^m$. However, to do this, we will have to generate many random bits at each of the m times that we run our statistical test. It should seem quite convincing (but this impression is false...) that this cannot be achieved with fewer random bits. (Why we want to save on randomness - in a second).

How to hit a hiding elephant? (cont'd.)

Let us reformulate the expansion property for a bipartite $G = (L, R, E)$ graph as follows. Fix any set $B \subseteq R$ of "bad" vertices on the right side of G. If B is not too big relative to R, then the following set $S \subseteq L$ is tiny. Namely, S is the set of those vertices $x \in L$ on the left side all of whose neighbors belong to B.



Hiding elephants III

So here is an efficient way of finding at least one member in W with overwhelming probability. We use a bipartite expander $H = (L, R, E)$ where R is identified with the set of random strings r.
The algorithm is extremely simple: We sample a random vertex v in L (for this we need to generate only $\log |L|$ random bits) and go over the list of x's neighbors in R. Remember that these vertices correspond to some strings r. We run the randomized algorithm with these strings.
Again, since W is large, its complement $B := R \setminus W$ is not too large.
As we said above, there is only a tiny set S of vertices v which would fail us. Namely, only those all neighbors of which are in the bad set B. With proper choice of parameters we can reduce the probability of failure to the desired level without having to generate lots of random bits.

And why try to save on randomness?

Two main reasons:

- It is hard to model the properties of physical random sources.
- (The real reason) It's a fundamental question when randomness is inherently helpful in computation. (Likewise with quantum-based randomness etc.)

The eigenvalue connection

This starts with a seemingly very naive idea.

A standard way of representing a graph G is through its adjacency matrix $A = A_G$. Namely, $a_{ij} = 1, 0$ according to whether the i-th and j-th vertex are adjacent or not.
Well, we know what to do with matrices - compute eigenvectors, eigenvalues etc.
Seems pretty mechanical and unimaginative, ha?

Reminder - Eigenvalues, Rayleigh quotients etc.

Here are some facts from your (advanced) linear algebra classes:
If A is a real symmetric matrix, then all its eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ are real.
If A is a matrix of a d-regular graph, then $\lambda_1 = d$, and the corresponding eigenvector is $\mathbf{1}$, the all-ones vector.

$$\lambda_2 = \max \frac{x^t A x}{\|x\|^2}$$

where the maximum is over vectors x that are perpendicular to $\mathbf{1}$. (This expression is called a Rayleigh quotient).

What does this have to do with expansion?

Let $G = (V, E)$ be an n-vertex d-regular graph, and let $S \subseteq V$ be a set with $\leq n/2$ vertices. We'd like to know whether there are many edges connecting S to $V \setminus S$. Here is a way of deriving this information using Rayleigh quotients. Let

$$x = |V \setminus S|\chi_S - |S|\chi_{V \setminus S}$$

where χ_Y is the characteristic function of the set Y (i.e. $\chi_Y(u) = 1, 0$ depending on whether u is in Y or not.)

Rayleigh quotients and expansion

Here are some easy facts:

- x is perpendicular to $\mathbf{1}$ (obvious).
- $\|x\|^2 = n|S|(n - |S|)$ (one line of calculation).
- $x^t A x$ depends only (and simply) on the number of edges between S and its complement. (a three-line calculation).

So this is telling us that

A large spectral gap implies high expansion, and in fact, expansion is just a special case of spectral gap between $\lambda_1 = d$ and λ_2. (But wait till you see the reverse implication).

The second eigenvector reveals bottlenecks in G: This is only the beginning of another fascinating story that we won't touch on that goes under "the nodal region theorem".
(See J. Friedman's work on the graph-theoretic version of this theorem.)

A tip of something much larger...

We won't have enough time to discuss this, but the things that we do here have close connections with

Geometry, where expansion goes under the name of Cheeger's constant.

Mathematical analysis, with the adjacency matrix of G being a close relative of the Laplace operator.

Representation theory....

Probability Theory: rapidly mixing Markov Chains. The spectral gap of a Markov Chain's transition matrix controls the speed at which the chain converges to its limit distribution.

Spectral gap, expansion et. al.

What should come as a bigger surprise is that there is also an implication in the opposite direction. Namely, expansion implies spectral gap.

Why is this surprising? Spectral gap means a small Rayleigh quotient for all (uncountably many) vectors $x \perp \mathbf{1}$, whereas expansion means a small Rayleigh quotient for a finite number of such x. How can you draw such a strong conclusion from such a weak assumption?

...and yet...

Such phenomena were previously known in differential geometry (Cheeger's Theorem).
A conceivable argument in favor of this claim: If the Rayleigh quotient is a slowly changing function of x, and if the discrete set on which it's known to be small is dense enough, then we can argue by some kind of continuity.

Not so fast....

Indeed, Cheeger's proof can be emulated in graph-theoretic terms (N. Alon). The proof is quite difficult and the bounds are pretty weak.

This is unavoidable, though. A graph can be an excellent expander and still have only a very meager spectral gap.

For example: Take the best $(\frac{d}{10})$-regular expander, and fill in the "missing" vertex degree as a collection of disjoint cliques (complete graphs) of size $\frac{9d}{10}+1$ each. The expansion can only go up from this addition, but the spectral gap is very small. To see this, I should show you an $x \perp \mathbf{1}$ for which the Rayleigh quotient is close to G. Indeed, take $x = \chi_Y - \chi_Z$ where Y and Z are two of these cliques.

Is there a combinatorial equivalent to spectral gap?

Granted, it is fairly surprising that expansion (a combinatorial condition) and spectral gap (a linear-algebraic condition) are qualitatively equivalent. However, for a long time I was wondering if the spectral gap can also be quantitatively captured in combinatorial terms.
This was recently solved in joint work with Yonatan Bilu. It is based on the notion of discrepancy.

Recall what we said that expanders often behave as if they were random graphs. One major realization of this principle is:

Theorem 2 (The expander mixing lemma). *Let $G = (V, E)$ be a d-regular graph on n vertices, and let $\lambda = \lambda_2(G)$ be the second largest eigenvalue of (the adjacency matrix of) G.*
Then for any two disjoint sets of vertices $X, Y \subseteq V$, the number of edges connecting X and Y satisfies

$$|e(X,Y) - \frac{d}{n}|X||Y|| \leq \lambda\sqrt{|X||Y|}$$

Discrepancy and spectral gap are essentially equivalent

Theorem 3 (Yonatan Bilu, L.). *Let $G = (V, E)$ be a d-regular graph on n vertices and suppose that for any two disjoint sets of vertices $X, Y \subseteq V$, the number of edges connecting X and Y satisfies*

$$|e(X,Y) - \frac{d}{n}|X||Y|| \leq \alpha\sqrt{|X||Y|}$$

for some $\alpha > 0$. Then the second eigenvalue of G is at most $O(\alpha \log(\frac{d}{\alpha}))$. The bound is tight

What's a "large" spectral gap?

An instinct that most combinatorialists have is to ask about any new parameter they learn "how large can it be?", "how small can it be?" (pretty childish, if you ask me...).
So, in the spirit of this great tradition, we ask:
Problem 1. *How small can λ_2 be in a d-regular graph? (i.e., how large can the spectral gap get)?*

This was answered as follows:
Theorem 4 (Alon, Boppana).

$$\lambda_2 \geq 2\sqrt{d-1} - o(1)$$

The largest possible spectral gap and the number $2\sqrt{d-1}$

Here is one good approach to extremal combinatorics and questions of the form "How small can this get, how large can that get". Guess where the extremum is attained (the ideal example), and show that there are no better instances.
What, then, is the ideal expander?
Funny answer: It's the infinite d-regular tree. Do you see why?
In fact, something like eigenvalues (spectrum) can also be defined for infinite graphs. It turns out that the supremum of the spectrum for the d-regular infinite tree is....

$$2\sqrt{d-1}$$

A recurring theme - Finite analogues for infinite graphs

Another instinct that comes with asking extremal problems:
When you have established a bound you always ask whether it is tight. So let us do it.

Problem 2. *Are there d-regular graphs with second eigenvalue*

$$\lambda_2 \leq 2\sqrt{d-1} \quad ?$$

When such graphs exist, they are called Ramanujan Graphs.

What we know about Ramanujan Graphs

Margulis; Lubotzky-Phillips-Sarnak: d-regular Ramanujan Graphs exist when $d-1$ is a prime power. The construction is easy, but the proof uses a lot of heavy mathematical machinery.

Friedman: If you are willing to settle for $\lambda_2 \leq 2\sqrt{d-1}+\epsilon$, they exist. Moreover, almost every d-regular graph satisfies this condition.

Experimentally, Novikov-Sarnak: For n large, about 52 percent of all d-regular graphs are Ramanujan. (No proof for this in sight...)

So where does this leave us?
Some open problems and current directions

- Can we construct d-regular Ramanujan Graphs for every d? Currently no one seems to know. $d = 7$ is the first unknown case.

- Can we find elementary methods of proof for graphs with large spectral gap (or even Ramanujan)? Some recent progress: Zigzag products (see below), and constructions based on random lifts of graphs (Bilu-L.) can get $\lambda_2 \leq O(\sqrt{d}\log^{3/2} d)$.

- Calculating actual expansion rates - Recent work by London-L. on Margulis's Expander.

- Expansion beyond the eigenvalue bounds: In general, the best expansion rate for a d-regular graph is about $\frac{d}{2}$. That this is tight follows easily when we consider cuts corresponding to sets of about $\frac{n}{2}$ vertices. However, small sets in random graphs have expansion rate $d - 1 - o(1)$, and this difference is very significant for many applications. Can we find explicit constructions of graphs with this behavior? The zig-zag method by Reingold-Vadhan-Wigderson goes a long way in this direction.

- The equivalence between expansion and eigenvalue separation gives at least some means of estimating the expansion ratio of a graph. Currently, the best approach to this problem combines combinatorics with geometry - finite metric spaces [LR], [LLR]. The current world record is in [ARV], and we still do not know how far it's possible to advance here.

- Local-global description of graphs. How can a large graph look at the micro level? How does this affect the graph's global properties?