Computational Complexity: A Modern Approach

Draft of a book: Dated January 2007
Comments welcome!

Sanjeev Arora and Boaz Barak
Princeton University
complexitybook@gmail.com

This is an Internet draft. Some chapters are more finished than others. References and attributions are very preliminary and we apologize in advance for any omissions (but hope you will nevertheless point them out to us).

Please send us bugs, typos, missing references or general comments to complexitybook@gmail.com — **Thank You!!**

Chapter 21

Pseudorandom constructions: expanders and extractors

"How hard could it be to find hay in a haystack?"
Howard Karloff

The probabilistic method is a powerful method to show the existence of objects (e.g., graphs, functions) with certain desirable properties. We have already seen it used in Chapter 6 to show the existence of functions with high circuit complexity, in Chapter 19 to show the existence of good error correcting codes, and in several other places in this book. But sometimes the mere *existence* of an object is not enough: we need an *explicit* and *efficient* construction. The paradigm of explicitly constructing objects with pseudorandom properties has been extremely useful in complexity theory and computer science at large. One example is the constructions of error correcting codes shown in Chapter 19. In this chapter we will see two more examples.

Expander graphs are graphs with certain "pseudorandom" properties that have had a vast number of applications, ranging from fast sorting networks, to counterexamples in metric space theory, to proving the **PCP** Theorem. In Section 21.1 we lay the groundwork for defining expanders, showing how random walks on graphs can be analyzed in terms of the adjacency matrix's eigenvalues. Then, in Section 21.2, we give two equivalent definitions for expander graphs and show their application for a randomness-efficient error reduction of probabilistic algorithms. In Section 21.3 we show an explicit construction of expander graphs. Finally, in Section 21.4, we use this construction to show a *deterministic* logspace algorithm for undirected graph connectivity.

Our second example of an explicit construction concerns the following issue: while randomized algorithms are modeled using a sequence of unbiased and independent coin tosses, real-life randomness sources are imperfect and have correlations and biases. In Section 21.5 we define *randomness extractors*— algorithms to extract (uncorrelated, unbiased) random bits from any *weakly random* source— and give explicit constructions for them. Then, in Section 21.6 we use extractors to derandomize probabilistic logspace computations, albeit at the cost of some increase in the space requirement. We emphasize that in contrast to the results of chapters 20 and 19, this derandomization (as well as all the other results of this chapter) is *unconditional* and uses no unproven assumptions.

Both the constructions and analysis of this chapter are somewhat involved. You might wonder why should coming up with explicit construction be so difficult. After all, a proof of existence via the probabilistic method shows not only that an object with the desired property exists but in fact the vast majority of objects have the property. As Karloff said (see quote above), how hard can it be to find a single one? But perhaps it's not so surprising this task is so difficult: after all, we know that almost all Boolean functions have exponential circuit complexity, but finding even a single one in **NP** with this property will show that $\mathbf{P} \neq \mathbf{NP}$!

21.1 Random walks and eigenvalues

In this section we study random walks on (undirected regular) graphs. We will see that we can use elementary linear algebra to relate parameters of the graph's adjacency matrix to the behavior of the random walk on that graph. As a corollary we obtain the proof of correctness for the random-walk space-efficient algorithm for undirected connectivity described in Section 7.7. **Note:** We restrict ourselves here to *regular* graphs, in which every vertex has the same degree. However, we do allow our graphs to have self-loops and parallel edges. Most of the definitions and results can be suitably generalized to undirected graphs that are not regular.

21.1.1 Distributions as vectors and the parameter $\lambda(G)$.

Let G be a d-regular n-vertex graph and let $\mathbf{p}$ be some probability distribution over the vertices of G. We can think of $\mathbf{p}$ as a (column) vector in $\mathbb{R}^n$ where $\mathbf{p}_i$ is the probability that vertex i is obtained by the distribution. Note that the L_1-norm of $\mathbf{p}$ (see Note 21.1), defined as $|\mathbf{p}|_1 = \sum_{i=1}^n |\mathbf{p}_i|$, is equal to 1. (In this case the absolute value is redundant since $\mathbf{p}_i$ is always non-negative.) Now let $\mathbf{q}$ represent the distribution of the following random variable: choose a vertex i in G according to $\mathbf{p}$, then take a random neighbor of i in G. We can easily compute $\mathbf{q}$, since the probability $\mathbf{q}_j$ that j is chosen is equal to the sum over all of j's neighbors i of the probability $\mathbf{p}_i$ that i is chosen times $1/d$ (where $1/d$ is the probability that, conditioned on i being chosen, the walk moves to $\mathbf{q}$). Thus $\mathbf{q} = A\mathbf{p}$, where $A = A(G)$ is the *normalized adjacency matrix* of G. That is, for every two vertices i, j, $A_{i,j}$ is equal to the number of edges between i and j divided by d. Note that A is a symmetric matrix,[1] with all its entries between 0 and 1, and the sum of entries in each row and column is exactly one. Such a matrix is called a symmetric *stochastic* matrix.

Let $\{\mathbf{e}^i\}_{i=1}^n$ be the *standard basis* of $\mathbb{R}^n$ (i.e. $\mathbf{e}^i$ has 1 in the i^{th} coordinate and zero everywhere else). Then for every $i \in [n]$, $A^\ell \mathbf{e}^i$ represents the distribution X_ℓ of the last step in a ℓ-step random walk from the i^{th} vertex. This already suggests that considering the adjacency matrix of a graph G could be very useful in analyzing random walks on G.

Definition 21.2 (The parameter $\lambda(G)$.)
Denote by $\mathbf{1}$ the vector $(1/n, 1/n, \ldots, 1/n)$ corresponding to the uniform distribution. Denote by $\mathbf{1}^\perp$ the set of vectors perpendicular to $\mathbf{1}$ (i.e., $\mathbf{v} \in \mathbf{1}^\perp$ if $\langle \mathbf{v}, \mathbf{1} \rangle = (1/n)\sum_i \mathbf{v}_i = 0$).
The parameter $\lambda(A)$, denoted also as $\lambda(G)$, is the maximum value of $\|A\mathbf{v}\|_2$ over all vectors $\mathbf{v} \in \mathbf{1}^\perp$ with $\|\mathbf{v}\|_2 = 1$.

The value $\lambda(G)$ is called the *second largest eigenvalue* of G. The reason is that since A is a symmetric matrix, we can find an orthogonal basis of eigenvectors $\mathbf{v}^1, \ldots, \mathbf{v}^n$ with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$ which we can sort to ensure $|\lambda_1| \geq |\lambda_2| \ldots \geq |\lambda_n|$. Note that $A\mathbf{1} = \mathbf{1}$. Indeed, for every i, $(A\mathbf{1})_i$ is equal to the inner product of the i^{th} row of A and the vector $\mathbf{1}$ which (since the sum of entries in the row is one) is equal to $1/n$. Thus, $\mathbf{1}$ is an *eigenvector* of A with the corresponding eigenvalue equal to 1. One can show that a symmetric stochastic matrix has all eigenvalues with absolute value at most 1 (see Exercise 21.3) and hence we can assume $\lambda_1 = 1$ and $\mathbf{v}^1 = \mathbf{1}$. Also, because $\mathbf{1}^\perp = \mathrm{Span}\{\mathbf{v}^2, \ldots, \mathbf{v}^n\}$, the value λ above will be maximized by (the normalized version of) $\mathbf{v}^2$, and hence $\lambda(G) = |\lambda_2|$. The quantity $1 - \lambda(G)$ is called the *spectral gap* of the graph. We note that some texts use *un-normalized* adjacency matrices, in which case $\lambda(G)$ is a number between 0 and d and the spectral gap is defined to be $d - \lambda(G)$.

One reason that $\lambda(G)$ is an important parameter is the following lemma:

[1] A matrix A is *symmetric* if $A = A^\dagger$, where $A^\dagger$ denotes the *transpose* of A. That is, $(A^\dagger)_{i,j} = A_{j,i}$ for every i, j.

Note 21.1 (L_p Norms)

A *norm* is a function mapping a vector $\mathbf{v}$ into a real number $\|\mathbf{v}\|$ satisfying **(1)** $\|\mathbf{v}\| \geq 0$ with $\|\mathbf{v}\| = 0$ if and only $\mathbf{v}$ is the all zero vector, **(2)** $\|\alpha\mathbf{v}\| = |\alpha| \cdot \|\mathbf{v}\|$ for every $\alpha \in \mathbb{R}$, and **(3)** $\|\mathbf{v} + \mathbf{u}\| \leq \|\mathbf{v}\| + \|\mathbf{u}\|$ for every vector $\mathbf{u}$. The third inequality implies that for every norm, if we define the *distance* between two vectors $\mathbf{u}$,$\mathbf{v}$ as $\|\mathbf{u} - \mathbf{v}\|$ then this notion of distance satisfies the triangle inequality.

For every $\mathbf{v} \in \mathbb{R}^n$ and number $p \geq 1$, the L_p *norm* of $\mathbf{v}$, denoted $\|\mathbf{v}\|_p$, is equal to $(\sum_{i=1}^n |\mathbf{v}_i|^p)^{1/p}$. One particularly interesting case is $p = 2$, the so-called *Euclidean norm*, in which $\|\mathbf{v}\|_2 = \sqrt{\sum_{i=1}^n \mathbf{v}_i^2} = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$. Another interesting case is $p = 1$, where we use the single bar notation and denote $|\mathbf{v}|_1 = \sum_{i=1}^n |\mathbf{v}_i|$. Another case is $p = \infty$, where we denote $\|\mathbf{v}\|_\infty = \lim_{p\to\infty} \|\mathbf{v}\|_p = \max_{i\in[n]} |\mathbf{v}_i|$.

Some relations between the different norms can be derived from the *Hölder inequality*, stating that for every p, q with $\frac{1}{p} + \frac{1}{q} = 1$, $\|\mathbf{u}\|_p \|\mathbf{v}\|_q \geq \sum_{i=1}^n |\mathbf{u}_i \mathbf{v}_i|$. To prove it, note that by simple scaling, it suffices to consider norm one vectors, and so it enough to show that if $\|\mathbf{u}\|_p = \|\mathbf{v}\|_q = 1$ then $\sum_{i=1}^n |\mathbf{u}_i||\mathbf{v}_i| \leq 1$. But $\sum_{i=1}^n |\mathbf{u}_i||\mathbf{v}_i| = \sum_{i=1}^n |\mathbf{u}_i|^{p(1/p)} |\mathbf{v}_i|^{q(1/q)} \leq \sum_{i=1}^n \frac{1}{p}|\mathbf{u}_i|^p + \frac{1}{q}|\mathbf{v}_i|^q = \frac{1}{p} + \frac{1}{q} = 1$, where the last inequality uses the fact that for every $a, b > 0$ and $\alpha \in [0, 1]$, $a^\alpha b^{1-\alpha} \leq \alpha a + (1 - \alpha)b$.

The Hölder inequality implies the following relations between the L_2, L_1 and L_∞ norms of every vector (Exercise 21.1):

$$|\mathbf{v}|_1 / \sqrt{n} \leq \|\mathbf{v}\|_2 \leq \sqrt{|\mathbf{v}|_1 \|\mathbf{v}\|_\infty} \tag{1}$$

Lemma 21.3 *Let G ba an n-vertex regular graph and $\mathbf{p}$ a probability distribution over G's vertices, then*

$$\|A^\ell \mathbf{p} - \mathbf{1}\|_2 \leq \lambda^\ell$$

PROOF: By the definition of $\lambda(G)$, $\|A\mathbf{v}\|_2 \leq \lambda\|\mathbf{v}\|_2$ for every $\mathbf{v} \perp \mathbf{1}$. Note that if $\mathbf{v} \perp \mathbf{1}$ then $A\mathbf{v} \perp \mathbf{1}$ since $\langle \mathbf{1}, A\mathbf{v}\rangle = \langle A^\dagger \mathbf{1}, \mathbf{v}\rangle = \langle \mathbf{1}, \mathbf{v}\rangle = 0$ (as $A = A^\dagger$ and $A\mathbf{1} = \mathbf{1}$). Thus A maps the space $\mathbf{1}^\perp$ to itself and since it shrinks any member of this space by at least λ, $\lambda(A^\ell) \leq \lambda(A)^\ell$. (In fact, using the eigenvalue definition of λ, it can be shown that $\lambda(A^\ell) = \lambda(A)^\ell$.)

Let $\mathbf{p}$ be some vector. We can break $\mathbf{p}$ into its components in the spaces parallel and orthogonal to $\mathbf{1}$ and express it as $\mathbf{p} = \alpha\mathbf{1} + \mathbf{p}'$ where $\mathbf{p}' \perp \mathbf{1}$ and α is some number. If $\mathbf{p}$ is a probability distribution then $\alpha = 1$ since the sum of coordinates in $\mathbf{p}'$ is zero. Therefore,

$$A^\ell \mathbf{p} = A^\ell(\mathbf{1} + \mathbf{p}') = \mathbf{1} + A^\ell \mathbf{p}'$$

Since $\mathbf{1}$ and $\mathbf{p}'$ are orthogonal, $\|\mathbf{p}\|_2^2 = \|\mathbf{1}\|_2^2 + \|\mathbf{p}'\|_2^2$ and in particular $\|\mathbf{p}'\|_2 \leq \|\mathbf{p}\|_2$. Since $\mathbf{p}$ is a probability vector, $\|\mathbf{p}\|_2 \leq |\mathbf{p}|_1 \cdot 1 \leq 1$ (see Note 21.1). Hence $\|\mathbf{p}'\|_2 \leq 1$ and

$$\|A^\ell \mathbf{p} - \mathbf{1}\|_2 = \|A^\ell \mathbf{p}'\|_2 \leq \lambda^\ell$$

■

It turns out that every connected graph has a noticeable spectral gap:

Lemma 21.4 ([**AlonSu00**]) *If G is a regular connected graph with self-loops at each vertex, then $\lambda(G) \leq 1 - \frac{1}{12n^2}$.*

PROOF: Let $\epsilon = \frac{1}{6n^2}$, let $\mathbf{u} \perp \mathbf{1}$ be a unit vector and let $\mathbf{v} = A\mathbf{u}$. We need to prove that $\|\mathbf{v}\|_2 \leq 1 - \epsilon/2$ and for this it suffices to prove that $1 - \|\mathbf{v}\|_2^2 \geq \epsilon$. (Indeed, if $\|\mathbf{v}\|_2 > 1 - \epsilon/2$ then $\|\mathbf{v}\|_2^2 > 1 - \epsilon$ and hence $1 - \|\mathbf{v}\|_2^2 < \epsilon$.) Since $\|\mathbf{u}\|_2 = 1$, $1 - \|\mathbf{v}\|_2^2 = \|\mathbf{u}\|_2^2 - \|\mathbf{v}\|_2^2$. We claim that this is equal to $\sum_{i,j} A_{i,j}(\mathbf{u}_i - \mathbf{v}_j)^2$ where i, j range from 1 to n. Indeed,

$$\sum_{i,j} A_{i,j}(\mathbf{u}_i - \mathbf{v}_j)^2 = \sum_{i,j} A_{i,j}\mathbf{u}_i^2 - 2\sum_{i,j} A_{i,j}\mathbf{u}_i\mathbf{v}_j + \sum_{i,j} A_{i,j}\mathbf{v}_j^2 =$$
$$\|\mathbf{u}\|_2^2 - 2\langle A\mathbf{u}, \mathbf{v}\rangle + \|\mathbf{v}\|_2^2 = \|\mathbf{u}\|_2^2 - 2\|\mathbf{v}\|_2^2 + \|\mathbf{v}\|_2^2 = \|\mathbf{u}\|_2^2 - \|\mathbf{v}\|_2^2\,,$$

where these equalities are due to the sum of each row and column in A equalling one, and to $\|\mathbf{v}\|_2^2 = \langle \mathbf{v}, \mathbf{v}\rangle = \langle A\mathbf{u}, \mathbf{v}\rangle = \sum_{i,j} A_{i,j}\mathbf{u}_i\mathbf{v}_j$.

Thus it suffices to show $\sum_{i,j} A_{i,j}(\mathbf{u}_i - \mathbf{v}_j)^2 \geq \epsilon$. Since $\mathbf{u}$ is a unit vector with coordinates summing to zero, there must exist vertices i, j such that $\mathbf{u}_i > 0, \mathbf{u}_j < 0$ and at least one of these coordinates has absolute value $\geq \frac{1}{\sqrt{n}}$, meaning that $\mathbf{u}_i - \mathbf{u}_j \geq \frac{1}{\sqrt{n}}$. Furthermore, because G is connected there is a path between i and j containing at most $D + 1$ vertices (where D is the *diameter*[2] of G). By renaming vertices, let's assume that $i = 1$, $j = D + 1$ and the coordinates $2, 3, \ldots, D$ correspond to the vertices on this path in order. Now, we have

$$\tfrac{1}{\sqrt{n}} \leq \mathbf{u}_1 - \mathbf{u}_{D+1} = (\mathbf{u}_1 - \mathbf{v}_1) + (\mathbf{v}_1 - \mathbf{u}_2) + \ldots + (\mathbf{v}_D - \mathbf{u}_{D+1}) \leq$$
$$\mathbf{u}_1 - \mathbf{u}_{D+1} = |\mathbf{u}_1 - \mathbf{v}_1| + |\mathbf{v}_1 - \mathbf{u}_2| + \ldots + |\mathbf{v}_D - \mathbf{u}_{D+1}| \leq$$
$$\sqrt{(\mathbf{u}_1 - \mathbf{v}_1)^2 + (\mathbf{v}_1 - \mathbf{u}_2)^2 + \ldots + (\mathbf{v}_D - \mathbf{u}_{D+1})^2}\sqrt{2D+1}\,, \quad (2)$$

where the last inequality follows by relating the L_2 and L_1 norms of the vector $(\mathbf{u}_1 - \mathbf{v}_1, \mathbf{v}_1 - \mathbf{u}_2, \ldots, \mathbf{v}_D - \mathbf{u}_{D+1})$; see Note 21.1. But this means that

$$\sum_{i,j} A_{i,j}(\mathbf{u}_i - \mathbf{v}_j)^2 \geq \tfrac{1}{dn(2D+1)}\,, \quad (3)$$

[2]The *diameter* of a graph G is the maximum distance (i.e., length of shortest path) between any pair of vertices in G. Note that the diameter of a connected n-vertex graph is always at most $n - 1$.

since the left hand side of (3) is a sum of non-negative terms and (2) implies that the terms of the form $A_{i,i}(\mathbf{u}_i - \mathbf{v}_i)^2$ and $A_{i,i+1}(\mathbf{v}_i - \mathbf{u}_{i+1})^2$ (for $i = 1, \ldots, D$) contribute at least $\frac{1}{dn(2D+1)}$ to this sum (both $A_{i,i}$ and $A_{i,i+1}$ are at least $1/d$ since they correspond to self-loops and edges of the graph).

Plugging in the trivial bound $D \leq n-1$ this already shows that $\lambda(G) \leq 1 - \frac{1}{4dn^2}$. To prove the lemma as stated, we use the fact (left as Exercise 21.2) that for every d-regular graph $D \leq 3n/(d+1)$. ■

The proof can be strengthened to show a similar result for every connected non-bipartite graph (not just those with self-loops at every vertex). Note that this condition is essential: if A is the adjacency matrix of a bipartite graph then one can find a vector $\mathbf{v}$ such that $A\mathbf{v} = -\mathbf{v}$.

21.1.2 Analysis of the randomized algorithm for undirected connectivity.

Together, lemmas 21.3 and 21.4 imply that, at least for regular graphs, if s is connected to t then a sufficiently long random walk from s will hit t in polynomial time with high probability.

Corollary 21.5 *Let G be a d-regular n-vertex graph with all vertices having a self-loop. Let s be a vertex in G. Let $\ell > 24n^2 \log n$ and let X_ℓ denote the distribution of the vertex of the ℓ^{th} step in a random walk from s. Then, for every t connected to s, $\Pr[X_\ell = t] > \frac{1}{2n}$.*

PROOF: By these lemmas 21.3 and 21.4 , if we consider the restriction of an n-vertex graph G to the connected component of s, then for every probability vector $\mathbf{p}$ over this component and $\ell \geq 13n^2$, $\|A^\ell \mathbf{p} - \mathbf{1}\|_2 < (1 - \frac{1}{12n^2})^{24n^2 \log n} < \frac{1}{n^2}$, where $\mathbf{1}$ here is the uniform distribution over this component. Using the relations between the L_1 and L_2 norms (see Note 21.1), $|A^\ell \mathbf{p} - \mathbf{1}|_1 < \frac{1}{n^{1.5}} < \frac{1}{2n}$ and hence every element in the connected component appears in $A^\ell \mathbf{p}$ with at least $1/n - 1/(2n) \geq 1/(2n)$ probability. ■

Note that Corollary 21.5 implies that if we repeat the $24n^2 \log n$ walk for $O(n \log n)$ times (or equivalently, if we take a walk of, say, length $100n^3 \log^2 n$) then we will hit *every* vertex t connected to s with high probability.

21.2 Expander graphs.

Expander graphs are extremely useful combinatorial objects, which we will encounter several times in the book. They can be defined in two equivalent ways. At a high level, these two equivalent definitions can be described as follows:

- *Combinatorial definition:* A constant-degree regular graph G is an *expander* if for every subset S of less than half of G's vertices, a constant fraction of the edges touching S are from S to its complement in G; see Figure 21.1.

- *Algebraic expansion:* A constant-degree regular graph G is an *expander* if its parameter $\lambda(G)$ bounded away from 1 by some constant. That is, $\lambda(G) \leq 1 - \epsilon$ for some constant $\epsilon > 0$.

What do we mean by a constant? By *constant* we refer to a number that is independent of the size of the graph. We will typically talk about graphs that are part of an infinite *family* of graphs, and so by constant we mean a value that is the same for all graphs in the family, regardless of their size. Below we make the definitions more precise, and show their equivalence.

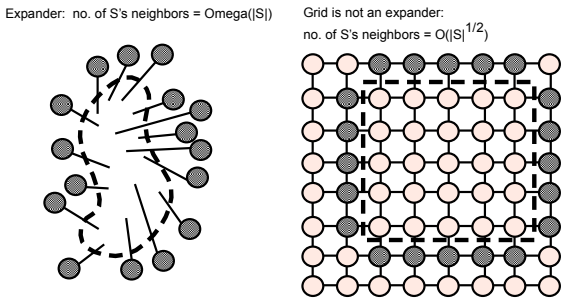


Figure 21.1: In a *combinatorial edge expander*, every subset S of the vertices that is not too big has at least $\Omega(|S|)$ edges to neighbors outside the set. The grid (and every other planar graph) is not a combinatorial edge expander as a $k \times k$ square in the grid has only $O(k)$ neighbors outside it.

21.2.1 The Algebraic Definition

The Algebraic definition of expanders is as follows:

Definition 21.6 ((n, d, λ)-expander graphs.)
If G is an n-vertex d-regular G with $\lambda(G) \leq \lambda$ for some number $\lambda < 1$ then we say that G is an (n, d, λ)-graph.
A family of graphs $\{G_n\}_{n\in\mathbb{N}}$ is an *expander graph family* if there are some constants $d \in \mathbb{N}$ and $\lambda < 1$ such that for every n, G_n is an (n, d, λ)-graph.

Many text use simply the name (n, d, λ)-graphs for such graphs. Also, many text use *unnormalized* adjacency matrices, and so have λ range between 0 and d. The smallest λ can be for a d-regular n-vertex graph is $(1 - o(1))\frac{2\sqrt{d-1}}{d}$ where $o(1)$ denotes a function tending to 0 as the number of vertices grows. This is called the Alon-Boppana bound and graphs meeting this bound are called *Ramanujan graphs* (see Exercise 21.7 for a weaker bound).

Explicit constructions. As we will see in Section 21.2.2, it is not hard to show that expander families exist using the probabilistic method. But this does not yield *explicit* constructions of such graphs which are often needed for applications. We say that an expander family $\{G_n\}_{n\in\mathbb{N}}$ is *explicit* if there is a polynomial-time algorithm that on input 1^n outputs the adjacency matrix of G_n. We say that the family is *strongly explicit* if there is a polynomial-time algorithm that on inputs $\langle n, v, i\rangle$ where $v \in [n], i \in [d]$ outputs the (index of the) i^{th} neighbor of v. Note that in the strongly explicit case, the lengths of the algorithm's inputs and outputs are $O(\log n)$ and so it runs in time polylog(n).

Fortunately, several explicit and strongly explicit constructions of expander graphs are known. Some of these constructions are very simple and efficient, but their analysis is highly non-trivial and uses relatively deep mathematics.[3] In Section 21.3 we will see a strongly explicit construction of expanders with elementary analysis. This construction also introduces a tool that we'll use to derandomize the random-walk algorithm for UPATH in Section 21.4.

21.2.2 Combinatorial expansion and existence of expanders.

We now describe a combinatorial criteria that is roughly equivalent to Definition 21.6. One advantage of this criteria is that it makes it easy to prove that a non-explicit expander family exists using the probabilistic method. It is also quite useful in several applications.

[3] An example is the following 3-regular expander graph: the vertices are the numbers 1 to $p-1$ for some prime p, and each number x is connected to $x+1$,$x-1$ and $x^{-1} \pmod p$.

Definition 21.7 (Combinatorial (edge) expansion)
An n-vertex d-regular graph $G = (V, E)$ is called an (n, d, ρ)-*combinatorial edge expander* if for every subset S of vertices satisfying $|S| \leq n/2$,

$$|E(S, \overline{S})| \geq \rho d|S|\,,$$

where $\overline{S}$ denotes the complement of S and for subsets S, T of vertices, $E(S, T)$ denotes the set of edges $\overline{ij}$ with $i \in S$ and $j \in T$.

Note that in this case the bigger ρ is the better the expander. We'll loosely use the term expander for any (n, d, ρ)-combinatorial edge expander with ρ a positive constant (independent of n). Using the probabilistic method, one can prove the following theorem: (Exercise 21.8 asks you to prove a slightly weaker version)

Theorem 21.8 (Existence of expanders)
Let $\epsilon > 0$ be some constant. Then there exists $d = d(\epsilon)$ and $N \in \mathbb{N}$ such that for every $n > N$ there exists an $(n, d, 1/2 - \epsilon)$-combinatorial edge expander.

Theorem 21.8 is tight in the sense that there is no (n, d, ρ)-combinatorial edge expander for $\rho > 1/2$ (Exercise 21.9). The following theorem relates combinatorial expansion with our previous Definition 21.6

Theorem 21.9 (Combinatorial and algebraic expansion [**Alon86**])
1. *If G is an (n, d, λ)-expander graph then it is an $(n, d, (1 - \lambda)/2)$-combinatorial edge expander.*
2. *If G is an (n, d, ρ)-combinatorial edge expander then its second largest eigenvalue (*without *taking absolute values) is at most $1 - \frac{\rho^2}{2}$. If furthermore G has all self loops then it is an $(n, d, 1 - \epsilon)$-expander where $\epsilon = \min\left\{\frac{2}{d}, \frac{\rho^2}{2}\right\}$.*

The condition that G has all the self-loops of Theorem 21.9 is used again to rule out bipartite graphs, which can be very good combinatorial edge expanders but have one eigenvalue equal to -1 and hence a spectral gap of zero.

21.2.3 Algebraic expansion implies combinatorial expansion.

The first part of Theorem 21.9 follows immediately from the following lemma:

Lemma 21.10 *Let G be an (n, d, λ) graph, S a subset of G's vertices and T its complement. Then*

$$|E(S, T)| \geq (1 - \lambda)\frac{d|S||T|}{|S| + |T|}\,.$$

PROOF: Let $\mathbf{x} \in \mathbb{R}^n$ denote the following vector:

$$\mathbf{x}_i = \begin{cases} +|T| & i \in S \\ -|S| & i \in T \\ 0 & \text{otherwise} \end{cases}\,.$$

Note that $\|\mathbf{x}\|_2^2 = |S||T|^2 + |T||S|^2 = |S||T|(|S| + |T|)$ and $\mathbf{x} \perp \mathbf{1}$.

Let $Z = \sum_{i,j} A_{i,j}(x_i - x_j)^2$. On the one hand $Z = \frac{2}{d}|E(S,T)|(|S| + |T|)^2$, since every edge $\overline{i\,j}$ with $i \in S$ and $j \in T$ appears twice in this sum, each time contributing $\frac{1}{d}(|S| + |T|)^2$ to the total. On the other hand,

$$Z = \sum_{i,j} A_{i,j}x_i^2 - 2\sum A_{i,j}x_i x_j + \sum_{i,j} A_{i,j}x_j^2 = 2\|\mathbf{x}\|_2^2 - 2\langle \mathbf{x}, A\mathbf{x}\rangle$$

(using the fact that A's rows and columns sum up to one). Since $\mathbf{x} \perp \mathbf{1}$ and $\|A\mathbf{x}\|_2 \leq \lambda\|\mathbf{x}\|_2$, we get that

$$\tfrac{1}{d}|E(S,T)|(|S| + |T|)^2 \geq (1-\lambda)\|x\|_2^2 .$$

Plugging in $\|x\|_2^2 = |S||T|(|S| + |T|)$ completes the proof. ■

Algebraic expansion also allows us to obtain an estimate on the number of edges between not-too-small subsets S and T, even if they are not disjoint:

Lemma 21.11 (Expander Mixing Lemma) *Let $G = (V,E)$ be an (n,d,λ)-expander graph. Let $S, T \subseteq V$, then*

$$\left| |E(S,T)| - \frac{d}{n}|S||T| \right| \leq \lambda d\sqrt{|S||T|}$$

Note that $\frac{d}{n}|S||T|$ is the number of edges one would expect between S and T if each vertex of S chooses its d neighbors at random. We leave the proof of Lemma 21.11 as Exercise 21.10.

21.2.4 Combinatorial Expansion Implies Algebraic Expansion

We now prove the second part of Theorem 21.9. Let $G = (V,E)$ be an n-vertex d-regular graph such that for every subset $S \subseteq V$ with $|S| \leq n/2$, there are $\rho|S|$ edges between S and $\overline{S} = V \setminus S$, and let A be G's normalized adjacency matrix.

Let λ be the second largest eigenvalue of A (not taking absolute values). We need to prove that $\lambda \leq 1 - \rho^2/2$. By the definition of an eigenvalue there exists a vector $\mathbf{u} \perp \mathbf{1}$ such that $A\mathbf{u} = \lambda\mathbf{u}$. Write $\mathbf{u} = \mathbf{v} + \mathbf{w}$ where $\mathbf{v}$ is equal to $\mathbf{u}$ on the coordinates on which $\mathbf{u}$ is positive and equal to 0 otherwise, and $\mathbf{w}$ is equal to $\mathbf{u}$ on the coordinates on which $\mathbf{u}$ is negative, and equal to 0 otherwise. (Since $\mathbf{u} \perp \mathbf{1}$, both $\mathbf{v}$ and $\mathbf{w}$ are nonzero.) We can assume that $\mathbf{v}$ is nonzero on at most $n/2$ of its coordinates (otherwise take $-\mathbf{u}$ instead of $\mathbf{u}$). Let $Z = \sum_{i,j} A_{i,j}(\mathbf{v}_i^2 - \mathbf{v}_j^2)$. Part 2 of the theorem (except for the "furthermore" clause) follows immediately from the following two claims:

Claim 1: $Z \geq 2\rho\|\mathbf{v}\|_2^2$.

Claim 2: $Z \leq \sqrt{8(1-\lambda)}\|\mathbf{v}\|_2^2$.

Proof of Claim 1: Sort the coordinates of $\mathbf{v}$ so that $\mathbf{v}_1 \geq \mathbf{v}_2 \geq \cdots \geq \mathbf{v}_n$ (with $\mathbf{v}_i = 0$ for $i > n/2$). Then, using $\mathbf{v}_i^2 - \mathbf{v}_j^2 = \sum_{k=i}^{j+1}(\mathbf{v}_k^2 - \mathbf{v}_{k+1}^2)$,

$$Z = \sum_{i,j} A_{i,j}(\mathbf{v}_i^2 - \mathbf{v}_j^2) = 2\sum_{i<j} A_{i,j}\sum_{k=i}^{j-1}(\mathbf{v}_k^2 - \mathbf{v}_{k+1}^2) .$$

Note that every term $(\mathbf{v}_k^2 - \mathbf{v}_{k+1}^2)$ appears in this sum once (with a weight of $2/d$) per each edge $\overline{i\,j}$ such that $i \leq k < j$. Since $\mathbf{v}_k = 0$ for $k > n/2$, this means that

$$Z = \tfrac{2}{d}\sum_{k=1}^{n/2} |E(\{1..k\}, \{k+1..n\})|(\mathbf{v}_k^2 - \mathbf{v}_{k+1}^2) \geq \tfrac{2}{d}\sum_{k=1}^{n/2} \rho k(\mathbf{v}_k^2 - \mathbf{v}_{k+1}^2) ,$$

by G's expansion. But, rearranging the terms (and using the fact that $\mathbf{v}_k = 0$ for $k > n/2$), the last sum is equal to

$$\tfrac{2}{d}d\rho \sum_{k=1}^{n/2} k\mathbf{v}_k^2 - (k-1)\mathbf{v}_k^2 = 2\sum_{k=1}^{n} \mathbf{v}_k^2 = 2\rho\|\mathbf{v}\|_2^2 .$$

■

PROOF OF CLAIM 2: Since $A\mathbf{u} = \lambda\mathbf{u}$ and $\langle \mathbf{v}, \mathbf{w}\rangle = 0$,

$$\langle A\mathbf{v}, \mathbf{v}\rangle + \langle A\mathbf{w}, \mathbf{v}\rangle = \langle A(\mathbf{v}+\mathbf{w}), \mathbf{v}\rangle = \langle A\mathbf{u}, \mathbf{v}\rangle = \langle \lambda(\mathbf{v}+\mathbf{w}), \mathbf{v}\rangle = \lambda\|\mathbf{v}\|_2^2 .$$

Since $\langle A\mathbf{w}, \mathbf{v}\rangle$ is not positive, $\langle A\mathbf{v}, \mathbf{v}\rangle / \|\mathbf{v}\|_2^2 \geq \lambda$, meaning that

$$1 - \lambda \geq 1 - \frac{\langle A\mathbf{v}, \mathbf{v}\rangle}{\|\mathbf{v}\|_2^2} = \frac{\|\mathbf{v}\|_2^2 - \langle A\mathbf{v}, \mathbf{v}\rangle}{\|\mathbf{v}\|_2^2} = \frac{\sum_{i,j} A_{i,j}(\mathbf{v}_i - \mathbf{v}_j)^2}{2\|\mathbf{v}\|_2^2} , \tag{4}$$

where the last equality is due to $\sum_{i,j} A_{i,j}(\mathbf{v}_i - \mathbf{v}_j)^2 = \sum_{i,j} A_{i,j}\mathbf{v}_i^2 - 2\sum_{i,j} A_{i,j}\mathbf{v}_i\mathbf{v}_j + \sum_{i,j} A_{i,j}\mathbf{v}_j^2 = 2\|\mathbf{v}\|_2^2 - 2\langle A\mathbf{v}, \mathbf{v}\rangle$. (We use here the fact that each row and column of A sums to one.)

Multiply both numerator and denominator of the last term in (4) by $\sum_{i,j} A_{i,j}(\mathbf{v}_i^2 + \mathbf{v}_j^2)$. The new numerator is

$$\left(\sum_{i,j} A_{i,j}(\mathbf{v}_i - \mathbf{v}_j)^2\right)\left(\sum_{i,j} A_{i,j}(\mathbf{v}_i + \mathbf{v}_j)^2\right) \geq \left(\sum_{i,j} A_{i,j}(\mathbf{v}_i - \mathbf{v}_j)(\mathbf{v}_i + \mathbf{v}_j)\right)^2 .$$

using the Cauchy-Schwartz inequality.[4] Hence, using $(a-b)(a+b) = a^2 - b^2$,

$$1 - \lambda \geq \frac{\left(\sum_{i,j} A_{i,j}(\mathbf{v}_i^2 - \mathbf{v}_j^2)\right)^2}{2\|\mathbf{v}\|_2^2 \sum_{i,j} A_{i,j}(\mathbf{v}_i + \mathbf{v}_j)^2} = \frac{Z^2}{2\|\mathbf{v}\|_2^2\left(\sum_{i,j} A_{i,j}\mathbf{v}_i^2 + 2\sum_{i,j} A_{i,j}\mathbf{v}_i\mathbf{v}_j + \sum_{i,j} A_{i,j}\mathbf{v}_j^2\right)} =$$

$$\frac{Z^2}{2\|\mathbf{v}\|_2^2\left(2\|\mathbf{v}\|_2^2 + 2\langle A\mathbf{v}, \mathbf{v}\rangle\right)} \geq \frac{Z^2}{8\|\mathbf{v}\|_2^4} ,$$

where the last inequality is due to the fact that A is a symmetric stochastic matrix, and hence $\|A\mathbf{v}\|_2 \leq \|\mathbf{v}\|_2$ for every $\mathbf{v}$, implying that $\langle A\mathbf{v}, \mathbf{v}\rangle \leq \|\mathbf{v}\|_2^2$.

The "furthermore" part is obtained by noticing that adding all the self-loops to a $d-1$-regular graph is equivalent to transforming its normalized adjacency matrix A into the matrix $\frac{d-1}{d}A + \frac{1}{d}I$ where I is the identity matrix. Since A's smallest eigenvalue (not taking absolute values) is at least -1, the new smallest eigenvalue is at least $-\frac{d-1}{d} + \frac{1}{d} = -1 + \frac{2}{d}$.■

21.2.5 Error reduction using expanders.

Before constructing expanders, let us see one application for them in the area of probabilistic algorithms. Recall that in Section **[expand:sec:errorred]** we saw that we can reduce the error of a probabilistic algorithm from, say, 1/3 to $2^{-\Omega(k)}$ by executing it k times independently and taking the majority value. If the algorithm utilized m random coins, this procedure will use $m \cdot k$ random coins, and it seems hard to think of a way to save on randomness. Nonetheless, using expanders we can obtain such error reduction using only $m + O(k)$ random coins.

The idea is simple: take an expander graph G from a very explicit family that is an $(M = 2^m, d, 1/10)$-expander graph for some constant d.[5] Choose a vertex v_1 at random, and take a length

[4] The Cauchy-Schwartz inequality says that for every two vectors $\mathbf{x}, \mathbf{y}$, $\langle \mathbf{x}, \mathbf{y}\rangle \leq \|\mathbf{x}\|_2\|\mathbf{y}\|_2$. Here we index over (i,j), and use $\mathbf{x}_{i,j} = \sqrt{A_{i,j}}(\mathbf{v}_i^2 - \mathbf{v}_j^2)$ and $\mathbf{y}_{i,j} = \sqrt{A_{i,j}}(\mathbf{v}_i^2 + \mathbf{v}_j^2)$.

[5] Note that we can use graph powering to transform any explicit expander family into an expander family with parameter $\lambda < 1/10$ (see also Section 21.3).

$k-1$ long random walk on G to obtain vertices $v_2, \ldots, v_k$ (note that choosing a random neighbor of a vertex requires $O(\log d) = O(1)$ random bits). Invoke the algorithm k times using $v_1, \ldots, v_k$ for the random coins (we identify the set $[M]$ of vertices with the set $\{0,1\}^m$ of possible random coins for the algorithm) and output the majority answer.

To keep things simple, we analyze here only the case of algorithms with one-sided error. For example, consider an **RP** algorithm that will never output "accept" if the input is not in the language, and for inputs in the language will output "accept" with probability 1/2 (the case of a **coRP** algorithm is analogous). For such an algorithm the procedure will output "accept" if the algorithm accepts even on a single set of coins v_i. If the input is not in the language, the procedure will never accept. If the input is in the language, then let $\mathcal{B} \subseteq [M]$ denote the "bad" set of coins on which the algorithms rejects. We know that $|\mathcal{B}| \leq \frac{M}{3}$. Plugging in $\beta = 1/3$ and $\lambda = 1/10$ in the following theorem immediately implies that the probability the above procedure will reject an input in the language is bounded by $2^{-\Omega(k)}$:

Theorem 21.12 (Expander walks [**derandomizedGraphProducts**])
Let G be an (n,d,λ) graph, and let $\mathcal{B} \subseteq [n]$ satisfying $|\mathcal{B}| \leq= \beta n$ for some $\beta \in (0,1)$. Let $X_1, \ldots, X_k$ be random variables denoting a $k-1$-step random walk in G from X_1, where X_1 is chosen uniformly in $[n]$. Then,

$$\Pr[\underbrace{\forall_{1\leq i\leq k} X_i \in B}_{(*)}] \leq ((1-\lambda)\sqrt{\beta}+\lambda)^{k-1}$$

Note that if λ and β are both constants smaller than 1 then so is the expression $(1-\lambda)\sqrt{\beta}+\lambda$.

PROOF: For $1 \leq i \leq k$, let B_i be the event that $X_i \in \mathcal{B}$. Note that the probability $(*)$ we're trying to bound is

$$\Pr[\wedge_{i=1}^k B_i] = \Pr[B_1]\Pr[B_2|B_1]\cdots\Pr[B_k|B_1,\ldots,B_{k-1}]\,.$$

Denote by B the linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$ that "zeroes out" the coordinates that are not in $\mathcal{B}$. That is, for every $i \in [n]$, $(B\mathbf{u})_i = \mathbf{u}_i$ if $i \in \mathcal{B}$ and $(B\mathbf{u})_i = 0$ otherwise. It's not hard to verify that for every probability vector $\mathbf{p}$ over $[n]$, $B\mathbf{p}$ is a vector whose coordinates sum up to the probability that a vertex i chosen according to $\mathbf{p}$ is in $\mathcal{B}$. Furthermore, if we normalize the vector $B\mathbf{p}$ to sum up to one, we get a the probability vector corresponding to $\mathbf{p}$ conditioned on this event.

Thus, if we let $\mathbf{1} = (1/n, \ldots, 1/n)$ denote the uniform distribution over $[n]$ and $\mathbf{p}^i \in \mathbb{R}^N$ be the distribution of X_i conditioned on the events $B_1, \ldots, B_i$, then

$$\mathbf{p}^1 = \frac{1}{\Pr[B_1]}B\mathbf{1}$$
$$\mathbf{p}^2 = \frac{1}{\Pr[B_2|B_1]}\frac{1}{\Pr[B_1]}BAB\mathbf{1}$$

and more generally

$$\mathbf{p}^i = \frac{1}{\Pr[B_i|B_{i-1}\ldots B_1]\cdots\Pr[B_1]}(BA)^{i-1}B\mathbf{1}\,.$$

Since every probability vector $\mathbf{p}$ satisfies $|\mathbf{p}|_1 = 1$,

$$(*) = |(\hat{B}A)^{k-1}\hat{B}\mathbf{1}|_1 \tag{5}$$

We bound the norm on the right-hand side of (5) by showing

$$\|(\hat{B}A)^{k-1}B\mathbf{1}\|_2 \leq \frac{((1-\lambda)\sqrt{\beta}+\lambda)^{k-1}}{\sqrt{n}}\,, \tag{6}$$

which suffices since for every $\mathbf{v} \in \mathbb{R}^N$, $|\mathbf{v}|_1 \leq \sqrt{N}\|\mathbf{v}\|_2$ (see Note 21.1). To prove (6), we will use the following definition and Lemma:

Definition 21.13 (Matrix norm)
For every matrix A, the *matrix norm* of A, denoted by $\|A\|$, is defined as the maximum of $\|A\mathbf{v}\|_2$ over all vectors $\mathbf{v}$ satisfying $\|\mathbf{v}\|_2 = 1$.

Exercises 21.3 and 21.4 ask you to prove that the norm of every normalized adjacency matrix is 1, and that for every two n by n matrices A, B, $\|A+B\| \leq \|A\| + \|B\|$ and $\|AB\| \leq \|A\|\|B\|$.

Lemma 21.14 *Let A be a normalized adjacency matrix of an (n, d, λ)-expander graph G. Let J be the adjacency matrix of the n-clique with self loops (i.e., $J_{i,j} = 1/n$ for every i, j). Then*

$$A = (1-\lambda)J + \lambda C \tag{7}$$

where $\|C\| \leq 1$.

Note that for every probability vector $\mathbf{p}$, $J\mathbf{p}$ is the uniform distribution, and so this lemma tells us that in some sense, we can think of a step on a (n, d, λ)-expander graph as going to the uniform distribution with probability $1-\lambda$, and to a different distribution with probability λ. This is of course not completely accurate, as a step on a d-regular graph will only go the one of the d neighbors of the current vertex, but we'll see that for the purposes of our analysis, the condition (7) will be just as good.[6]

Proof of Lemma 21.14: Indeed, simply define $C = \frac{1}{\lambda}(A - (1-\lambda)J)$. We need to prove $\|C\mathbf{v}\|_2 \leq \|\mathbf{v}\|_2$ for very $\mathbf{v}$. Decompose $\mathbf{v}$ as $\mathbf{v} = \mathbf{u} + \mathbf{w}$ where $\mathbf{u} = \alpha\mathbf{1}$ for some $\alpha \in \mathbb{R}$ and $\mathbf{w} \perp \mathbf{1}$. Since $A\mathbf{1} = \mathbf{1}$ and $J\mathbf{1} = \mathbf{1}$ we get that $C\mathbf{u} = \frac{1}{\lambda}(\mathbf{u} - (1-\lambda)\mathbf{u}) = \mathbf{u}$. Now, let $\mathbf{w}' = A\mathbf{w}$. Then $\|\mathbf{w}'\|_2 \leq \lambda\|\mathbf{w}\|_2$ and, as we saw in the proof of Lemma 21.3, $\mathbf{w}' \perp \mathbf{1}$. In other words, the sum of the coordinates of $\mathbf{w}$ is zero, meaning that $J\mathbf{w} = \mathbf{0}$. We get that $C\mathbf{w} = \frac{1}{\lambda}\mathbf{w}'$. Since $\mathbf{w}' \perp \mathbf{u}$, $\|C\mathbf{v}\|_2^2 = \|\mathbf{u} + \frac{1}{\lambda}\mathbf{w}'\|_2^2 = \|\mathbf{u}\|_2^2 + \|\frac{1}{\lambda}\mathbf{w}'\|_2^2 \leq \|\mathbf{u}\|_2^2 + \|\mathbf{w}\|_2^2 = \|\mathbf{v}\|_2^2$, where we use twice the Pythagorean theorem that for $\mathbf{u} \perp \mathbf{w}$, $\|\mathbf{u}+\mathbf{w}\|_2^2 = \|\mathbf{u}\|_2^2 + \|\mathbf{w}\|_2^2$. ■

Returning to the proof of Theorem 21.12, we can write $BA = B\big((1-\lambda)J + \lambda C\big)$, and hence $\|BA\| \leq (1-\lambda)\|BJ\| + \lambda\|BC\|$. Since J's output is always a vector of the form $\alpha\mathbf{1}$, and it can be easily verified that $\|B\mathbf{1}\|_2 = \sqrt{\frac{\beta n}{n^2}} = \frac{\sqrt{\beta}}{\sqrt{n}} = \sqrt{\beta}\|\mathbf{1}\|_2$, $\|BJ\| = \sqrt{\beta}$. Also, because B is an operation that merely zeros out some parts of its input, $\|B\| \leq 1$ implying that $\|BC\| \leq 1$. Thus, $\|BA\| \leq (1-\lambda)\sqrt{\beta} + \lambda$. This means that $\|(BA)^{k-1}B\mathbf{1}\|_2 \leq ((1-\lambda)\sqrt{\beta} + \lambda)^{k-1}\frac{\sqrt{\beta}}{\sqrt{n}}$, establishing (6). ■

The success of the error reduction procedure for *two-sided error* algorithms is obtained by the following theorem, whose proof we omit:

Theorem 21.15 (Expander Chernoff Bound [??])
Let G be an (N, d, λ)-expander graph and $B \subseteq [N]$ with $|B| = \beta N$. Let $X_1, \ldots, X_k$ be random variables denoting a $k-1$-step random walk in G (where X_1 is chosen uniformly). For every $i \in [k]$, define B_i to be 1 if $X_i \in B$ and 0 otherwise. Then, for every $\delta > 0$,

$$\Pr\left[\left|\frac{\sum_{i=1}^k B_i}{k} - \beta\right| > \delta\right] < 2e^{(1-\lambda)\delta^2 k/60}$$

[6] Algebraically, the reason (7) is not equivalent to going to the uniform distribution in each step with probability $1-\lambda$ is that C is not necessarily a stochastic matrix, and may have negative entries.

21.3 Explicit construction of expander graphs

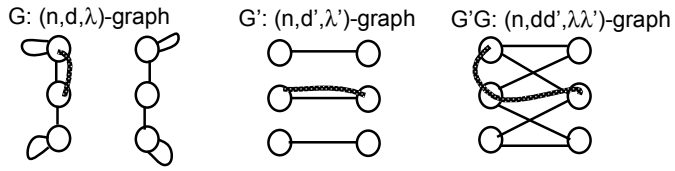
We now show a construction of a very explicit expander graph family. The main tools in our construction will be several types of graph products. A *graph product* is an operation that takes two graphs G, G' and outputs a graph H. Typically we're interested in the relation between properties of the graphs G, G' to the properties of the resulting graph H. In this section we will mainly be interested in three parameters: the number of vertices (denoted n), the degree (denoted d), and the 2^{nd} largest eigenvalue of the normalized adjacency matrix (denoted λ), and study how different products affect these parameters. We then use these products to obtain a construction of a strongly explicit expander graph family. In the next section we will use the same products to show a *deterministic* logspace algorithm for undirected connectivity.

21.3.1 Rotation maps.

In addition to the adjacency matrix representation, we can also represent an n-vertex degree-d graph G as a function $\hat{G}$ from $[n] \times [d]$ to $[n]$ that given a pair $\langle v, i \rangle$ outputs u where the i^{th} neighbor of v in G. In fact, it will be convenient for us to have $\hat{G}$ output an additional value $j \in [d]$ where j is the index of v as a neighbor of u. Given this definition of $\hat{G}$ it is clear that we can invert it by applying it again, and so it is a permutation on $[n] \times [d]$. We call $\hat{G}$ the *rotation map* of G. For starters, one may think of the case that $\hat{G}(u, i) = (v, i)$ (i.e., v is the i^{th} neighbor of u iff u is the i^{th} neighbor of v). In this case we can think of $\hat{G}$ as operating only on the vertex. However, we will need the more general notion of a rotation map later on.

We can describe a graph product in the language of graphs, adjacency matrices, or rotation maps. Whenever you see the description of a product in one of this forms (e.g., as a way to map two graphs into one), it is a useful exercise to work out the equivalent descriptions in the other forms (e.g., in terms of adjacency matrices and rotation maps).

21.3.2 The matrix/path product

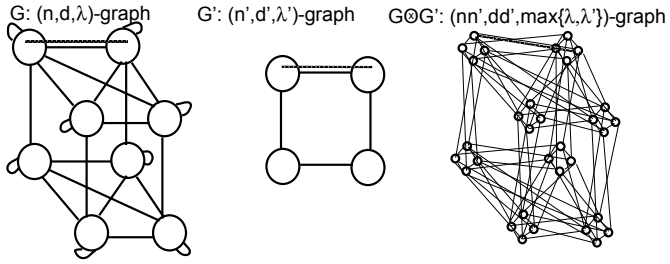


For every two n vertex graphs G, G' with degrees d, d' and adjacency matrices A, A', the graph $G'G$ is the graph described by the adjacency matrix $A'A$. That is, $G'G$ has an edge (u, v) for every length 2-path from u to v where the first step in the path is taken on en edge of G and the second is on an edge of G'. Note that G has n vertices and degree dd'. Typically, we are interested in the case $G = G'$, where it is called *graph squaring*. More generally, we denote by G^k the graph $G \cdot G \cdots G$ (k times). We have already encountered this case before in Lemma 21.3, and similar analysis yields the following lemma (whose proof we leave as Exercise 21.6):

Lemma 21.16 (Matrix product improves expansion) $\lambda(G'G) \leq \lambda(G')\lambda(G')$

Note that one can easily compute the rotation map of $G'G$ using the rotation maps of G and G'.

21.3.3 The tensor product



Let G and G' be two graphs with n (resp n') vertices and d (resp. d') degree, and let $\hat{G} : [n] \times [d] \to [n] \times [d]$ and $\hat{G'} : [n'] \times [d'] \to [n'] \times [d']$ denote their respective *rotation maps*. The *tensor product* of G and G', denoted $G \otimes G'$, is the graph over nn' vertices and degree dd' whose rotation map $\widehat{G \otimes G'}$ is the permutation over $([n] \times [n']) \times ([d] \times [d'])$ defined as follows

$$\widehat{G \otimes G'}(\langle u, v\rangle, \langle i, j\rangle) = \langle u', v'\rangle, \langle i', j'\rangle\,,$$

where $(u', i') = \hat{G}(u, i)$ and $(v', j') = \hat{G'}(v, j)$. That is, the vertex set of $G \otimes G'$ is pairs of vertices, one from G and the other from G', and taking a the step $\langle i, j\rangle$ on $G \otimes G'$ from the vertex $\langle u, v\rangle$ is akin to taking two independent steps: move to the pair $\langle u', v'\rangle$ where u' is the i^{th} neighbor of u in G and v' is the i^{th} neighbor of v in G'.

In terms of adjacency matrices, the tensor product is also quite easy to describe. If $A = (a_{i,j})$ is the $n \times n$ adjacency matrix of G and $A' = (a'_{i',j'})$ is the $n' \times n'$ adjacency matrix of G', then the adjacency matrix of $G \otimes G'$, denoted as $A \otimes A'$, will be an $nn' \times nn'$ matrix that in the $\langle i, i'\rangle^{th}$ row and the $\langle j, j'\rangle$ column has the value $a_{i,j} \cdot a'_{i',j'}$. That is, $A \otimes A'$ consists of n^2 copies of A', with the $(i, j)^{th}$ copy scaled by $a_{i,j}$:

$$A \otimes A' = \begin{pmatrix} a_{1,1}A' & a_{1,2}A' & \dots & a_{1,n}A' \\ a_{2,1}A' & a_{2,2}A' & \dots & a_{2,n}A' \\ \vdots & & & \vdots \\ a_{n,1}A' & a_{n,2}A' & \dots & a_{n,n}A' \end{pmatrix}$$

The tensor product can also be described in the language of graphs as having a cluster of n' vertices in $G \otimes G'$ for every vertex of G. Now if, u and v are two neighboring vertices in G, we will put a bipartite version of G' between the cluster corresponding to u and the cluster corresponding to v in G. That is, if (i, j) is an edge in G' then there is an edge between the i^{th} vertex in the cluster corresponding to u and the j^{th} vertex in the cluster corresponding to v.

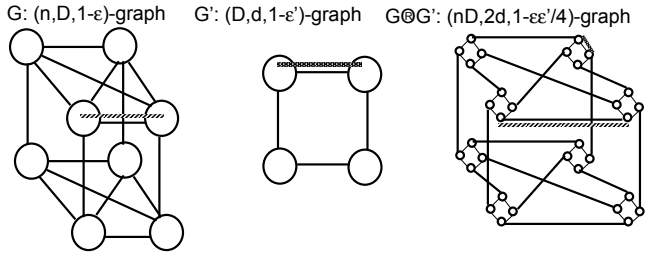
Lemma 21.17 (Tensor product preserves expansion) *Let $\lambda = \lambda(G)$ and $\lambda' = \lambda(G')$ then $\lambda(G \otimes G') \leq \max\{\lambda, \lambda'\}$.*

One intuition for this bound is the following: taking a T step random walk on the graph $G \otimes G'$ is akin to taking two independent random walks on the graphs G and G'. Hence, if both walks converge to the uniform distribution within T steps, then so will the walk on $G \otimes G'$.

Proof of Lemma 21.17: Given some basic facts about tensor products and eigenvalues this is immediate since if $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A (where A is the adjacency matrix of G) and $\lambda'_1, \dots, \lambda'_{n'}$ are the eigenvalues of A (where A' is the adjacency matrix of G'), then the eigenvalues of $A \otimes A'$ are all numbers of the form $\lambda_i \cdot \lambda'_j$, and hence the largest ones apart from 1 are of the form $1 \cdot \lambda(G')$ or $\lambda(G) \cdot 1$ (see also Exercise 21.19). ■

We note that one can show that $\lambda(G \otimes G') \leq \lambda(G) + \lambda(G')$ without relying on any knowledge of eigenvalues (see Exercise 21.20). This weaker bound suffices for our applications.

21.3.4 The replacement product



In both the matrix and tensor products, the degree of the resulting graph is larger than the degree of the input graphs. The following product will enable us to reduce the degree of one of the graphs. Let G, G' be two graphs such that G has n vertices and degree D, and G' has D vertices and degree d. The *balanced replacement product* (below we use simply *replacement product* for short) of G and G' is denoted by $G \circledR G'$ is the nn'-vertex $2d$-degree graph obtained as follows:

1. For every vertex u of G, the graph $G \circledR G'$ has a copy of G' (including both edges and vertices).
2. If u, v are two neighboring vertices in G then we place d parallel edges between the i^{th} vertex in the copy of G' corresponding to u and the j^{th} vertex in the copy of G' corresponding to v, where i is the index of v as a neighbor of u and j is the index of u as a neighbor of v in G. (That is, taking the i^{th} edge out of u leads to v and taking the j^{th} edge out of v leads to u.)

(Some texts reserve the term replacement product for the variant that uses only a single edge (as opposed to d parallel edges) in Item 2 above. The addition of parallel edges ensures that a random step from a vertex v in $G \circledR G'$ will move to a neighbor within the same cluster and a neighbor outside the cluster with the same probability.)

The replacement product also has a simple description in terms of rotation maps: since $G \circledR G'$ has nD vertices and $2d$ degree, its rotation map $G \hat{\circledR} G'$ is a permutation over $([n] \times [D]) \times ([d] \times \{0,1\})$ and so can be thought of as taking four inputs u, v, i, b where $u \in [n]$, $v \in [D]$, $i \in [d]$ and $b \in \{0,1\}$. If $b = 0$ then it outputs $u, \hat{G}'(v,i), b$ and if $b = 1$ then it outputs $\hat{G}(u,v), i, b$. That is, depending on whether b is equal to 0 or 1, the rotation map either treats v as a vertex of G' or as an edge label of G.

In the language of adjacency matrices the replacement product is described as follows:

$$A \circledR A' = 1/2\hat{A} + 1/2(I_n \otimes A'), \tag{8}$$

where A, A' denote the normalized adjacency matrices of G and G' respectively, and $\hat{A}$ denotes the permutation matrix corresponding to the rotation map of G. That is, $\hat{A}$ is an $(nD) \times (nD)$ matrix whose $(i,j)^{th}$ column is all zeroes except a single 1 in the $(i',j')^{th}$ place where $(i',j') = \hat{G}(i,j)$.

If $D \gg d$ then the replacement product's degree will be significantly smaller than G's degree. The following Lemma shows that this dramatic degree reduction does not cause too much of a deterioration in the graph's expansion:

Lemma 21.18 (Expansion of replacement product) *If $\lambda(G) \leq 1-\epsilon$ and $\lambda(H) \leq 1-\delta$ then $\lambda(G \circledR H) \leq 1 - \frac{\epsilon\delta^2}{24}$.*

The intuition behind Lemma 21.18 is the following: Think of the input graph G as a good expander whose only drawback is that it has a too high degree D. This means that a k step random

walk on G' requires $O(k \log D)$ random bits. However, as we saw in Section 21.2.5, sometimes we can use fewer random bits if we use an expander. So a natural idea is to generate the edge labels for the walk by taking a walk using a smaller expander G' that has D vertices and degree $d \ll D$. The definition of $G \textcircled{R} G'$ is motivated by this intuition: a random walk on $G \textcircled{R} G'$ is roughly equivalent to using an expander walk on G' to generate labels for a walk on G. In particular, each step a walk over $G \textcircled{R} G'$ can be thought of as tossing a coin and then, based on its outcome, either taking a a random step on G', or using the current vertex of G' as an edge label to take a step on G. Another way to gain intuition on the replacement product is to solve Exercise 21.21, that analyzes the *combinatorial* (edge) expansion of the resulting graph as a function of the edge expansion of the input graphs.

PROOF OF LEMMA 21.18: It suffices to show that $\lambda(G \textcircled{R} H)^3 \leq 1 - \frac{\epsilon\delta^2}{8}$. Since for every graph F, $\lambda(F^k) = \lambda(F)^k$, we will do so by showing $\lambda((G \textcircled{R} H)^3) \leq 1 - \frac{\epsilon\delta^2}{8}$. Let A be the $n \times n$ normalized adjacency matrix of G (with $\hat{A}$ the $(nD) \times (nD)$ permutation matrix corresponding to the rotation map $\hat{G}$), let B be the $D \times D$ normalized adjacency matrix of H, and let C be the adjacency matrix of $(G \textcircled{R} H)^3$. Then, (8) implies that

$$C = (1/2\hat{A} + 1/2(I_n \otimes B))^3 \,, \tag{9}$$

Now Lemma 21.14 implies that $B = (1-\delta)B' + \delta J_D$ for some matrix B' with norm at most 1 (where J_D is the $D \times D$ all $1/D$ matrix). We plug this into (9), expand all terms and then collect together all the terms except for the one corresponding to $1/2\delta(I_n \otimes J)1/2\hat{A}1/2\delta(I_n \otimes J)$. The reader can verify that all terms correspond to matrices of norm at most 1 and hence (9) becomes

$$C = (1 - \tfrac{\delta^2}{8})C' + \tfrac{\delta^2}{8}(I_n \otimes J_D)\hat{A}(I_n \otimes J_D) \,, \tag{10}$$

where C' is some $(nD) \times (nD)$ matrix of norm at most 1. The lemma will follow from the following claim:

CLAIM: $(I_n \otimes J_D)\hat{A}(I_n \otimes J_D) = A \otimes J_D$

PROOF: Indeed, the left-hand side is the normalized adjacency matrix of the graph on nD vertices on which a step from a vertex (i,j) corresponds to: 1) choosing a random $k \in [D]$ 2) letting i' be the k^{th} neighbor of i in G 3) choosing j' at random in $[D]$ moving to the vertex (i,k). We can equivalently describe this as going to a random neighbor i' of i in G and choosing j' at random in $[D]$, which is the graph corresponding to the matrix $A \otimes J_D$. ■

The claim concludes the proof since $\lambda(A \otimes J_D) \leq \max\{\lambda(A), \lambda(J_D)\} = \max\{\lambda(A), 0\}$. The lemma follows by plugging this into (10) and using the fact that $\lambda(C') \leq 1$ for every matrix of norm at most 1. ■

21.3.5 The actual construction.

We now use the three graph products of described above to show a strongly explicit construction of an expander graph family. That is, we prove the following theorem:

Theorem 21.19 (Explicit construction of expanders)
There exists a strongly-explicit λ, d-expander family for some constants d and $\lambda < 1$.

Note that using the matrix/graph product Theorem 21.19 can be improved to yield a strongly-explicit λ, d-expander family for *every* $\lambda > 0$ (albeit at the expense of allowing d to be an arbitrarily large constant depending on λ.

PROOF: We will start by showing something slightly weaker: a very explicit family of graphs $\{G_k\}$ where G_k is not a graph on k vertices but on roughly c^k vertices for some constant c. That is, rather than showing a family of graphs for *every* size n, we will only show a family of graphs for certain sizes n. We will then sketch how the construction can be improved to yield a graph family containing a graph of every size n.

The construction is recursive: we start by a finite size graph G_1 (which we can find using brute force search), and construct the graph G_k from the graph G_{k-1}. On a high level the construction is as follows: each of the three product will serve a different purpose in the construction. The *Tensor product* allows us to take G_{k-1} and increase its number of vertices, at the expense of increasing the degree and possibly some deterioration in the expansion. The *replacement product* allows us to dramatically reduce the degree at the expense of additional deterioration in the expansion. Finally, we use the *Matrix/Path product* to regain the loss in the expansion at the expense of a mild increase in the degree. The actual definition is as follows:

- Let H be a $(D = (2d)^{100}, d, 0.01)$-expander graph, which we can find using brute force search. (We choose d to be a large enough constant that such a graph exists) We let G_1 be a $((2d)^{100}, 2d, 1/2)$-expander graph and G_2 be a $((2d)^{200}, 2d, 1/2)$-expander graphs (again, such graphs can be easily found via brute force).
- For $k > 2$ define
$$G_k = \left(G_{\lfloor \frac{k-1}{2} \rfloor} \otimes G_{\lfloor \frac{k-1}{2} \rfloor}\right).$$

We prove the following claim:

CLAIM: For every k, G_k is a $((2d)^{100k}, 2d, 1 - 1/50)$-expander graph. Furthermore, there is a poly(k)-time algorithm that given a label of a vertex i in G_k and an index j in $[2d]$ finds the j^{th} neighbor of i in G_k.

PROOF: We prove the first part by induction. Verify directly that it holds for $k = 1, 2$. For $k > 2$, if we let n_k be the number of vertices of G_k then $n_k = n^2_{\lfloor \frac{k-1}{2} \rfloor}(2d)^{100}$. By induction we assume $n_{\lfloor \frac{k-1}{2} \rfloor} = (2d)^{100\lfloor \frac{k-1}{2} \rfloor}$ which implies that $n_k = (2d)^{100k}$ (using the fact that $2\lfloor \frac{k-1}{2} \rfloor + 1 = k$). It's also easy to verify that G_k has degree $2d$ for every j: if G has degree $2d$ then $G \otimes G$ has degree $(2d)^2$, $(G \otimes G)^{50})$ has degree $(2d)^{100}$ and $(G \otimes G)^{50}) \circledR H$ has degree $(2d)$. The eigenvalue analysis also follows by induction: if $\lambda(G) \leq 1 - 1/50$ then $\lambda(G \otimes G)^{50} \leq 1/e < 1/2$. Hence, by Lemma 21.18, $\lambda((G \otimes G)^{50} \circledR H) \leq 1 - 1/2(0.99)^2/24 \leq 1 - 1/50$.

For the furthermore part, note that there is a natural algorithm to compute the neighborhood function of G_k that makes 100 recursive calls to the neighborhood function of $G_{\lfloor \frac{k-1}{2} \rfloor}$, thus running in time roughly $n^{\log 100}$. ■

The above construction and analysis yields an expander graph family containing an n vertex graph for every n of the form c^k for some constant c. The proof of Theorem 21.19 is completed by observing that one can transform an (n, d, λ)-expander graph to an (n', c^2d', λ)-expander graph for any $n/c \leq n' \leq n$ by joining together into a "mega-vertex" sets of at most c vertices (Exercise 21.12). ■

There are other known constructions of expanders that are more efficient in terms of computation time and relation between degree and the parameter λ than the product-based construction above. However, the proofs for these constructions are often more complicated and require deeper mathematical tools. Also, the replacement product (and its close cousin, the zig-zag product) have found applications beyond the constructions of expander graphs. One such application is the deterministic logspace algorithm for undirected connectivity described in the next section. The replacement product also found another application in the construction of combinatorial expanders with a greater expansion of small sets that what is implied by the parameter λ [**CapalboReVaWi02**].

21.4 Deterministic logspace algorithm for undirected connectivity.

This section describes a recent result of Reingold, showing that at least the most famous randomized logspace algorithm, the random walk algorithm for the problem UPATH of s-t-connectivity in undirected graphs (see Chapter 7) can be completely "derandomized."

Theorem 21.20 (Reingold's theorem [**Reingold05**])
UPATH $\in \mathbf{L}$.

Reingold describes a set of poly(n) walks starting from s such that if s is connected to t then one of the walks is guaranteed to hit t. The *existence* of such a small set of walks can be shown using the probabilistic method and Lemma 21.11. The point is that Reingold's enumeration of walks can be carried out deterministically in logspace.

21.4.1 Outline of idea.

As before we are interested in undirected graphs that may have parallel edges. When analyzing will assume that the input graph for the s-t connectivity problem is d-regular for say $d = 4$. This is without loss of generality: if a vertex has degree $d'' < 3$ we add a self-loop of multiplicity to bring the degree up to d, and if the vertex has degree $d' \geq 3$ we can replace it by a cycle of d' vertices, and each of the d' edges that were incident to the old vertex then attach to one of the cycle nodes. Of course, a logspace machine does not have space to store the modified graph, but it can pretend that these modifications have taken place, since it can perform them on the fly whenever it accesses the graph. (Formally speaking, the transformation is implicitly computable in logspace as per Definition 4.16.) In fact, the proof below will perform a series of other local modifications on the graph, each with the property that the logspace algorithm can perform them on the fly.

Recall that checking connectivity in *expander* graphs is easy. Specifically, if every connected component in G is an expander, then there is a number $\ell = O(\log n)$ such that if s and t are connected then they are connected with a path of length at most ℓ. Indeed, Lemma 21.3 implies that in every n-vertex regular graph G, the distribution of the ℓ^{th} vertex in a random walk is within $\sqrt{n}\lambda^\ell$ statistical (or L_1) distance from the uniform distribution. In particular this means that if each connected component H of G is an *expander* graph, having $\lambda(H)$ bounded away from 1, then a random walk of length $\ell = O(\log n)$ from a vertex u in H will reach every vertex of H with positive probability.

The idea behind Reingold's algorithm is to transform the graph G (in an implicitly computable in logspace way) to a graph G' such that every connected component in G becomes an expander in G', but two vertices that were not connected will stay unconnected.

21.4.2 The logspace algorithm for connectivity (proof of Theorem 21.20)

By adding more self-loops we may assume that the input graph G is of degree d^{50} for some constant d that is sufficiently large to ensure the existence of a $(d^{50}, d/2, 0.01)$-expander graph H. Since the size of H is some constant, we assume the algorithm has access to it (either H could be "hardwired" into the algorithm or the algorithm could perform brute force search to discover it). Let $G_0 = G$ and for $k \geq 1$, define $G_k = (G_{k-1} \circledR H)^{50}$, where $\circledR$ denotes the balanced replacement product defined in Section 21.3.4.

If G_{k-1} is an N-vertex graph with degree d^{50}, then $G_{k-1} \circledR H$ is a $d^{50}N$-vertex graph with degree d and thus $G_k = (G_{k-1} \circledR H)^{50}$ is a $d^{50}N$-vertex graph with degree d. Note also that if two vertices were connected (resp., disconnected) in G_{k-1}, then they are still connected (resp., disconnected) in

G_k. The key observation that the graph $G_{10\log n}$ is an expander, and therefore an easy instance of UPATH. Specifically, we have:

CLAIM: For every k, G_k is an $(d^{50k}n, d^{20}, \max\{1-1/20, 2^k/(12n^2)\})$-graph, where n denotes the number of vertices in $G = G_0$.

PROOF: Indeed, by Lemmas 21.16 and 21.18, for every $\epsilon < 1/20$ and D-degree graph F, if $\lambda(F) \leq 1-\epsilon$ then $\lambda(F \circledR H) \leq 1-\epsilon/25$ and hence $\lambda\left((F \circledR H)^{50}\right) \leq 1-2\epsilon$. By Lemma 21.4, every connected component of G has expansion parameter at most $1 - \frac{1}{12n^2}$ (note that n is at least as large as the number of vertices in the connect component). It follows that for $k = 10\log n$, in the graph G_k every connected component has expansion parameter at most $\max\{1-1/20, 2^k/(12n^2)\} = 1-1/20$. ■

Since $G_{10\log n}$ is an expander, to find whether a pair of vertices s, t are connected in $G_{10\log n}$ we simply need to enumerate over all $\ell = O(\log n)$-long paths from s in G_k and see whether any one of these hits t. The catch is of course that the graph we are given is G, not $G_{10\log n}$. Therefore the question is whether for $k = 10\log n$, given a description of a vertex s in G_k and an index $i \in [d^{20}]$, we can compute the i^{th} neighbor of s in G_k using only logarithmic space. (Once we can perform a single step of G_k it's easy to perform ℓ steps in logspace as well by repeating them one after the other, and reusing the same space to compute each step.)

The graph G_k is equal to $(G_{k-1} \circledR H)^{50}$ and thus it suffices to show that we can take a single step in the graph $G_{k-1} \circledR H$ in logspace (we can then repeat the same process for 50 times). Now by the definition of the replacement product, a vertex in $G_{k-1} \circledR H$ is represented by a pair $\langle u, v\rangle$ where u is a vertex of G_{k-1} and v is a vertex of H. The index of a neighbor of $\langle u, v\rangle$ is represented by a pair $\langle b, i\rangle$ where $b \in \{0,1\}$ and $i \in [d/2]$. If $b = 0$ then the $\langle b, i\rangle^{th}$ neighbor of $\langle u, v\rangle$ is $\langle u, v'\rangle$ where v' is the i^{th} neighbor of v' in H. If $b = 1$ then the $\langle b, i\rangle^{th}$ neighbor of $\langle u, v\rangle$ is the pair $\langle u', v'\rangle$ denoting the result of applying G_{k-1}'s rotation map to $\langle u, v\rangle$. (That is, u' is the v^{th} neighbor of u in G_{k-1}, and v' is the index of u as a neighbor of u' in G_{k-1}.) This description already implies an obvious recursive algorithm to compute the rotation map of G_k. Letting s_k denotes the space needed to compute a rotation map of G_k by this algorithm, we see that s_k satisfies the equation $s_k = s_{k-1} + O(1)$, implying that $s_{10\log n} = O(\log n)$.[7] ■

21.5 Weak Random Sources and Extractors

Suppose, that despite any philosophical difficulties, we are happy with probabilistic algorithms, and see no need to "derandomize" them, especially at the expense of some unproven assumptions. We still need to tackle the fact that real world sources of randomness and unpredictability rarely, if ever, behave as a sequence of perfectly uncorrelated and unbiased coin tosses. Can we still execute probabilistic algorithms using real-world "weakly random" sources?

21.5.1 Min Entropy

For starters, we need to define what we mean by a weakly random source.

Definition 21.21

Let X be a random variable. The *min entropy* of X, denoted by $H_\infty(X)$, is the largest real number k such that $\Pr[X = x] \leq 2^{-k}$ for every x in the range of X.

If X is a distribution over $\{0,1\}^n$ with $H_\infty(X) \geq k$ then it is called an (n,k)-*source*.

[7] When implementing the algorithm one needs to take care *not* to make a copy of the input when invoking the recursive procedure, but rather have all procedure operate on a globally accessible memory that contains the index k and the vertex and edge labels. For more details see the original paper [**Reingold05**] or [**goldreichbook**, Section 5.2.4].

It is not hard to see that if X is a random variable over $\{0,1\}^n$ then $H_\infty(X) \leq n$ with $H_\infty(X) = n$ if and only if X is distributed according to the uniform distribution U_n. Our goal in this section is to be able to execute probabilistic algorithms given access to a distribution X with $H_\infty(X)$ as small as possible. It can be shown that min entropy is a *minimal requirement* in the sense that in general, to execute a probabilistic algorithm that uses k random bits we need access to a distribution X with $H_\infty(X) \geq k$ (see Exercise [**expand:ex:minentnec**]).

Example 21.22
Here are some examples for distributions X over $\{0,1\}^n$ and their min-entropy:

- (Bit fixing and generalized bit fixing sources) If there is subset $S \subseteq [n]$ with $|S| = k$ such that X's projection to the coordinates in S is uniform over $\{0,1\}^k$, and X's projection to $[n] \setminus S$ is a fixed string (say the all-zeros string) then $H_\infty(X) = k$, we call such a random variable a *bit-fixing* source. The same holds if X's projection to $[n] \setminus S$ is a fixed deterministic function of its projection to S, in which case we say that X is a *generalized bit-fixing source*. For example, if the bits in the odd positions of X are independent and uniform and for every even position $2i$, $X_{2i} = X_{2i-1}$ then $H_\infty(X) = \lceil \frac{n}{2} \rceil$. This may model a scenario where we measure some real world data at too high a rate (think of measuring every second a physical event that changes only every minute).

- (Linear subspaces) If X is the uniform distribution over a linear subspace of $\mathrm{GF}(2)^n$ of dimension k, then $H_\infty(X) = k$. (In this case X is actually a generalized bit-fixing source — can you see why?)

- (Biased coins) If X is composed of n independent coins, each outputting 1 with probability $\delta < 1/2$ and 0 with probability $1-\delta$, then as n grows, $H_\infty(X)$ tends to $H(\delta)n$ where H is the Shannon entropy function. That is, $H(\delta) = \delta \log \frac{1}{\delta} + (1-\delta) \log \frac{1}{1-\delta}$.

- (Santha-Vazirani sources) If X has the property that for every $i \in [n]$, and every string $x \in \{0,1\}^{i-1}$, conditioned on $X_1 = x_1, \ldots, X_{i-1} = x_{i-1}$ it holds that both $\Pr[X_i = 0]$ and $\Pr[X_i = 1]$ are between δ and $1-\delta$ then $H_\infty(X) \geq H(\delta)n$. This can model sources such as stock market fluctuations, where current measurements do have some limited dependence on the previous history.

- (Uniform over subset) If X is the uniform distribution over a set $S \subseteq \{0,1\}^n$ with $|S| = 2^k$ then $H_\infty(X) = k$. As we will see, this is a very general case that "essentially captures" all distributions X with $H_\infty(X) = k$.

We see that min entropy is a pretty general notion, and distributions with significant min entropy can model many real-world sources of randomness.

Next we formalize what it means to *extract* random —more precisely, almost random— bits from an (n,k) source. We will use the notion of *statistical distance* (see Section A.3.6 in the appendix) to quality when two distributions are close.

21.5.2 Definition of randomness extractors

We can now define randomness extractors - these are functions that transform an (n,k) source into an almost uniform distribution. The extractor uses a small number of additional truly random bits, denoted by d in the definition below.

Definition 21.23 (Randomness extractors)
A function $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (k,ϵ) extractor if for any (n,k)-source X, the distribution $\mathsf{Ext}(X,U_d)$ is ϵ-close to U_m. (For every ℓ, U_ℓ denotes the uniform distribution over $\{0,1\}^\ell$.)

Why an additional input? Our stated motivation for extractors is to execute probabilistic algorithms without access to perfect unbiased coins. Yet, it seems that an extractor is not sufficient for this task, as we only guarantee that its output is close to uniform if it is given an additional input (called the *seed* of the extractor) that is uniformly distributed. First, we note that the requirement of an additional input is necessary: for every function $\mathsf{Ext} : \{0,1\}^n \to \{0,1\}^m$ and every $k \leq n-1$ there exists an (n,k)-source X such that the first bit of $\mathsf{Ext}(X)$ is constant (i.e, is equal to some value $b \in \{0,1\}$ with probability 1), and so is at least of statistical distance $1/2$ from the uniform distribution (Exercise 21.13). Second, if the length t of the second input is sufficiently short (e.g., $t = O(\log n)$) then, for the purposes of simulating probabilistic algorithms, we can do without any access to true random coins, by enumerating over all the 2^t possible inputs (see Section [**expand:subsec:extractsim**]). Clearly, d has to be somewhat short for the extractor to be non-trivial: for $d \geq m$, we can have a trivial extractor that ignores its first input and outputs the second input.

21.5.3 Existence proof for extractors.

It turns out that at least if we ignore issues of computational efficiency, very good extractors exist:

Theorem 21.24
For every k, n, ϵ, there exists a (k,ϵ)-extractor $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^k$ with $d = \log n + 2\log(1/\epsilon) + O(1)$

PROOF: Call an (n,k) source X *flat* if X is the uniform distribution over a 2^k-sized subset of $\{0,1\}^n$. In Exercise 19.7 it is shown that every (n,k) source can be expressed as a convex combination of flat (k,n)-sources, which means that it suffices to show a function Ext such that $\mathsf{Ext}(X,U_d)$ is close to the uniform distribution when X is an (n,k)-flat source.

We will prove the existence of such an extractor by the probabilistic method, choosing Ext as a random function from $\{0,1\}^n \times \{0,1\}^d \to \{0,1\}^k$. Let X be an (n,k) flat source and let f be a function from $\{0,1\}^k \to \{0,1\}$. If we choose Ext at random then the expectation $\mathsf{E}[f(\mathsf{Ext}(X,U_d))]$ is obtained by evaluating f on $2^k \times 2^d$ random points, and hence by the Chernoff bound the probability that this expectation deviates from $\mathsf{E}[f(U_k)]$ by more than ϵ is bounded by $2^{-2^{k+d}/4\epsilon^2}$. This means that if $d > \log n + 2\log(1/e) + 3$ then this probability is bounded by $2^{-2n(2^k)}$. But the number of flat distributions is at most $(2^n)^{2^k}$ and the number of functions from $\{0,1\}^k \to \{0,1\}$ is 2^{2^k} and hence the union bound implies that there is a choice of Ext guaranteeing

$$|\mathsf{E}[f(\mathsf{Ext}(X,U_d))] - \mathsf{E}[f(U_k)]| < \epsilon$$

for *every* (n,k)-flat source and function $f : \{0,1\}^k \to \{0,1\}$. In other words, $\mathsf{Ext}(X,U_d)$ is ϵ-close to U_k for every (n,k)-flat source and hence for every (n,k)-source. ■

This extractor is optimal in the sense that there is an absolute constant c such that every (k,ϵ) extractor that is non-trivial (has output longer than seed length and $\epsilon < 1/2$) must satisfy $d \geq \log(n-k) + 2\log(1/\epsilon) - c$ [**NisanZu96**, **RadhakrishnanTa00**].

21.5.4 Extractors based on hash functions

The non-explicit extractor of Theorem 21.24 is not very useful: for most applications we need *efficiently computable* extractors. One such efficiently computable extractor (though with very bad parameters) can be obtained using pairwise independent hash functions.

Recall that a collection $\mathcal{H}$ of functions from $\{0,1\}^n$ to $\{0,1\}^m$ is *pairwise independent* if for every $x \neq x'$ in $\{0,1\}^n$ and $y, y' \in \{0,1\}^m$, the probability that $h(x) = y$ and $h(x') = y'$ for a random $h \in_R \mathcal{H}$ is 2^{-2m} (see Section 8.2.1). There are efficiently computable such collections one can choose a random function in $\mathcal{H}$ by picking a string at random from $\{0,1\}^d$ (for $d = n+m$). We will use h to denote both this string and the function. We define an extractor $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$, where $m = k - 2\log(1/\epsilon)$ as follows:

$$\mathsf{Ext}(x,h) = h(x) \circ h, \tag{11}$$

where $\circ$ denotes concatenation of strings. Its analysis is given by the following famous lemma:

Lemma 21.25 (Leftover hash lemma [**ILL89**, **BC??**]) $\delta(\mathsf{Ext}(X, U_d), U_{m+d}) \leq \epsilon$

PROOF: We study the *collision probability* of $\mathsf{Ext}(X, U_d)$. The probability that $h(x) \circ h = h'(x') \circ h'$ for random $h, h' \in_R \mathcal{H}$ and $x, x' \in_R X$ is bounded by the probability that $h = h'$ (which is equal to 2^{-d}) times the probability that $h(x) = h(x')$. The latter is bounded by 2^{-k} (a bound on the probability that $x = x'$ implies by the fact that X is a (k,n)-source) plus 2^{-m} (the probability that $h(x) = h(x')$ for a random $h \in_R \mathcal{H}$ and $x \neq x'$). Thus the collision probability of $\mathsf{Ext}(X, U_d)$ is at most $2^{-d}(2^{-k} + 2^{-m}) = 2^{-(d+m)} + 2^{-d-k}$.

Now, treat this distribution as a probability vector $\mathbf{v} \in \mathbb{R}^{2^{d+m}}$. Then the collision probability is precisely the L_2-norm of $\mathbf{v}$ squared and hence $\|\mathbf{v} - \mathbf{1}\|_2 \leq 2^{-(d-k)/2}$, where $\mathbf{1}$ denotes the probability vector corresponding to the uniform distribution U_{m+d}. Applying the relation between the L_1 and L_2 norms we get that

$$\delta(\mathsf{Ext}(X, U_d), U_{m+d}) = 1/2|\mathbf{v} - \mathbf{1}|_1 \leq 1/2 2^{(m+d)/2}\|\mathbf{v} - \mathbf{1}\|_2 \leq 1/2 2^{k/2+d/2-\log(1/\epsilon)} 2^{-k/2-d/2} < \epsilon\,.$$

■

Lemma 21.26 (Leftover hash lemma) *If x is chosen from a distribution on $\{0,1\}^n$ with min-entropy at least k/δ and $\mathcal{H}$ has collision error δ, then $h(X) \circ h$ has distance at most $\sqrt{2\delta}$ to the uniform distribution.*

PROOF: Left as exercise. (Hint: use the relation between the L_2 and L_1 norms ■

21.5.5 Extractors based on random walks on expanders

We can also use construct explicit extractors using expander graphs:

Lemma 21.27 *Let $\epsilon > 0$. For every n and $k \leq n$ there exists a (k,ϵ)-extractor $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^n$ where $t = O(n - k + \log 1/\epsilon)$.*

PROOF: Suppose X is an (n,k)-source and we are given a sample a from it. Let G be a $(2^n, d, 1/2)$-expander graph for some constant d (see Definition 21.6 and Theorem 21.19).

Let z be a truly random seed of length $t = 10\log d(n - k + \log 1/\epsilon) = O(n - k + \log 1/\epsilon)$. We interpret z as a random walk in G of length $10(n - k + \log 1/\epsilon)$ starting from the node whose label is a. (That is, we think of z as $10(n - k + \log 1/\epsilon)$ labels in $[d]$ specifying the steps taken in the walk.) The output $\mathsf{Ext}(a, z)$ of the extractor is the label of the final node on the walk.

We have $\|X-\mathbf{1}\|_2^2 \leq \|X\|_2^2 = \sum_a \Pr[X=a]^2$, which is at most 2^{-k} by Lemma [**expand:lem:collprob**] since X is an (n,k)-source. Therefore, after a random walk of length t the distance to the uniform distribution is (by the upper bound in ([**expand:eqn:expandeexpand**])):

$$\|M^t X - \frac{1}{2^N}\mathbf{1}\|_1 \leq \lambda_2^t \|X - \frac{1}{2^N}\mathbf{1}\|_2 \sqrt{2^N} \leq \lambda_2^t 2^{(N-k)/2}.$$

When t is a sufficiently large multiple of $N-k+\log 1/\varepsilon$, this distance is smaller than ε. ■

21.5.6 Extractors from pseudorandom generators

For many years explicit constructions of randomness extractors fell very far behind the parameters achieved by the optimal non-explicit construction of Theorem 21.24. In particular, researchers had no construction of, say, an $(n^{1/3}, 1/10)$-extractor with $O(\log n)$ seed size and say $n^{1/4}$ bits of output (parameters which are useful if one wants to be able to simulate a probabilistic polynomial-time algorithm that uses m random bits using an $(m^{1.4}, m^4)$-source). Then in 1999 Trevisan showed a vastly improved extractor construction. But what was more stunning than the result itself was Trevisan's technique: he showed that *pseudorandom generators* such as the ones we've seen in chapters 20 and 19, when viewed in the right way, are in fact also randomness extractors. This was very surprising, since these pseudornadom generators rely on *hardness assumptions* (such as the existence of a function in **E** with high circuit complexity). Thus it would seem that these generators will not be useful in the context of randomness extractors, where we are looking for constructions with *unconditional* analysis and are not willing to make any unproven assumptions.

What Trevisan noticed is that while we normally think of a pseudorandom generator G as having only one input, we can think of it as a function that takes two inputs: a short seed and the truth table of a supposedly hard function f. While our theorems state that the pseudorandom generator works if f is a hard function, the proofs of these theorems are actually *constructive*: they transform a distinguisher D that distinguishes between the generator's output and a random string into a small circuit A that computes the function f. In fact, the circuit A only uses the distinguisher D as a *black-box*. Therefore we can apply this transformation even when the distinguisher D is an arbitrary function that does not necessarily have a small function.

This means that we can rephrase the result on pseudorandom generators from worst-case assumptions (Theorem 19.1) as follows:

Theorem 21.28 (Theorem 19.1, constructive version)
For every time-constructible function $S : \mathbb{N} \to \mathbb{N}$, there is a constant c and algorithms G and R satisfying the following:

- *On input a function $f : \{0,1\}^\ell \to \{0,1\}$ and a string $z \in \{0,1\}^{c\ell}$, algorithm G runs in $2^{O(\ell)}$ time and outputs a string of length $m = S(\ell)^{1/c}$, which we denote by $G^f(z)$.*
- *If $D : \{0,1\}^m \to \{0,1\}$ is a function such that $\left|\mathsf{E}[D(G^f(U_{c\ell}))] - \mathsf{E}[D(U_m)]\right| > 1/10$ then there is an advice string a of length at most $S(\ell)^{1/4}$ such that on every input x, $R^D(a,x) = f(x)$ and furthermore R runs in time at most $S(\ell)^{1/4}$.*

The algorithm R is the algorithm for the reduction that is implicit in the proof of all the pseudorandom generators that we've seen. If we know that f's circuit complexity is at least $S(\ell)$ then the existence of R means that there is no $S(\ell)^{1/4}$-sized distinguisher between $G(U_{c\ell})$ and U_m. Looking at pseudorandom generators in this way, Trevisan saw that they can be viewed as "extracting" or "distilling" the hardness that is present in the function f into a random-looking output. Thus he proved the following lemma:

Lemma 21.29 *For every k, n let* Ext *be the function that on input $f \in \{0,1\}^n$ and $z \in \{0,1\}^{c \log n}$ outputs $G^f(z)$, where G is the generator obtained by Theorem 19.1 for $S = k$ and considering f as a function $f : \{0,1\}^\ell \to \{0,1\}$ for $\ell = \log n$. Then,* Ext *is a $(k, 1/5)$ generator.*

PROOF: Suppose otherwise, that there is a (k,n)-source X and a function D that distinguishes between $\mathsf{Ext}(X, U_{c\ell})$ and U_m with bias at least $1/5$. Then, with probability at least $1/10$ over $f \in_R X$, D distinguishes between $G^f(U_{c\ell})$ and U_m with bias at least $1/10$. Let's call an f for which this happens "bad". Note that for every bad f there exists an advice string $a \in \{0,1\}^{k^{1/4}}$ such that f is computed by the algorithm $x \mapsto R^D(a,x)$. This means that the number of bad f's is at most $2^{k^{1/4}}$. But since X is a k-source this means that the probability of a random f being bad is at most $2^{k^{1/4}} 2^{-k} \ll 1/10$, and hence we've arrived at a contradiction. ■

21.6 Pseudorandom generators for space bounded computation

We now show how extractors can be used to obtain a pseudorandom generator for space-bounded randomized computation, which allows randomized logspace computations to be run with $O(\log^2 n)$ random bits. We stress that this generator does not use any unproven assumptions.

Theorem 21.30 (Nisan's pseudorandom generator [**Nisan??**])
For every d there is a $c > 0$ and a poly(n)-time computable function $g : \{0,1\}^{c \log^2 n} \to \{0,1\}^{n^d}$ such that for every space-bounded machine M that has a configuration graph of size $\leq n^d$ on inputs of size n:

$$\left| \Pr_{r \in \{0,1\}^{n^d}} [M(x,r) = 1] - \Pr_{z \in \{0,1\}^{c \log^2 n}} [M(x, g(z)) = 1] \right| < \frac{1}{10}. \qquad (12)$$

Nisan's theorem implies that there is a polynomial-time simulation of every algorithm in **BPL** using $O(\log^2 n)$ space. Note that Savitch's theorem (Theorem [**thm:savitch**]) also implies that $\mathbf{BPL} \subseteq \mathbf{SPACE}(\log^2 n)$, but the algorithm in Savitch's proof takes $n^{\log n}$ time. Saks and Zhou [**saksZ99**] improved Nisan's ideas to show that $\mathbf{BPL} \subseteq \mathbf{SPACE}(\log^{1.5} n)$, which leads many experts to conjecture that $\mathbf{BPL} = \mathbf{L}$ (i.e., randomness does not help logspace computations at all). Indeed, we've seen in Section 21.4 that the famous random-walk algorithm for undirected connectivity can be derandomized in logspace.

The proof of Theorem 21.30 we present uses randomness extractors and is due to Impagliazzo, Nisan, and Wigderson [**impagliazzoNW94**] (with further improvements by Raz and Reingold [**razR99**]). Nisan's original paper did not explicitly use extractors— the definition of extractors came later and was influenced by results such as Nisan's. The main intuition behind Nisan's construction —and also the conjecture $\mathbf{BPL} = \mathbf{L}$— is that the logspace machine has one-way access to the random string and only $O(\log n)$ bits of memory. So it can only "remember" $O(\log n)$ of the random bits it has seen. To exploit this we will use the following simple lemma, which shows how to recycle a random string about which only a little information is known.

Lemma 21.31 (Recycling lemma) *Let $f : \{0,1\}^n \to \{0,1\}^s$ be any function and* Ext $: \{0,1\}^n \times \{0,1\}^t \to \{0,1\}^m$ *be a $(k, \epsilon/2)$-extractor, where $k = n - (s+1) - \log \frac{1}{\epsilon}$. When $X \in_R \{0,1\}^n$, $W \in_R \{0,1\}^m$, $z \in_R \{0,1\}^t$, then*

$$f(X) \circ W \approx_\epsilon f(X) \circ \mathsf{Ext}(X, z).$$

When the lemma is used, $s \ll n$ and $n = m$. Thus $f(X)$, which has length s, contains only a small amount of information about X. The Lemma says that using an appropriate extractor (whose random seed can have length as small as $t = O(s + \log(1/\epsilon))$ if we use Lemma 21.27) we can get a new string $\mathsf{Ext}(X, z)$ that looks essentially random, even to somebody who knows $f(X)$. PROOF: For $v \in \{0,1\}^s$ we denote by X_v the random variable that is uniformly distributed over the set $f^{-1}(v)$. Then we can express $\delta(f(X) \circ W, f(X) \circ \mathsf{Ext}(X, z))$ as

$$= \frac{1}{2} \sum_{v,w} \left| \Pr[f(X) = v \wedge W = w] - \Pr_z[f(X) = v \wedge \mathsf{Ext}(X, z) = w] \right|$$

$$= \sum_v \Pr[f(X) = v] \cdot \delta(W, \mathsf{Ext}(X_v, z)) \tag{13}$$

Let $V = \left\{ v : \Pr[f(X) = v] \geq \epsilon/2^{s+1} \right\}$. If $v \in V$, then we can view X_v as a (n, k)-source, where $k \geq n - (s + 1) - \log \frac{1}{\epsilon}$. Thus by definition of an extractor, $\mathsf{Ext}(X_v, r) \approx_{\epsilon/2} W$ and hence the contributions from $v \in V$ sum to at most $\epsilon/2$. The contributions from $v \notin V$ are upperbounded by $\sum_{v \notin V} \Pr[f(X) = v] \leq 2^s \times \frac{\epsilon}{2^{s+1}} = \epsilon/2$. The lemma follows. ■

Now we describe how the Recycling Lemma is useful in Nisan's construction. Let M be a logspace machine. Fix an input of size n and view the graph of all configurations of M on this input as a *leveled branching program*. For some $d \geq 1$, M has $\leq n^d$ configurations and runs in time $L \leq n^d$. Assume without loss of generality —since unneeded random bits can always be ignored— that it uses 1 random bit at each step. Without loss of generality (by giving M a separate worktape that maintains a time counter), we can assume that the configuration graph is leveled: it has L levels, with level i containing configurations obtainable at time i. The first level contains only the start node and the last level contains two nodes, "accept" and "reject;" every other level has $W = n^d$ nodes. Each level i node has two outgoing edges to level $i + 1$ nodes and the machine's computation at this node involves using the next bit in the random string to pick one of these two outgoing edges. We sometimes call L the *length* of the configuration graph and W the *width*.

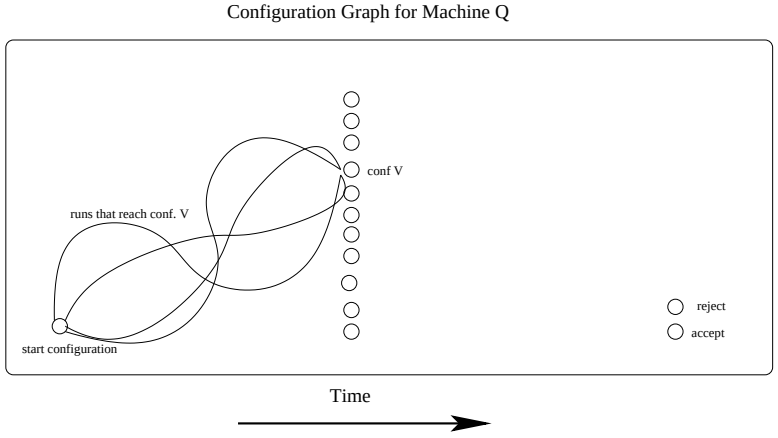


Figure 21.2: Configuration graph for machine M

For simplicity we first describe how to reduce the number of random bits by a factor 2. Think of the L steps of the computation as divided in two halves, each consuming $L/2$ random bits. Suppose we use some random string X of length $L/2$ to run the first half, and the machine is now at node v in the middle level. The only information known about X at this point is the index of v, which is a string of length $d \log n$. We may thus view the first half of the branching program as a (deterministic) function that maps $\{0,1\}^{L/2}$ bits to $\{0,1\}^{d \log n}$ bits. The Recycling Lemma

allows us to use a random seed of length $O(\log n)$ to recycle X to get an almost-random string $\mathsf{Ext}(X, z)$ of length $L/2$, which can be used in the second half of the computation. Thus we can run L steps of computation using $L/2 + O(\log n)$ bits, a saving of almost a factor 2. Using a similar idea recursively, Nisan's generator runs L steps using $O(\log n \log L)$ random bits.

Now we formally define Nisan's generator.

Definition 21.32 (Nisan's generator)
For some $r > 0$ let $\mathsf{Ext}_k : \{0,1\}^{kr} \times \{0,1\}^r \to \{0,1\}^{kr}$ be an extractor function for each $k \geq 0$. For every integer $k \geq 0$ the associated Nisan generator $G_k : \{0,1\}^{kr} \to \{0,1\}^{2^k}$ is defined recursively as (where $|a| = (k-1)r, |z| = r$)

$$G_k(a \circ z) = \begin{cases} z_1 \quad \text{(i.e., first bit of } z) & k = 1 \\ G_{k-1}(a) \circ G_{k-1}(\mathsf{Ext}_{k-1}(a, z)) & k > 1 \end{cases}$$

Now we use this generator to prove Theorem 21.30. We only need to show that the probability that the machine goes from the start node to the "accept" node is similar for truly random strings and pseudorandom strings. However, we will prove a stronger statement involving intermediate steps as well.

If nodes u is a node in the configuration graph, and s is a string of length 2^k, then we denote by $f_{u,2^k}(s)$ the node that the machine reaches when started in u and its random string is s. Thus if s comes from some distribution $\mathcal{D}$, we can define a distribution $f_{u,2^k}(\mathcal{D})$ on nodes that are 2^k levels further from u.

Theorem 21.33
Let $r = O(\log n)$ be such that for each $k \leq d \log n$, $\mathsf{Ext}_k : \{0,1\}^{kr} \times \{0,1\}^r \to \{0,1\}^{kr}$ is a $(kr - 2d\log n, \epsilon)$-extractor. For every machine of the type described in the previous paragraphs, and every node u in its configuration graph:

$$\delta(f_{u,2^k}(U_{2^k}), f_{u,2^k}(G_k(U_{kr}))) \leq 3^k \epsilon, \tag{14}$$

where U_l denotes the uniform distribution on $\{0,1\}^l$.

To prove Theorem 21.30 let $u = u_0$, the start configuration, and $2^k = L$, the length of the entire computation. Choose $3^k\epsilon < 1/10$ (say), which means $\log 1/\epsilon = O(\log L) = O(\log n)$. Using the extractor of Section 21.5.5 as Ext_k, we can let $r = O(\log n)$ and so the seed length $kr = O(r \log L) = O(\log^2 n)$.

Proof: (Theorem 21.33) Let ϵ_k denote the maximum value of the left hand side of (14) over all machines. The lemma is proved if we can show inductively that $\epsilon_k \leq 2\epsilon_{k-1} + 2\epsilon$. The case $k = 1$ is trivial. At the inductive step, we need to upper bound the distance between two distributions $f_{u,2^k}(\mathcal{D}_1)$, $f_{u,2^k}(\mathcal{D}_4)$, for which we introduce two distributions $\mathcal{D}_2, \mathcal{D}_3$ and use triangle inequality:

$$\delta(f_{u,2^k}(\mathcal{D}_1), f_{u,2^k}(\mathcal{D}_4)) \leq \sum_{i=1}^{3} \delta(f_{u,2^k}(\mathcal{D}_i), f_{u,2^k}(\mathcal{D}_{i+1}))\,. \tag{15}$$

The distributions will be:

$$\begin{aligned} \mathcal{D}_1 &= U_{2^k} \\ \mathcal{D}_4 &= G_k(U_{kr}) \\ \mathcal{D}_2 &= U_{2^{k-1}} \circ G_{k-1}(U_{(k-1)r}) \\ \mathcal{D}_3 &= G_{k-1}(U_{(k-1)r}) \circ G_{k-1}(U'_{(k-1)r}) \qquad (U, U' \text{ are identical but independent}). \end{aligned}$$

We bound the summands in (15) one by one.

Claim 1: $\delta(f_{u,2^k}(\mathcal{D}_1) - f_{u,2^k}(\mathcal{D}_2)) \leq \epsilon_{k-1}$.
Denote $\Pr[f_{u,2^{k-1}}(U_{2^{k-1}}) = w]$ by $p_{u,w}$ and $\Pr[f_{u,2^{k-1}}(G_{k-1}(U_{(k-1)r})) = w]$ by $q_{u,w}$. According to the inductive assumption,

$$\frac{1}{2}\sum_w |p_{u,w} - q_{u,w}| = \delta(f_{u,2^{k-1}}(U_{2^{k-1}}), f_{u,2^{k-1}}(G_{k-1}(U_{(k-1)r}))) \leq \epsilon_{k-1}.$$

Since $\mathcal{D}_1 = U_{2^k}$ may be viewed as two independent copies of $U_{2^{k-1}}$ we have

$$\delta(f_{u,2^k}(\mathcal{D}_1), f_{u,2^k}(\mathcal{D}_2)) = \sum_v \frac{1}{2}\left|\sum_w p_{uw}p_{wv} - \sum_w p_{uw}q_{wv}\right|$$

where w, v denote nodes 2^{k-1} and 2^k levels respectively from u

$$\begin{aligned} &= \sum_w p_{uw}\frac{1}{2}\sum_v |p_{wv} - q_{wv}| \\ &\leq \epsilon_{k-1} \quad \text{(using inductive hypothesis and } \sum_w p_{uw} = 1) \end{aligned}$$

Claim 2: $\delta(f_{u,2^k}(\mathcal{D}_2), f_{u,2^k}(\mathcal{D}_3)) \leq \epsilon_{k-1}$.

The proof is similar to the previous case and is omitted.

Claim 3: $\delta(f_{u,2^k}(\mathcal{D}_3), f_{u,2^k}(\mathcal{D}_4)) \leq 2\epsilon$.
We use the Recycling Lemma. Let $g_u : \{0,1\}^{(k-1)r} \to [1, W]$ be defined as $g_u(a) = f_{u,2^{k-1}}(G_{k-1}(a))$. (To put it in words, apply the Nisan generator to the seed a and use the result as a random string for the machine, using u as the start node. Output the node you reach after 2^{k-1} steps.) Let $X, Y \in U_{(k-1)r}$ and $z \in U_r$. According to the Recycling Lemma,

$$g_u(X) \circ Y \approx_\epsilon g_u(X) \circ \mathsf{Ext}_{k-1}(X, z),$$

and then Part 5 of Lemma A.26 implies that the equivalence continues to hold if we apply a (deterministic) function to the second string on both sides. Thus

$$g_u(X) \circ g_w(Y) \approx_\epsilon g_u(X) \circ g_w(\mathsf{Ext}_{k-1}(X, z))$$

for all nodes w that are 2^{k-1} levels after u. The left distribution corresponds to $f_{u,2^k}(\mathcal{D}_3)$ (by which we mean that $\Pr[f_{u,2^k}(\mathcal{D}_3) = v] = \sum_w \Pr[g_u(X) = w \wedge g_w(Y) = v]$) and the right one to $f_{u,2^k}(\mathcal{D}_4)$ and the proof is completed. ■

What have we learned?

- Often we can easily show that a random object has certain attractive properties, but it's non-trivial to come up with an *explicit* construction of an object with these properties. Yet, once found, such explicit constructions are often extremely useful.
- The behavior of random walks on a graph is tightly related to the eigenvalues of its adjacency matrix.
- An *expander* graph family is a collection of constant-degree graphs whose second largest eigenvalue is bounded away from 1. Such families can be shown to exist using the probabilistic method, but we also know of *explicit* constructions.
- An ℓ-step random walk on an expander graph is to a certain extend "pseudorandom" and behaves similarly to ℓ randomly chosen vertices under certain measures. This fact has been found useful in a variety of setting, from the randomness efficient error reduction procedure for **BPP** to the logspace algorithm for undirected connectivity.
- Extractors are functions that transform a distribution with a large min-entropy into (close to) the uniform distribution.
- Pseudorandom generators with a "black-box" analysis of their correctness can be used to construct randomness extractors, even though the latter are based on no unproven assumptions or lower bounds.

Chapter notes and history

Expanders were first studied by Pinsker [**Pinsker73**] who defined them and proved their existence with the first explicit construction given by Margulis [**Margulis73**]. However, his proof was existential in the sense that it showed that a certain explicit family of graphs satisfied $\lambda(G) < 1$ for every graph G in the family, but gave no explicit bound on that parameter λ. Gabber and Galil [**GabberGa81**] improved Margulis' analysis and gave an explicit bound on $\lambda(G)$, a bound that was later improved by Jimbo and Marouka [**JimboMa87**]. Lubotzky, Phillips and Sarnak [**LubotzkyPhSa88**] constructed *Ramanujan* graphs, that are expander with an optimal dependence between the parameter λ and their degree.

Lemma 21.4 (every connected graph has some spectral gap) is from Alon and Sudakov [**AlonSu00**] and is an improved version of a result appearing as Problem 11.29 in [**Lovasz07book**]. The Alon-Boppanna lower bound on the second eigenvalue of a d-regular graph was first stated in [**Alon86**]; a tight bound on the $o(1)$ error term was given in [**Nilli04**].

The relation between the algebraic (eigenvalue-based) and combinatorial definitions of expanders was developed by Dodziuk, Alon, and Milman in the papers [**Dodziuk84**, **AlonMi84**, **AlonMi85**, **Alon86**]. These results can be viewed as a discrete version of a result by Cheeger **[Cheeger70]** on compact Riemannian manifolds.

The application of expanders to pseudorandomness was first described by Ajtai, Komlos, and Szemeredi [**AKS87**]. Then Cohen and Wigderson [**cohenW89**] and Impagliazzo-Zuckerman [**ImpagliazzoZu89**] showed how to use them to "recycle" random bits as described in Section 21.2.5.

The explicit construction of expanders presented in Sectionřefexpand:sec:explicit is due to Reingold, Vadhan and Wigderson [**reingoldVW00**], although we chose to present it using the replacement product as opposed to the closely related zig-zag product used there. Our analysis of the

replacement product using Lemma 21.14 follows Rozenman and Vadhan [**RozenmanVa05**]. The expansion properties of the replacement product were also analyzed in a particular case of products of two cubes by Gromov [**Gromov83**] and for general graphs (in a somewhat different context) by Martin and Randall [**MartinRa00**].

The deterministic logspace algorithm for undirected connectivity is due to Reingold [**Reingold05**]. Trifonov [**Trifonov05**] proved concurrently and independently the slightly weaker result of an $O(\log n \log\log n)$-space algorithm for the same problem.

Hoory, Linial and Wigderson [**HooLinWig06**] give an excellent introduction to expander graphs and their computer science applications. The Alon-Spencer book [**AlonSp00**] also contains several results on expanders.

Weak random sources were first considered in the 1950s by von Neumann [**jpb:vonNeumann61**]. The second volume of Knuth's seminal work studies real-life pseudorandom generators and their limitations. The study of weak random sources as defined here started with Blum [**blum84**]. Progressively weaker models were then defined, culminating in the "correct" definition of an (N, k) source in Zuckerman [**zuckerman90**]. Zuckerman also observed that this definition generalizes all models that had been studied to date. (See [**srinivasanZ99**] for an account of various models considered by previous researchers.) He also gave the first simulation of probabilistic algorithms with such sources assuming $k = \Omega(N)$. A succession of papers has improved this result; for some references, see the paper of Lu, Reingold, Vadhan, and Wigderson [**LRVW03**], the current champion in this area (though very likely dethroned by the time this book appears).

The earliest work on extractors —in the guise of *leftover hash lemma* (Lemma 21.25) of Impagliazzo, Levin, and Luby [**ILL89**] and a related precursor by Bennett, Brassard and Robert [**BennettBrRo88**]— took place in context of cryptography. Nisan [**nisan92**] then showed that hashing could be used to obtain provably good pseudorandom generators for logspace. The notion of an extractor was first formalized by Nisan and Zuckerman [**nisanZ96**]. Trevisan [**trevisan01**] pointed out that any "black-box" construction of a pseudorandom generator gives an extractor, and in particular used the Nisan-Wigderson generator to construct extractors as described in the chapter. Since then a sequence of beautiful papers gave explicit constructions of extractors with improved parameters. In particular Lu, Reingold, Vadhan and Wigderson [**LRVW03**] were the first to obtain an extractor with the seed length and output length within a constant factor of the optimal non-explicit extractor of Theorem 21.24. The current state of the art extractor construction is by Guruswami, Umans and Vadhan [**GuruswamiUmVa07**].

Trevisan's insight about using pseudorandom generators to construct extractors has been greatly extended. It is now understood that three combinatorial objects studied in three different fields are very similar: pseudorandom generators (cryptography and derandomization), extractors (weak random sources) and list-decodable error-correcting codes (coding theory and information theory). Constructions of any one of these objects often gives constructions of the other two. For a survey, see Vadhan's lecture notes [**vadhan03**] and survey [**Vadhan07**].

Exercises

21.1. (a) Prove Hölder's Inequality (see Note 21.1): For every p, q with $\frac{1}{p} + \frac{1}{q} = 1$, $\|\mathbf{u}\|_p \|\mathbf{v}\|_q \geq \sum_{i=1}^{n} |\mathbf{u}_i \mathbf{v}_i|$.

Hint: Use the fact that the log function is concave (has negative second derivative) implying that for $a, b > 0$, $\alpha \log a + (1-\alpha)\log b \leq \log(\alpha a + (1-\alpha)b)$.

(b) Prove the inequality (1): $|\mathbf{v}|_1/\sqrt{n} \leq \|\mathbf{v}\|_2 \leq \sqrt{|\mathbf{v}|_1 \|\mathbf{v}\|_\infty}$ for every $\mathbf{v}$.

Hint: For the first inequality, use Hölder's inequality with $p = q = 2$ (deriving the Cauchy-Schwartz Inequality). For the second, use $p = 1$ and $q = \infty$.

21.2. Prove that for every n-vertex d-regular graph G, the diameter of G (maximum over all pairs of distinct vertices i, j in G of the length of the shortest path in G between i and j) is at most $3n/(d+1)$.

> **Hint:** Show that for every shortest path between two vertices, if we pick any third vertex in the path then the $d+1$-sized neighborhoods of all the picked vertices are disjoint.

21.3. Recall that the *norm* of a matrix A, denoted $\|A\|$, is defined as the maximum of $\|A\mathbf{v}\|_2$ for every unit vector $\mathbf{v}$. Let A be a symmetric stochastic matrix: i.e., $A = A^\dagger$ and every row and column of A has non-negative entries summing up to one. Prove that $\|A\| \leq 1$.

> **Hint:** first show that $\|A\|$ is at most say n^2. Then, prove that for every $k \geq 1$, A^k is also stochastic and $\|A^{2k}\mathbf{v}\|_2 \geq \|A^k\mathbf{v}\|_2^2$ using the equality $\langle \mathbf{w}, B\mathbf{z}\rangle = \langle B^\dagger \mathbf{w}, \mathbf{z}\rangle$ and the inequality $\langle \mathbf{w}, \mathbf{z}\rangle \leq \|\mathbf{w}\|_2 \|\mathbf{z}\|_2$.

21.4. Let A, B be two $n \times n$ matrices.

(a) Prove that $\|A + B\| \leq \|A\| + \|B\|$.

(b) Prove that $\|AB\| \leq \|A\|\|B\|$.

21.5. Let A, B be two symmetric stochastic matrices. Prove that $\lambda(A + B) \leq \lambda(A) + \lambda(B)$.

21.6. Prove Lemma 21.16.

> **Hint:** Use the fact that if A is a normalized adjacency matrix of a graph and $\mathbf{v} \perp \mathbf{1}$ then $A\mathbf{v} \perp \mathbf{1}$.

21.7. Recall that the *trace* of a Matrix A, denoted $\mathrm{tr}(A)$, is the sum of the entries along its diagonal.

(a) Prove that if an $n \times n$ matrix A has eigenvalues $\lambda_1, \ldots, \lambda_n$, then $\mathrm{tr}(A) = \sum_{i=1}^n \lambda_i$.

(b) Prove that if A is a normalized adjacency matrix of an n-vertex graph G, and $k \geq 1$, then $\mathrm{tr}(A^k)$ is equal to n times the probability that a if we select a vertex i uniformly at random and take a k step random walk from i, then we end up back in i.

(c) Prove that for every n-vertex d-degree graph G, $\lambda(G) \geq \frac{1}{\sqrt{d}}(1 + o(1))$, where $o(1)$ denotes a term, depending on n and d that tends to 0 as n grows.

> **Hint:** Setting $\lambda = \lambda(G)$, use the previous two items to prove that $1 + (n-1)\lambda^2 \geq n/d$.

21.8. Let an n, d random graph be an n-vertex graph chosen as follows: choose d random permutations $\pi_1, \ldots, \pi_d$ from $[n]$ to $[n]$. Let the the graph G contains an edge (u, v) for every pair u, v such that $v = \pi_i(u)$ for some $1 \leq i \leq d$. Prove that a random n, d graph is an $(n, 2d, \frac{1}{10})$ combinatorial edge expander with probability $1 - o(1)$ (i.e., tending to one with n).

> **Hint:** For every set $S \subseteq n$ with $|S| \leq n/2$, try to bound probability that the number of edges between S and $\bar{S}$ deviates strongly from its expectation.

21.9. Prove that for every n-vertex d-regular graph, there exists a subset S of $n/2$ vertices, such that $E(S, \bar{S}) \leq dn/4$. Conclude that there does not exist an (n, d, ρ)-combinatorial edge expander for $\rho > 1/2$.

Hint: Use the probabilistic method - choose S to be a random $n/2$-sized subset of the vertices. For every pair of distinct vertices u, v, the probability that $u \in S$ and $v \in \bar{S}$ or vice versa is at most $1/2$ (it would be exactly half if we chose S with replacements). Hence, since there are $dn/2$ edges in the graph, the expected value of in $E(S, \bar{S})$ is at most $dn/4$.

21.10. Prove the Expander Mixing Lemma (Lemma 21.11).

Hint: You can use Lemma 21.14.

21.11. [**Tanner84**]

(a) Prove that if $\mathbf{p}$ is a probability vector then $\|\mathbf{p}\|_2^2$ is equal to the probability that if i and j are chosen from $\mathbf{p}$, then $i = j$.

(b) Prove that if $\mathbf{s}$ is the probability vector denoting the uniform distribution over some subset S of vertices of a graph G with normalized adjacency matrix A, then $\|A\mathbf{p}\|_2^2 \geq 1/|\Gamma(S)|$, where $\Gamma(S)$ denotes the set of S's neighbors.

(c) Prove that if G is an (n, d, λ)-expander graph, and S is a subset of ϵn vertices, then

$$|\Gamma(S)| \geq \frac{|S|}{2\lambda^2\left((1-\epsilon)^2 - 2\epsilon/\lambda^2\right)}\,.$$

Hint: Show that if $\mathbf{s}$ is the uniform distribution over S then $\|A\mathbf{s}\|_2^2 \leq 1/n + \lambda^2(\epsilon n + 1/n)$.

A graph where $|\Gamma(S)| \leq c|S|$ for every not-too-big set S (say, $|S| \leq n/(10d)$) is said to have *vertex expansion* c. This exercise shows that graphs with the minimum possible second eigenvalue $\frac{2}{\sqrt{d}}(1 + o(1))$ have vertex expansion roughly $d/4$. It is known that such graphs have in fact vertex expansion roughly $d/2$ [**Kahale??**], and there are counterexamples showing this is tight. In contrast, random d-regular graphs have vertex expansion $(1 - o(1))d$.

21.12. If G is a graph and S is a subset of G's vertices then by *contracting* S we mean transforming G into a graph H where all of S's members are replaced by a single vertex s with an edge $\overline{s\,v}$ in H for every edge $\overline{u\,v}$ in G where $u \in S$. Prove that if G is an (n, d, λ)-graph, and $c \in \mathbb{N}$ divides n, then the graph H obtained by dividing the vertices of G to n/c sets of size c and contracting all these sets is an $(n/c, c^2 d, \lambda)$-graph.

Hint: Think about how this transformation affects the adjacency matrix of G.

21.13. Prove that for every function $\mathsf{Ext} : \{0,1\}^n \to \{0,1\}^m$ and there exists an $(n, n-1)$-source X and a bit $b \in \{0,1\}$ such that $\Pr[\mathsf{Ext}(X)_1 = b] = 1$ (where $\mathsf{Ext}(X)_1$ denotes the first bit of $\mathsf{Ext}(X)$). Prove that this implies that $\delta(\mathsf{Ext}(X), U_m) \geq 1/2$.

21.14. Show that there is a constant $c > 0$ such that if an algorithm runs in time T and requires m random bits, and $m > k + c\log T$, then it is not possible in general to simulate it in a blackbox fashion using an (N, k) source and $O(\log n)$ truly random bits.

Hint: For each source show that there is a randomized algorithm —it need not be efficient, since it is being used as a "black box"— for which the simulation fails.

21.15. A *flat* (N, k) source is a (N, k) source where for every $x \in \{0,1\}^N$ p_x is either 0 or exactly 2^{-k}.

Show that a source X is an (N, k)-source iff it is a distribution on flat sources. In other words, there is a set of flat (N, k)-sources $X_1, X_2, \ldots$ and a distribution $\mathcal{D}$ on them such that drawing

a sample of X corresponds to picking one of the X_i's according to $\mathcal{D}$, and then drawing a sample from X_i.

> **Hint:** You need to view a distribution as a point in a 2^N-dimensional space, and show that X is in the convex hull of the points that represent all possible flat sources.

21.16. Use Nisan's generator to give an algorithm that produces universal traversal sequences for n-node graphs (see Definition **[expand:def:uts]**) in $n^{O(\log n)}$-time and $O(\log^2 n)$ space.

21.17. Suppose Boolean function f is (S, ϵ)-hard and let D be the distribution on m-bit strings defined by picking inputs $x_1, x_2, \ldots, x_m$ uniformly at random and outputting $f(x_1)f(x_2)\cdots f(x_m)$. Show that the statistical distance between D and the uniform distribution is at most ϵm.

21.18. Prove Lemma 21.25.

21.19. Let A be an $n \times n$ matrix with eigenvectors $\mathbf{u}^1, \ldots, \mathbf{u}^n$ and corresponding values $\lambda_1, \ldots, \lambda_n$. Let B be an $m \times m$ matrix with eigenvectors $\mathbf{v}^1, \ldots, \mathbf{v}^m$ and corresponding values $\alpha_1, \ldots, \alpha_m$. Prove that the matrix $A \otimes B$ has eigenvectors $\mathbf{u}^i \otimes \mathbf{v}^j$ and corresponding values $\lambda_i \cdot \alpha_j$.

21.20. Prove that for every two graphs G, G', $\lambda(G \otimes G') \leq \lambda(G) + \lambda(G')$ without using the fact that every symmetric matrix is diagonalizable.

> **Hint:**Use Lemma 21.14.

21.21. Let G be an n-vertex D-degree graph with ρ combinatorial edge expansion for some $\rho > 0$. (That is, for every a subset S of G's vertices of size at most $n/2$, the number of edges between S and its complement is at least $\rho d|S|$.) Let G' be a D-vertex d-degree graph with ρ' combinatorial edge expansion for some $\rho' > 0$. Prove that $G \circledR G'$ has at least $\rho^2\rho'/1000$ edge expansion.

> **Hint:**Every subset of the replacement product of G and G' can be thought of as n subsets of the individual clusters. Treat differently the subsets that take up more than $1 - \rho/10$ portion of their clusters and those that take up less than that. For the former use the expansion of G, while for the latter use the expansion of G'.