CSCI 5060: Techniques in the theory of computing | **Lecture 11**
The Chinese University of Hong Kong | 11 April 2012

In the last lecture we saw an application of expander graphs. There are many others in computer science, and you can read about some of them in the book by Hoory, Linial, and Wigderson. However we left out an important question unanswered: Do infinite families of expander graphs even exist?

Obtaining expanding graph families requires some work, and I know of three different approaches for going about it. Each of them has some advantages and some drawbacks. Recall that we are looking for a family of d-regular graphs $\{G_n\}$ on n vertices such that $\lambda(G_n) > 1 - \varepsilon$ for some $\varepsilon > 0$ and all n.

Here is a quick summary of the three approaches for obtaining such families:

- **Probabilistic:** Using the probabilistic method, we can argue that a random d-regular graph on n vertices satisfies $\lambda(G_n) > 1 - \varepsilon$ with good probability. One advantage of the probabilistic approach is that it shows the existence of graphs with very strong vertex expansion properties: For every set S of vertices that is not too large, the number of neighbors of S is very close to the maximum possible value of $d|S|$. The main disadvantage of the probabilistic approach is that it is not explicit as it requires the use of randomness, and bad choices of the randomness yield non-expanding graphs.
- **Combinatorial:** This is an iterative construction of infinite expander families. Starting from a single graph G_0, one iteratively applies the operations of powering and zig-zag product (or replacement product), thereby obtaining a sequence of graphs $G_0, G_1, G_2, \ldots$ on an increasing number of vertices but with bounded degree. Powering makes sure that $\lambda(G_n) \leq \lambda(G_{n-1})$, while the zig-zag product keeps the degree bounded. This approach has one important application in complexity theory: It leads to a space-efficient algorithm for connectivity in undirected graphs.
- **Algebraic:** This approach exploits the algebraic and geometric properties of Cayley graphs, which are graphs defined from generators of finite groups (or infinite groups of which finite quotients are taken). One highlight of this approach is the construction of infinite families of Ramanujan graphs, i.e. graphs G_n such that $\lambda(G_n) \leq 2\sqrt{d-1}/d$. (Recall that $\lambda(G_n) \geq 2\sqrt{d-1}/d - o_n(1)$.) Some of the expander families obtained using this approach are very easy to describe (and implement).

In this lecture and the next one I want to show some ideas of the algebraic approach for constructing expander families. This is a deep theory of which we will barely scratch the surface, but we will see some unexpected connections with a seemingly unrelated object – small-biased sets.

1 Cayley graphs, Abelian groups, and small-biased sets

Recall that a *group* is a set with an operation which is associative ($(ab)c = a(bc)$) with an identity (there is an element 1 such that $a1 = 1a = a$ for every a) and inverses for all elements (for every a

there is an a^{-1} such that $aa^{-1} = a^{-1}a = 1$). A set S of group elements is called a *generating set* if every element of G can be written as a finite product of elements in S. Here we will only worry about finite groups.

Let S be a generating set for a group G which is closed under inverse (for every a in S, a^{-1} is also in S. The *Cayley graph* $\mathrm{Cay}(G, S)$ is the $|S|$-regular graph where there is a vertex for every element in G and an edge from g to sg for every $g \in G$ and $s \in S$. (Parallel edges and loops are allowed.) Since S is closed under inverse, this graph is undirected.

We are interested in constructing infinite families of Cayley graphs which are expanding. It is common to start with a specific family of groups $\{G_n\}$ and try to construct a set of generators S_n, $|S_n| \leq d$ for G_n so that $\lambda(\mathrm{Cay}(G_n, S_n)) \leq 1 - \varepsilon$.

To illustrate the connection between the algebra of the groups G_n and the expansion of the corresponding Cayley graphs let us start with some groups we already have some experience with, namely $G_n = \mathbb{Z}_2^n$. Unfortunately it will not be possible to obtain expander families out of these groups. Nevertheless, they will serve as a good introduction to Cayley graphs.

Let $S = \{s_1, \dots, s_d\}$ be a subset of $\mathbb{Z}_2^n$. (In $\mathbb{Z}_2^n$ every element is its own inverse, so S is automatically closed under inverse.) Notice that S is a generating set for Z_2^n if and only if the rank of $s_1, \dots, s_d$ viewed as vectors in $\mathbb{Z}_2^n$ is n, which is only possible if $d \geq n$. So it is not even possible to generate $\mathbb{Z}_n$ with a number of elements independent of n, much less make it into an expanding family. But let us anyway try to answer the following question:

> How small can $\lambda = \lambda(\mathrm{Cay}(\mathbb{Z}_2^n, S))$ get among all sets S of size d?

We saw that $\lambda(\mathrm{Cay}(\mathbb{Z}_2^n, S)) = 1$ unless $d \geq n$, so let's see what happens when d becomes larger. Last time we showed that

$$\lambda^t \geq \|\mathbf{p}^t - \mathbf{u}\|$$

for every $t > 0$, where $\mathbf{p}^t$ is the distribution of a random walk after t steps starting from some vertex s. Because of commutativity, after $t = \alpha d$ steps the random walk can reach at most

$$\binom{d}{0} + \binom{d}{1} + \cdots + \binom{d}{t} \leq 2^{dH(\alpha)}$$

vertices. Let us choose $t \leq d/2$ so that $dH(\alpha) = n - 1$. Then at least 2^{n-1} of the vertices have not been reached with t steps and

$$\|\mathbf{p}^t - \mathbf{u}\| \geq \sqrt{2^{n-1} \cdot (0 - 2^{-n})^2} = 2^{-(n+1)/2}.$$

and so

$$\lambda \geq 2^{-(n+1)/2t} = 2^{-(n-1)/2\alpha d - O(1/t)} = 2^{-H(\alpha)/2\alpha - O(1/t)}.$$

Applying the upper bound $H(\alpha) \leq \alpha \log_2(1/\alpha) + O(\alpha)$, we obtain

$$\lambda \geq 2^{-\log_2(1/\alpha)/2 - O(1) - O(1/t)} = \Omega(\sqrt{\alpha}).$$

from where $H(\alpha) = O(\lambda^2/\log(1/\lambda))$, and so $d = \Omega(n/\lambda^2 \log(1/\lambda))$. It turns out this is tight up to the $\Omega(\log(1/\lambda))$ factors by the following lemma:

Lemma 1. $\lambda(\mathrm{Cay}(\mathbb{Z}_2^n, S)) = \max_{a \neq 0} |\mathrm{E}_{s \sim S}[\chi_a(s)]|$.

This equation says that $\lambda(\mathrm{Cay}(\mathbb{Z}_2^n, S)) \leq \delta$ if and only if the uniform distribution over the set S is δ-biased! In Lecture 2 we showed we can achieve $|S| = O(n/\delta^2)$, and we just saw that it is necessary to have $|S| = \Omega(n/\delta^2 \log(1/\delta))$. It is not known if the logarithmic factor is necessary.

Proof. Let A be the normalized $2^n \times 2^n$ adjacency matrix of $\mathrm{Cay}(\mathbb{Z}_2^n, S)$. Then we can write

$$A = \frac{1}{d} \sum_{s \in S} A_s$$

where $A_s(g, h) = 1$ if $h = s + g$ (using additive notation for the operation in Z_2^n, and 0 otherwise).

A very nice property of abelian groups is that all the matrices A_s have the same eigenvectors, and so these must also be the eigenvectors of A. The 2^n eigenvectors of A_s are the character functions χ_a viewed as vectors in $\mathbb{Z}_2^n$:

$$(\chi_a A_s)(h) = \sum_g \chi_a(g) A_s(g, h) = \chi_a(s + h) = \chi_a(s)\chi_a(h)$$

because the only g for which $A_s(g, h)$ is nonzero is $g = s + h$. So χ_a is an eigenvector of A_s with eigenvalue $(-1)^{\langle a,s \rangle}$, and by linearity χ_a is an eigenvector of A with eigenvalue

$$\lambda_a = \frac{1}{d} \sum_{s \in S} (-1)^{\langle a,s \rangle}.$$

This gives us a formula for all 2^n eigenvalues of A. When $a = 0$, $\lambda_0 = 1$, and so

$$\lambda(\mathrm{Cay}(\mathbb{Z}_2^n, S)) = \max_{a \neq 0} \frac{1}{d} \left| \sum_{s \in S} \chi_a(s) \right| = \max_{a \neq 0} |\mathrm{E}_{s \sim S}[\chi_a(s)]|. \qquad \square$$

Getting back to our objective of constructing an expanding family of Cayley graphs over $\mathbb{Z}_2^n$, we see that this is impossible as $\lambda(\mathrm{Cay}(\mathbb{Z}_2^n, S)) = \Omega(\sqrt{d/n})$. The situation is similar over other Abelian groups, and to make this approach work we have to turn to non-Abelian groups.

2 Group actions, Schreier graphs, and Kazhdan constants

The expander family we will give will not be a family of Cayley graphs, but a generalization called Schreier graphs. Let G be a group and A be a set. A *group action* of G on A is a map from $G \times A$ To A such that (1) $(gh)a = g(ha)$ and (2) $1a = a$ for every $a \in A$. From (1) and (2) it follows that if $a = gb$, then $b = g^{-1}a$. We say the group action is *transitive* if for every $a, b \in A$ there exists a $g \in G$ such that $ga = b$.

One kind of example is obtained by taking $A = G$ and the group action is just applying the group operation. More interesting examples, including the ones we will work with, come from geometry: Here A can be the points of a geometric space and G can be a group of transformations that acts on this space.

Let G be a group acting transitively on A and S a set of generators for G closed under inverse. The Schreier graph $\mathrm{Sch}(A, S)$ has vertex set A and edges (a, sa) for every $a \in A$ and $s \in S$. In particular, every Cayley graph is a Schreier graph (with the group acting on itself).

Let $S = \{s_1, \ldots, s_d\}$. To analyze the expansion of $\mathrm{Sch}(A, S)$, we write the adjacency matrix A as $\frac{1}{d}\sum_{s\in S} A_s$, where $A_s(a, b) = 1$ if $a = sb$ and 0 otherwise. As in the Abelian case, we want to use this formula to analyze the eigenvalues of A. Unfortunately, it is no longer true that the matrices A_s share the same basis of eigenvectors so we cannot expect to have a nice formula for the eigenvalues of A.

However we do not need to know the exact eigenvalues of A, but we merely need to argue that all but the highest one are bounded away from 1. We can ensure that $\lambda_n \geq -1 + 1/d$ by including the identity 1 into S (this creates a loop around every vertex), so it remains to upper bound λ_2. Recall that

$$\lambda_2 = \max_{\mathbf{v}\perp\mathbf{u}, \|\mathbf{v}\|=1} \langle \mathbf{v}A, \mathbf{v}\rangle.$$

On the other hand,

$$\langle \mathbf{v}A, \mathbf{v}\rangle = \frac{1}{d}\sum_{s\in S} \langle \mathbf{v}A_s, \mathbf{v}\rangle.$$

Since every A_s is a permutation matrix, we have

$$\langle \mathbf{v}A_s, \mathbf{v}\rangle^2 \leq \|\mathbf{v}A_s\| \cdot \|\mathbf{v}\| \leq \|\mathbf{v}\|^2 = 1$$

from where we get

$$\lambda_2 = \max_{\mathbf{v}\perp\mathbf{u}, \|\mathbf{v}\|=1} \langle \mathbf{v}A, \mathbf{v}\rangle \leq \frac{d - 1 + \kappa}{d} \tag{1}$$

where

$$\kappa = \kappa(\mathrm{Sch}(A, S)) = \max_{\mathbf{v}\perp\mathbf{u}, \|\mathbf{v}\|=1} \min_{s\in S} \langle \mathbf{v}A_s, \mathbf{v}\rangle$$

is called the *Kazhdan constant* of the Schreier graph $\mathrm{Sch}(A, S)$.

Suppose we are given an infinite family of Schreier graphs $\mathrm{Sch}(A_n, S_n)$, where $|S_n| = d$ for all n. By the inequality (1), to show that this family is expanding, it is sufficient to prove that $\kappa \leq 1 - \varepsilon$, that is: For every $\mathbf{v} \perp \mathbf{u}$, $\|\mathbf{v}\| = 1$ there exists some $s \in S$ such that $\langle \mathbf{v}A_s, \mathbf{v}\rangle \leq 1 - \varepsilon$, where $\varepsilon > 0$ is some constant independent of n. In the contrapositive: If $\|\mathbf{v}\| = 1$ and $\langle \mathbf{v}A_s, \mathbf{v}\rangle > 1 - \varepsilon$ for every $s \in S$, then $\mathbf{v} \not\perp \mathbf{u}$ (where $S = S_n, A = A_n$).

3 The Margulis-Gabber-Galil expander family

We now instantiate the analysis from the previous section to a specific family of Schreier graphs. Let n be a prime and consider a group G_n of affine transformations $z \to Az + b$ generated by

$$e_x = (1, 0) \quad e_y = (0, 1) \quad T = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}.$$

Then G_n is generated by the transformations $z \to z + e_x, z \to z + e_y, z \to Tz, z \to Bz$ and their four inverses $z \to z - e_x, z \to z - e_y, z \to T^{-1}z, z \to B^{-1}z$, plus the identity.[1]. Let S_n be the set of these nine transformations.

Theorem 2. *$\{\mathrm{Sch}(\mathbb{Z}_n^2, S_n)\}$ is an expanding graph family.*

[1] This is a bit different in letter but similar in spirit to the Margulis-Gabber-Galil construction

Fix n and let $S = S_n$. To prove this theorem, we assume that $\|\mathbf{v}\| = 1$ and $\langle \mathbf{v}A_s, \mathbf{v}\rangle > 1 - \varepsilon$ for all $s \in S$, where ε is a sufficiently small constant (to be set later) and show that $\mathbf{v} \not\perp \mathbf{u}$. We write

$$\langle \mathbf{v}A_s, \mathbf{v}\rangle = \sum_{z \in \mathbb{Z}_n^2} \mathbf{v}(z)\mathbf{v}(sz).$$

The condition that this expression is at least $1 - \varepsilon$ can be rewritten as

$$\sum_{z \in \mathbb{Z}_n^2} (\mathbf{v}(z) - \mathbf{v}(sz))^2 \leq 2\varepsilon. \tag{2}$$

We can make better sense of this expression via Fourier analysis. First we need to show how to do Fourier analysis in $\mathbb{Z}_n^2$. This is not so different from Fourier analysis over the Boolean cube; we only have to be a bit more careful because when $n > 2$ we have to deal with complex numbers. We also use a different normalization.

Let $\omega = e^{2\pi i/n}$ be a root of unity. The character functions

$$\omega_a(x) = \omega_{a_1,a_2}(x_1, x_2) = \frac{1}{n}\omega^{\langle a,x\rangle}$$

form an orthonormal basis over the vector space of functions from $\mathbb{Z}_n^2$ to $\mathbb{C}$ with respect to the inner product

$$\sum_{x \in \mathbb{Z}_2^n} f(x)\overline{g(x)}$$

where the crossbar denotes complex conjugation. This gives the Fourier expansion

$$f(x) = \sum_{a \in \mathbb{Z}_n^2} \hat{f}(a)\omega_a(x) \quad \text{where} \quad \hat{f}(a) = \sum_{a \in \mathbb{Z}_n^2} f(x)\overline{\omega_a(x)}.$$

By orthogonality we have Parseval's identity

$$\sum_x |f(x)|^2 = \sum_a |\hat{f}(a)|^2.$$

Applying Parseval's identity to (2) we obtain $\sum_a |\hat{\mathbf{v}}(a)|^2 = 1$ and

$$\sum_{a \in \mathbb{Z}_n^2} |\hat{\mathbf{v}}(a) - \hat{\mathbf{v}}_s(a)|^2 \leq 2\varepsilon \quad \text{for every } s \in S \tag{3}$$

where $\hat{\mathbf{v}}_s(a) = \sum_{z \in \mathbb{Z}_n^2} \mathbf{v}(sz)\omega_a(z)$.

To see what this means, let us calculate the Fourier transforms $\hat{\mathbf{v}}_s$ for various s in S. When $s = (z \to z + e_x)$ we have

$$\hat{\mathbf{v}}_s(a) = \sum_{z \in \mathbb{Z}_n^2} \mathbf{v}(z + e_x)\omega_a(z) = \sum_{z \in \mathbb{Z}_n^2} \mathbf{v}(z)\omega_a(z - e_x) = \omega^{-a_1}\hat{\mathbf{v}}(a)$$

so we get

$$\sum_{a \in \mathbb{Z}_n^2} |1 - \omega^{-a_1}|^2 |\hat{\mathbf{v}}(a)|^2 \leq 2\varepsilon.$$

Using the fact that $|1-e^{-i\theta}|^2 = 2(1-\cos\theta) \geq 2(\theta/\pi)^2$ for $\theta \in [-\pi,\pi]$ we have

$$\sum_{a\in\mathbb{Z}_n^2} \frac{8a_1^2}{n^2} \cdot |\hat{\mathbf{v}}(a)|^2 \leq 2\varepsilon.$$

Let us represent the elements of $\mathbb{Z}_n$ by the integers between $-(n-1)/2$ and $(n-1)/2$. Then

$$8\sqrt{\varepsilon} \sum_{a\colon |a_1|>\sqrt[4]{\varepsilon}n} |\hat{\mathbf{v}}(a)|^2 \leq \sum_{a\colon |a_1|>\sqrt[4]{\varepsilon}n} \frac{8a_1^2}{n^2} \cdot |\hat{\mathbf{v}}(a)|^2 \leq 2\varepsilon$$

from where $\sum_{a\colon |a_1|>\sqrt[4]{\varepsilon}n} |\hat{\mathbf{v}}(a)|^2 \leq \sqrt{\varepsilon}/4$. We can apply the same argument to $s = (z \to z + e_y)$ and obtain the same formula for a_2. Putting the two together we have

$$\sum_{a\colon |a_1| \text{ or } |a_2|>\sqrt[4]{\varepsilon}n} |\hat{\mathbf{v}}(a)|^2 = O(\sqrt{\varepsilon}). \tag{4}$$

Let us now look at the Fourier transform of $s = (z \to Tz)$:

$$\hat{\mathbf{v}}_s(a) = \sum_{z\in\mathbb{Z}_n^2} \mathbf{v}(Tz)\omega_a(z) = \sum_{z\in\mathbb{Z}_n^2} \mathbf{v}(z)\omega_a(T^{-1}z) = \sum_{z\in\mathbb{Z}_n^2} \mathbf{v}(z)\omega_{B^{-1}a}(z) = \hat{\mathbf{v}}(B^{-1}a).$$

By the same argument we also obtain the formulas

$$\hat{\mathbf{v}}_{z\to T^{-1}z}(a) = \hat{\mathbf{v}}(Ba) \quad \text{and} \quad \hat{\mathbf{v}}_{z\to B^{-1}z}(a) = \hat{\mathbf{v}}(Ta).$$

To use these formulas, we apply the Cauchy-Schwarz inequality to (3) to get

$$\begin{aligned} \sum_{a\in\mathbb{Z}_n^2} \left||\hat{\mathbf{v}}(a)|^2 - |\hat{\mathbf{v}}_s(a)|^2\right| &\leq \sqrt{\sum(|\hat{\mathbf{v}}(a)| + |\hat{\mathbf{v}}_s(a)|)^2} \cdot \sqrt{\sum(|\hat{\mathbf{v}}(a)| - |\hat{\mathbf{v}}_s(a)|)^2} \\ &\leq \sqrt{\sum(2|\hat{\mathbf{v}}(a)|^2 + 2|\hat{\mathbf{v}}_s(a)|^2)} \cdot \sqrt{\sum|\hat{\mathbf{v}}(a) - \hat{\mathbf{v}}_s(a)|^2} \leq 2 \cdot \sqrt{2\varepsilon} = \sqrt{8\varepsilon} \quad \text{for every } s \in S. \end{aligned}$$

from where

$$\sum_{a\in\mathbb{Z}_n^2} \left||\hat{\mathbf{v}}(a)|^2 - |\hat{\mathbf{v}}(Ta)|^2\right| \leq \sqrt{8\varepsilon} \quad \text{and} \quad \sum_{a\in\mathbb{Z}_n^2} \left||\hat{\mathbf{v}}(a)|^2 - |\hat{\mathbf{v}}(Ba)|^2\right| \leq \sqrt{8\varepsilon}. \tag{5}$$

We now want to use (4) and (5) to argue that for ε small enough $\hat{\mathbf{v}}(0,0) \neq 0$, which is the same as saying $\mathbf{v} \not\perp \mathbf{u}$. Let $\mathbf{p}(a) = |\hat{\mathbf{v}}(a)|^2$. Then $\mathbf{p}$ is a probability distribution over $\{-(n-1)/2,\dots,(n-1)/2\}^2$. We will write $\mathbf{p}(A) = \sum_{a\in A}\mathbf{p}(a)$ for the probability of an event A. Then we can interpret (4) as saying that $\mathbf{p}(R) = 1 - O(\sqrt{\varepsilon})$, where

$$R = \{a\colon |a_1|, |a_2| \leq \sqrt[4]{\varepsilon}n\}$$

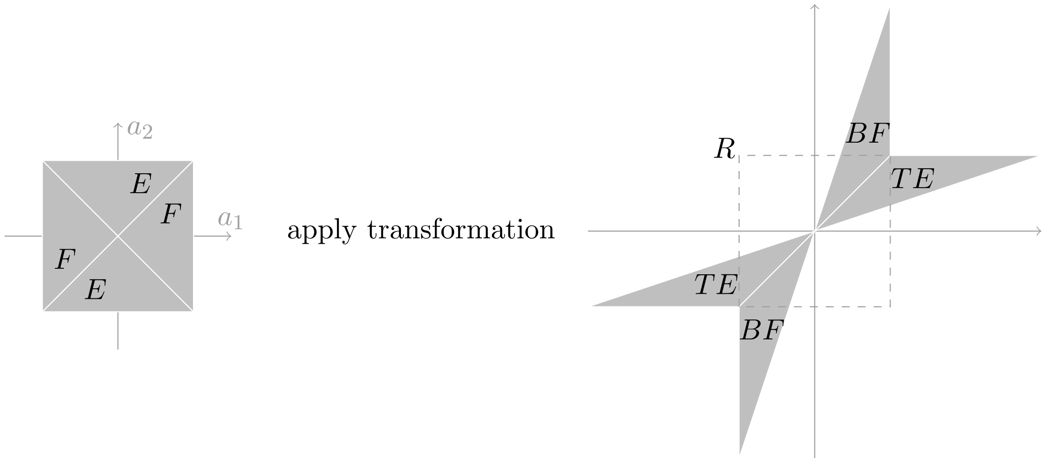
We can also interpret the inequalities (5) as upper bounds on statistical distances between two distributions. For example the first one says that the statistical distance between the distributions $\mathbf{p}$ and $\mathbf{p}_T$ which assigns to a the probability $\mathbf{p}(Ta)$ is $O(\sqrt{\varepsilon})$. This means for every A,

$$|\mathbf{p}(A) - \mathbf{p}(TA)| = O(\sqrt{\varepsilon}) \quad \text{and} \quad |\mathbf{p}(A) - \mathbf{p}(BA)| = O(\sqrt{\varepsilon}). \tag{6}$$

where TA is the set obtained by applying T to all points in A and BA is defined similarly.

Now let E be the event "$|a_1| < |a_2| \leq \sqrt[4]{\varepsilon}n$" and F be the event "$|a_2| < |a_1| \leq \sqrt[4]{\varepsilon}n$.[2]

The following picture illustrates what happens when we apply T to E and B to F. The dashed square indicates the region R. Inside this region, TE falls completely inside F and BF falls completely inside E.



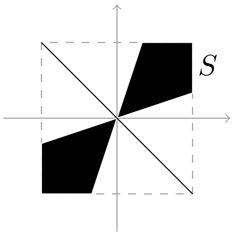
By the above facts we have that

$$\mathbf{p}(E) \leq \mathbf{p}(TE) + O(\sqrt{\varepsilon}) \leq \mathbf{p}(TE \cap R) + O(\sqrt{\varepsilon}) \leq \mathbf{p}(F) + O(\sqrt{\varepsilon}).$$

Now

$$\mathbf{p}(F) \leq \mathbf{p}(BF) + O(\sqrt{\varepsilon}) \leq \mathbf{p}(BF \cap R) + O(\sqrt{\varepsilon}).$$

Since $BF \cap R$ is contained in E, we get that $\mathbf{p}(E - BF \cap R) = O(\sqrt{\varepsilon})$. Analogously $\mathbf{p}(F - TE \cap R) = O(\sqrt{\varepsilon})$. It follows that $\mathbf{p}(S) = 1 - O(\sqrt{\varepsilon})$, where S is the set depicted below:



Now you can check that $T^{10}S \cap R$ intersects S only at $(0,0)$. Therefore

$$\mathbf{p}(S) \leq \mathbf{p}(TS \cap R) + O(\sqrt{\varepsilon}) \leq \cdots \leq \mathbf{p}(T^{10}S \cap R) + O(\sqrt{\varepsilon}) \leq \mathbf{p}(0,0) + O(\sqrt{\varepsilon})$$

and so $\mathbf{p}(0,0) = 1 - O(\sqrt{\varepsilon}) > 0$ provided ε is chosen sufficiently small.

[2] I do not really understand the reasons why things work out the way they do, but presumably one can come up with an argument along these lines, which I found in Terence Tao's lecture notes on expanders, by playing around with the geometry of the linear maps T and B.