CS395T Pseudorandomness and Combinatorial Constructions
Lectures 21 and 22, April 3rd and 5th, 2001
Lecturer: Prof. David Zuckerman, Scribe: Jean-Philippe Martin

1 Introduction

This lecture focuses on weak random sources and extractors, and partially follows the slides in the presentation titled "Extractors for Weak Random Sources and their Applications". Illustrations from the presentation have been copied into this document to make it easier to understand.

2 Weak Random Sources

One approach to pseudo-randomness is to consider a task which uses n truly random bits and determine how we can replace those bits with n bits that are "less random" (e.g. pairwise independent) to provide results that are sufficiently close to the original. We then generate these n bits from m truly random bits ($m < n$).

Another approach is to start with n bits which are not truly random (but we do not necessarily know exactly how much they differ from random bits) and figure out ways to use these bits. This is what we will concentrate on for this section.

A source X is a random variable.

There are several different models of what the probability distribution of sources could look like, for example distributions where some bits are fixed, or the new bits depend on the value of the previous bits, or the bits are biased.

2.1 Min-Entropy

This model is very general and subsume essentially all other models, including the Independent Identically Distributed model which we will discuss in the next section.

Definition 1 *The* min-entropy *of a source X is written $H_\infty(X)$. It is defined as follows.*

$$H_\infty(X) \geq k \Leftrightarrow (\forall x) \Pr[X = x] \leq 2^{-k}$$

Equivalently, $H_\infty(X) = \min_x (-\log_2(\Pr[X = x]))$.

An intuitive view of $H_\infty(X)$ is that k is the number of "truly random" bits contained in X. For example suppose that the strings generated by a source Y contain some random bits and some fixed bits, and $H_\infty(Y) = k$. In this case, we know that the strings generated by Y must contain k random bits.

2.2 Independent Identically Distributed

This simple model was proposed by Von Neumann.

Definition 2 *The source X is* independent identically distributed *(iid) iff there is some probability p such that for all the bits X_i generated by X*

$$\Pr[X_i = 1] = p$$

In other words, if X is an iid source then each bit it generates has probability p (not necessarily $1/2$) of being 1.

These bits can be turned into unbiased random bits by grouping them in pairs, discarding the 00 and 11 and outputting "1" for 10 and "0" for 01.

00	drop
01	0
10	1
11	drop

This works because $\Pr[01] = \Pr[10] = p(1-p)$. The resulting string will be shorter (by an amount which depends on p) but the bits will be unbiased and truly random.

Can we do this for other, harder models? It turns out that this is not the case.

Claim: we cannot deterministically extract even one bit of randomness from some source X with $H_\infty(X) \leq n-1$.

Proof: (by contradiction)

Suppose that there exists a function $f : \{0,1\}^n \mapsto \{0,1\}$, which extracts one bit of randomness out of X. That is, $f(X)$ is close to uniform for any source X with $H_\infty(X) \leq n-1$.

Without loss of generality, suppose that $|f^{-1}(1)| \geq |f^{-1}(0)|$. If X were uniformly distributed inside of $f^{-1}(1)$ then f would not be random anymore. However, such an X would still have $H_\infty(X) \geq n-1$ and such X could very well have this configuration. Therefore f cannot exist and we cannot extract even one bit of randomness. $\square$

Even though no randomness can be extracted from this source X, if we use other random bits Y (the catalyst) we can output new random bits, which are independent of Y.

2.3 Hash Construction

The worst case (i.e. furthest from uniform) for a source X with $H_\infty(X) = k$ is to have a few strings with probability 2^{-k} and all others with zero probability. Let us call the set of strings A. Note that $|A| = 2^k$ and $A \subseteq \{0,1\}^n$.

Using the Leftover Hash Lemma, we can use $O(n)$ new random bits to choose a random hash function and generate m bits which are close to uniform.

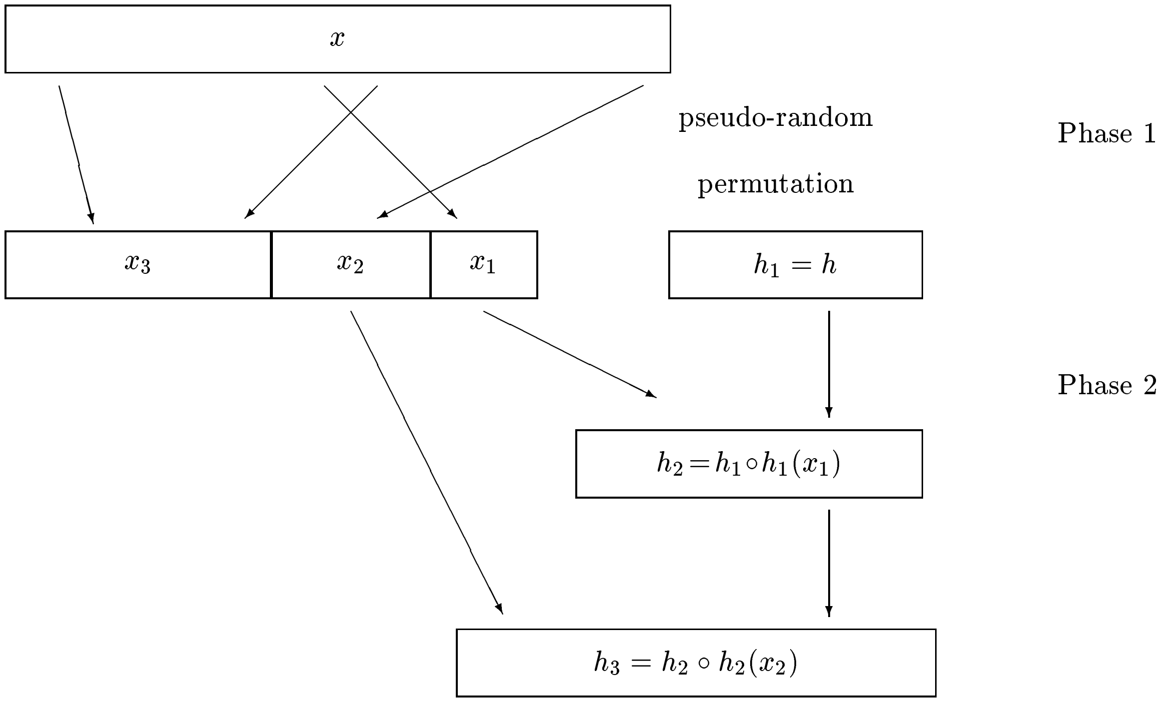
2.4 Extractor Construction

We can use extractors to extract randomness from the weak source using only $O(\log(n))$ bits of randomness. This result is optimal.

We will outline a simpler construction, which uses a catalyst of $O(\log^3(n))$ random bits.

This construction uses two phases. The first is a pseudo-random permutation of the bits in X. Each block is $\log(n)$-wise independent: it costs us $\log^2(n)$ bits for each block.

In the second phase, a hash function h_1 is chosen at random. From this, random bits are built : $h_2 = h_1 \circ h_1(x_1)$. These random bits are then mapped to the corresponding hash function, and we compute $h_3 = h_2 \circ h_2(x_2)$, and so on. There are $\log(n)$ hash functions.



3 Extractors

Best Known Extractors

error $\epsilon = 1/n^{\Theta(1)}$, $\alpha > 0$ arbitrary

due to	min-entropy k	d random bits	output m
goal	any k	$O(\log n)$	k
[Z]	$\Omega(n)$	$O(\log n)$	$(1-\alpha)k$
[TUZ]	$n^{O(1/(\log\log n)^2)}$	$O(\log n)$	$(1-\alpha)k$
[RSW]	any k	$O(\log n)$	$k/\log n$
[RSW]	any k	$O(\log n(\log\log n)^2)$	$(1-\alpha)k$
+[WZ]	any k	$O(\log^2 n(\log\log n)^2)$	k

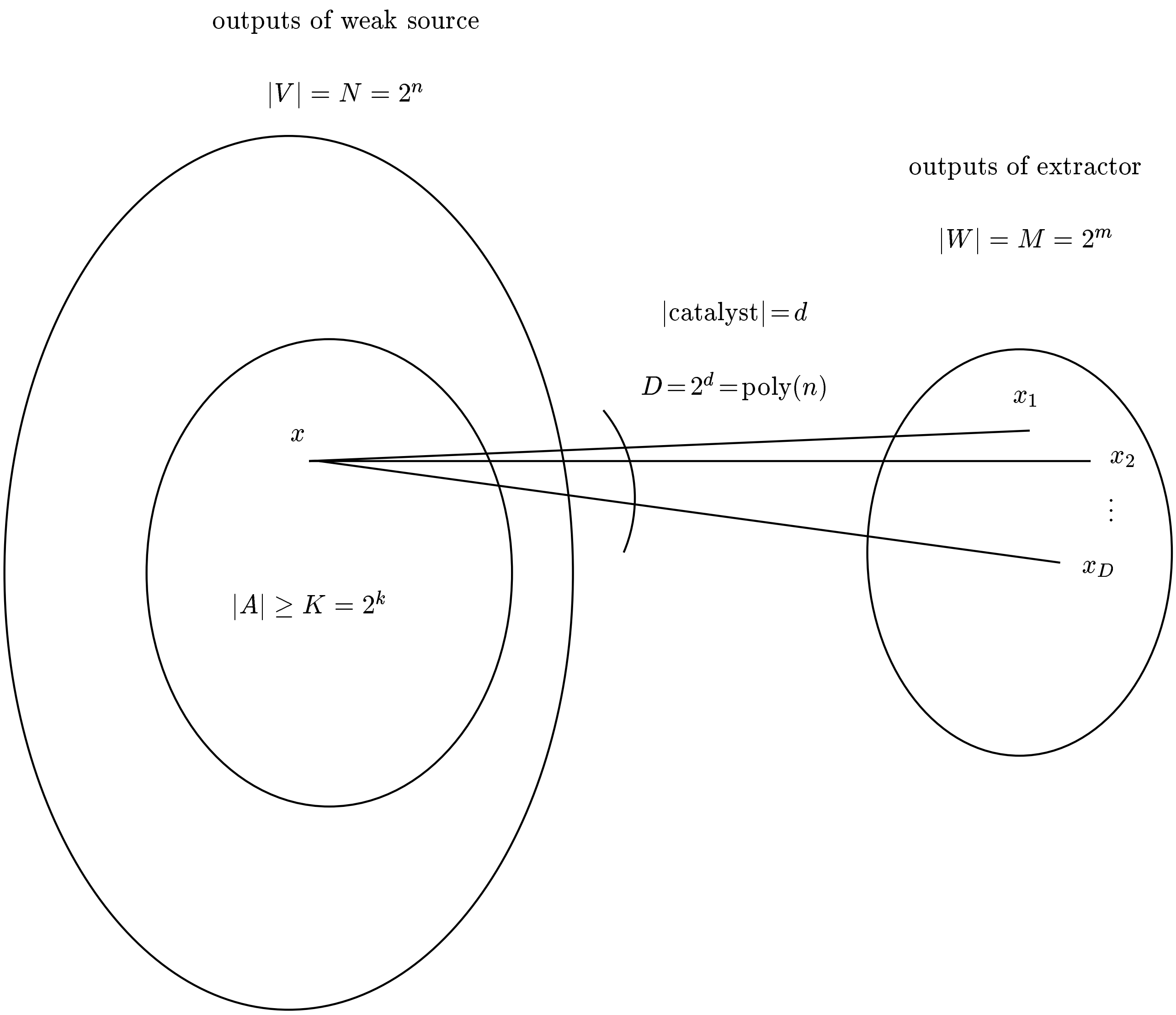
Best Known Dispersers

due to	min-ent. k	d random bits	output m
goal	any k	$O(\log n)$	k
[TUZ]	any k	$O(\log n)$	$k - O(\log n)$

3.1 Applications of Extractors

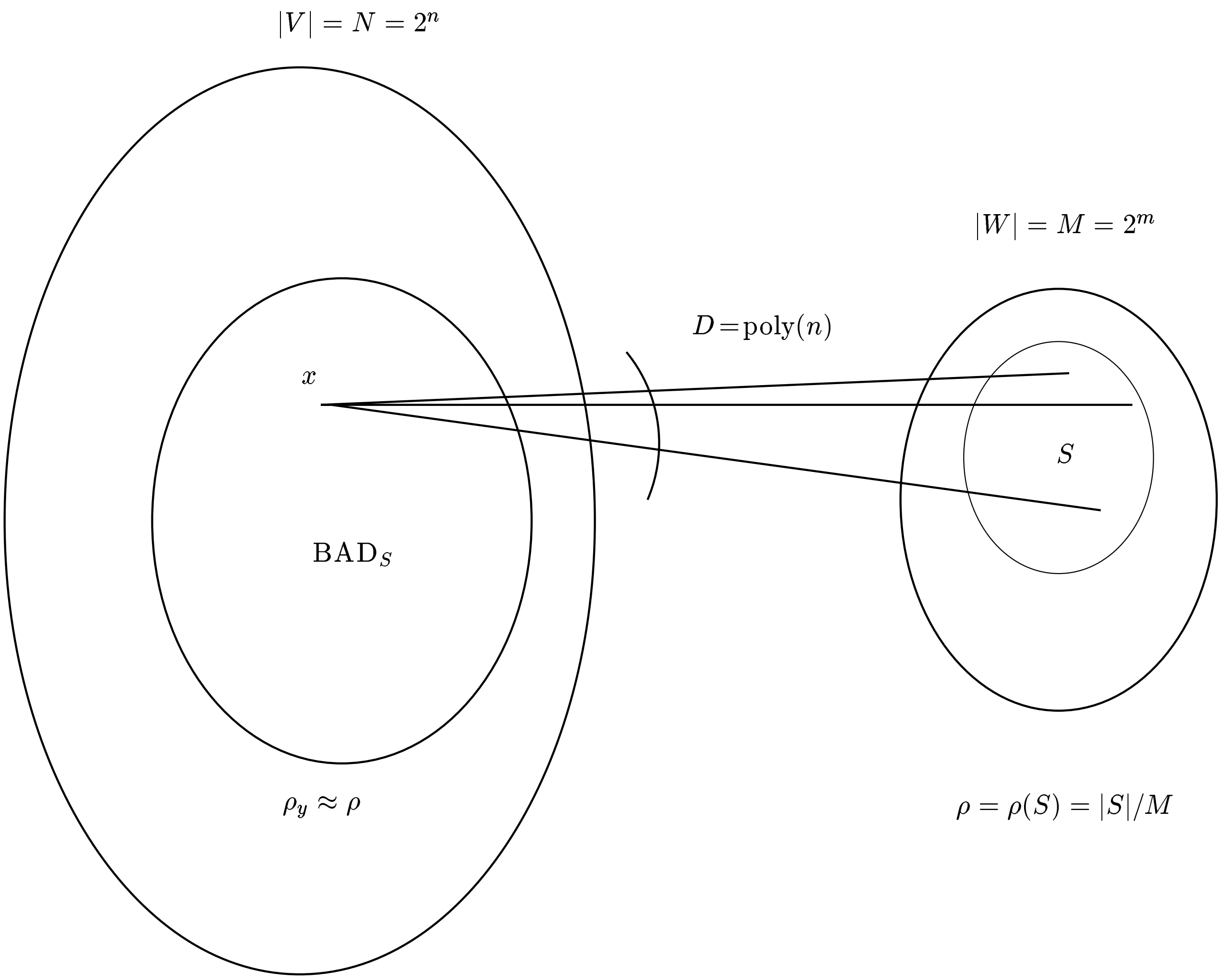
- Simulating randomized algorithms using weak random sources
- Random sampling using few random bits
- Expanders that beat the eigenvalue bound
- Error-correcting codes with good list decoding properties

3.2 Graph-Theoretic View of Extractors (Definitional)



In this graph, $\Gamma(x) = \{h_1(x), \ldots, h_D(x)\} = \{x_1, \ldots, x_D\}$.
Extractor property: $x \in_U A, x_i \in_U \Gamma(x) \Rightarrow x_i$ ϵ-uniform on W

3.3 Alternate View



$$\begin{aligned} \text{BAD}_S^+ &= \{x \in V : \rho_x > \rho + \epsilon\} \\ \text{BAD}_S^- &= \{x \in V : \rho_x < \rho - \epsilon\} \end{aligned}$$

$$\text{BAD}_S = \text{BAD}_S^+ \cup \text{BAD}_S^-$$

Alternate view: For all $S \subseteq W$, $|\text{BAD}_S| \leq K'$.

We shall now prove that these two views are equivalent.

Claim: The definitional view implies the alternate view.

Proof: Set $K' = 2K$. It is sufficient to show that if the extractor property holds then for any S, $|\text{BAD}_S| < 2K = K'$. In fact we will only prove that $|\text{BAD}_S^+| < K$. The proof that $|\text{BAD}_S^-| < K$ is similar.

Suppose that for some S, $A = |\mathrm{BAD}_S^+| \geq K$. Because this set is sufficiently large we can apply the extractor property to A and deduce that the image of any point in A is ϵ-uniform. On the other hand, the definition of $|\mathrm{BAD}_S^+|$ allows us to compute ρ_A, the fraction of neighbors of A in S:

$$\rho_A = \frac{1}{|A|} \sum_{x \in A} \rho_x$$

We know that each $x \in A$ has $\rho_x > \rho + \epsilon$. It follows that $\rho_A > \rho + \epsilon$. Since A has too many neighbors in S, their distribution cannot be ϵ-uniform. From this contradiction we conclude that $\forall S : |\mathrm{BAD}_S^+| < K$.

□

Claim: The graph-theoretic view is equivalent to the alternate view.

Proof: It suffices to show that a graph satisfying the alternate view is an extractor.

The idea is to choose A of size sufficiently larger than BAD_S so that the distribution on W is close to uniform.

Suppose we have a graph $G = (V, E)$ which satisfies
$\forall S : BAD_S = \{x \in V : |\rho_x - \rho| > \epsilon'\}$ has size $\leq K'$. We claim that this can be used to build a (k, ϵ)-extractor where $k = \log_2 K, K = \frac{K'}{\epsilon'}$, and $\epsilon = 2\epsilon'$.

Consider any $S \subseteq W$ (W is the right hand graph) and any $A \subseteq V, |A| \geq K$.

It suffices to show $|\rho_A - \rho| \leq \epsilon$, because this implies that the x_i are ϵ-uniform on W.

$$|\rho_A - \rho| \leq \frac{1}{|A|} \sum_{x \in A} |\rho_x - \rho|$$

$$= \frac{1}{|A|} \sum_{x \in BAD \cap A} \overbrace{|\rho_x - \rho|}^{\leq 1} + \frac{1}{|A|} \sum_{x \in \overline{BAD} \cap A} \overbrace{|\rho_x - \rho|}^{\leq \epsilon'}$$

$$\leq \frac{|BAD_S|}{|A|} + \epsilon' \leq 2\epsilon' = \epsilon$$

□

Suppose now that we have a weak random source and want to use it for a randomized algorithm. What if we don't have any random bits to use as catalyst? We can run the algorithm over all values in S (i.e. cycle over all the possible values for the truly random bits) — there is "only" a polynomial number of them.

3.4 Amplification of BPP

An example of such an application is the amplification of a BPP algorithm with a success probability $\geq 2/3$ using no random bits. Suppose that the source provides n weakly random bits and we need m.

The extractor we will use for the amplification is a $(k', \epsilon = 1/6)$-extractor $E : \{0,1\}^n \times \{0,1\}^d \mapsto \{0,1\}^m$ (d is the size of the catalyst, i.e. E has degree 2^d).

Fix the real input range to A. Let $W_a = \{\text{random strings giving correct answer on } a\}$. We are given a weak source X with $H_\infty(X) \geq k$.

Now run the BPP algorithm using $E(X, y)$ as the random bits for all $y \in \{0,1\}^d$. Take the majority vote as the answer.

If $|\rho_x - \rho| < \frac{1}{6}$ then we get the correct answer at x. Pr[error]$\leq |BAD_{W_a}| \cdot 2^{-k}$ since no element x has probability $> 2^{-k}$.

$|BAD_{w_a}| < 2^{k'+1}$ because, by property of the expander, $|BAD_{w_a}| < 2K'$. It follows from this that Pr[error]$< 2^{k'+1-k}$.

3.5 Random Sampling

The goal here is to guess the density ρ of a set $S \subseteq \{0,1\}^m$, $\rho = |S|/2^m$ using a black box $\boxed{x \in S?}$.

This is done by choosing several points $x_1, ..., x_D$ and sampling the black box at these points (i.e. estimating whether $x_i \in S$). These results are then combined to form the estimate $\hat{\rho}$.

The most immediate approach is to use the "oblivious" estimator, $\hat{\rho} = |\{x_1, ..., x_D\} \cap S|/D$. This is a good estimate but there are others as well.

Phrased more formally, the goal of this problem is to find an estimate which is accurate with high probability. $\hat{\rho}$ must satisfy: $\Pr[|\hat{\rho} - \rho| \geq \epsilon] \leq \gamma$.

Think of $\epsilon = 1/\text{poly}(m)$, $\gamma = 2^{-m/200}$.

At least m random bits are necessary. It turns out that one can achieve the required estimate using $1.01m$ random bits!

Pick $x \in_U \{0,1\}^n$, $\Gamma(x) = \{x_1, \ldots, x_D\}$, $\hat{\rho} = \rho_x$

$\Pr_x[|\rho_x - \rho| \geq \epsilon] \leq |BAD_S| \cdot 2^{-n}$

3.6 Extractors are dispersers

In this section we show the relationship between extractor graphs and dispersers. Recall that an (N, M, D, K, ϵ)-disperser is a bipartite graph with N nodes on the left side, each with degree D, and M nodes on the right side (W), such that every subset of K nodes on the left side is connected to more than $1 - \epsilon$ fraction of the nodes on the right.

Definition 3 *The* neighbors of a set *A, denoted $\Gamma(A)$, are defined as all the nodes in A and all the nodes which have a neighbor in A.*

Claim: Any (K, ϵ)-extractor is also a (K, ϵ)-disperser

Proof: Suppose that we have a (K, ϵ)-extractor E. We know that for any A ($|A| \geq K$), choosing a random $x \in_U A$ and then a neighbor z of x, z will be within ϵ of uniform on W. Let us use $\Gamma(A)$ as our test set. Since $E(x, y) \in \Gamma(A)$ with probability 1,

$$\left| 1 - \frac{|\Gamma(A)|}{|W|} \right| \leq \epsilon$$

Hence $1 - \epsilon \leq \frac{|\Gamma(A)|}{|W|}$, which concludes our proof of $|\Gamma(A)| \geq (1 - \epsilon) \cdot |W|$. $\square$

3.7 Highly Expanding Graphs

Definition 4 *A graph $G = (V, E)$ is said to be* K-expanding *for some value K if any two subsets $A, B \subseteq V$ of size at least K are linked by an edge in G.*

This condition is equivalent to the condition that if $|A| \geq K$, then $|\Gamma(A)| > N - K$, where $N = |V|$. For if $|\Gamma(A)| \leq N - K$, then the K vertices not in $\Gamma(A)$ would constitute a set of size K that did not share an edge with A. The other direction is similar.

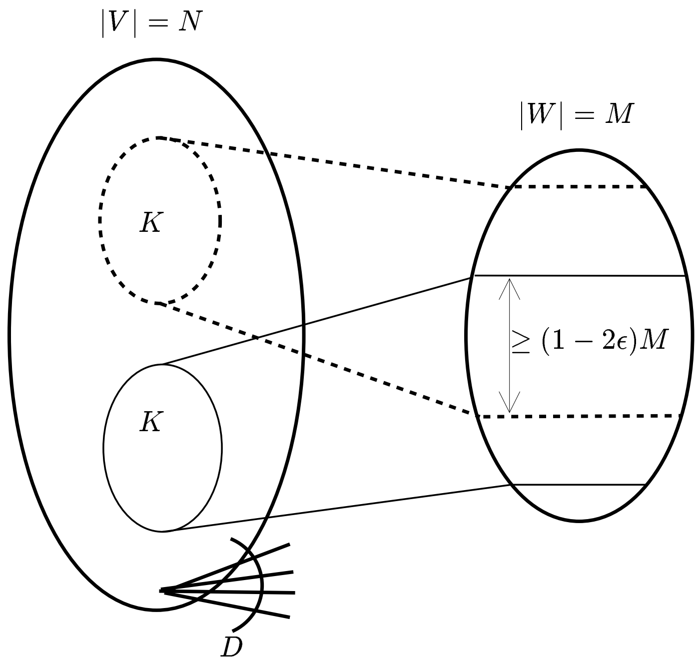
Think of $K = \sqrt{N}$. How small can we make the average degree of these highly expanding graphs? From the equivalent expansion condition, the average degree $d_{avg} > (N - K)/K \approx N/K$.

Using the second eigenvalue method would yield a graph with average degree $d_{avg} \approx (N/K)^2$.

Using extractors, highly expanding graphs can be constructed which achieve $d_{avg} = N^{1+o(1)}/K$.

3.7.1 Constructing a highly expanding graph from an extractor

We show how to construct a highly expanding graph from a $(k, \epsilon < \frac{1}{4})$-extractor.



The diagram above shows a (k, ϵ)-extractor with degree D. Since extractors are also dispersers, each set of size K on the left has at least $(1 - \epsilon) \cdot M$ neighbors on the right. The intersection of the sets on the right has size at least $(1 - 2\epsilon)M$.

To build a highly connected graph from this extractor, we will add an edge between two vertices a and b in V (the left hand side) if they share a neighbor. We know that the vertices on the left (originally) have degree D. The vertices on W (the right side) have a small average degree $\bar{D}$ because $ND = M\bar{D}$. However, we have no guarantee on the individual vertices. It could be the case that a vertex on the right is a neighbor with every single vertex on the left side!

We will remove all vertices in W with degree $> 2\bar{D}$. This removes at most half of the vertices in W. The neighbor sets now have size $\geq \frac{1}{2} - \epsilon$. If we choose $\epsilon < \frac{1}{4}$ then we are still guaranteed to create a highly expanding graph.

To understand why eigenvalues do poorly for this problem, suppose an r-regular graph on N vertices. $E(S,T)$=number of edges between S and T. We have seen before that

$$\left| E(S,T) - \frac{r}{N}|S||T| \right| \leq \lambda_2 \sqrt{|S||T|}$$

If we substitute $|S| = |T| = K$, and using $\lambda_2 \geq \sqrt{r}$ (which is true up to a small constant),we have

$$\left| E(S,T) - \frac{r}{N}K^2 \right| \leq \sqrt{r} \cdot K$$

$$E(S,T) > 0 \Rightarrow \sqrt{r} \cdot K < \frac{r}{N}K^2 \Rightarrow 1 < \frac{\sqrt{r}}{N}K \Rightarrow$$

$$r > \left(\frac{N}{K}\right)^2$$