COM S 485
Prof. John Hopcroft
Lecture 26 3/29/2006
Scribes: Byron Roberts & Brendan Rehon

Derandomization

When generating random numbers, one of the main difficulties is getting the random bits. In addition, there are certain situations in which one would not want to use a pseudorandom generator (e.g. encryption). Here, we will see how to minimize the number of random bits required for a random algorithm.

Reference: Randomized Algorithms by Motwani and Raghavan

Consider the following game: We have Bob and Alice, who will choose a total of n integers. Alice goes first: she selects $(1-\varepsilon)n$ integers. Bob goes next: he selects d integers ($d < \varepsilon n$). In order for Bob to win, he must select at least one of the same integers that Alice does.

n= 1000 ε= 1/10 So Alice chooses 900 integers.

If d=6, Bob will lose with probability $(1/10)^6$

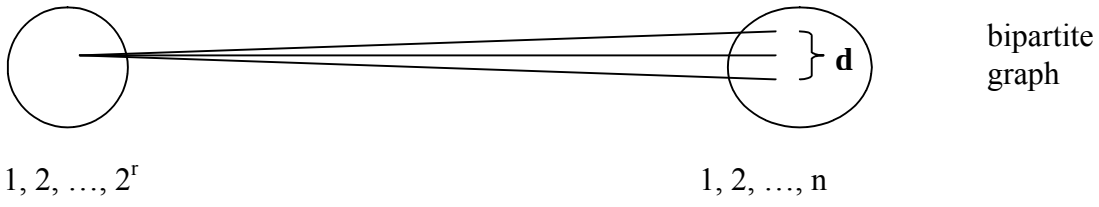
Suppose Bob uses a deterministic algorithm to pick his integers. If Alice discovers the algorithm, she can always win. So if Bob is clever, he will use a random algorithm.

To pick an integer between 1 and n requires log(n) bits. So to pick d integers requires dlog(n) bits.

There are ways that Bob can reduce the number of bits that he needs, such as using an expander.

Suppose Bob needs r bits:

Create map ϕ: $2^r \rightarrow$ subsets of {1, 2, …,n} of size less than or equal to d.



Generate r random bits, then "follow the edges" to get d random integers

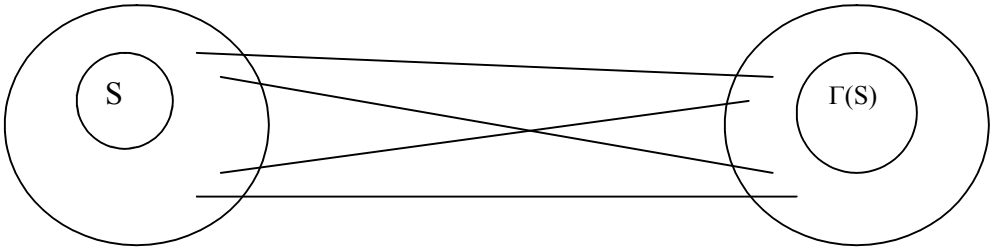
Bob chooses r bits, uniformly at random, and uses ϕ to give him a subset of d integers.

Let p be the probability that none of Bob's d integers lies in Alice's (1-ε)n integers (p is the probability that Bob loses.)

Task: construct ϕ so that $p \leq \varepsilon^d$.
*constructing the mapping ϕ is equivalent to constructing a bipartite graph.

Suppose there was a bipartite graph with vertex sets U and V such that for all $S \leq U$ such that $|S| > p|U|$

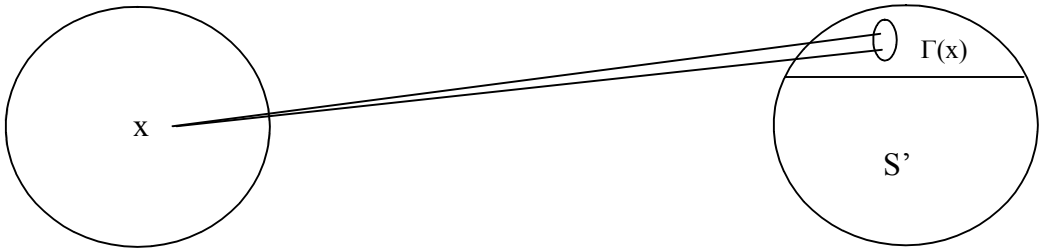
Set of neighbors $\Gamma(S) \leq V$ has size $|\Gamma(S)| \geq \varepsilon|V|$:



U
$|S| \geq p|U|$

V
$|\Gamma(S)| \geq \varepsilon|V|$

Then, for any subset $S' \leq V$ of size at least $(1-\varepsilon)|V|$ and any x in U the probability that $\Gamma(x)$ lies outside of S' is $\leq p$.



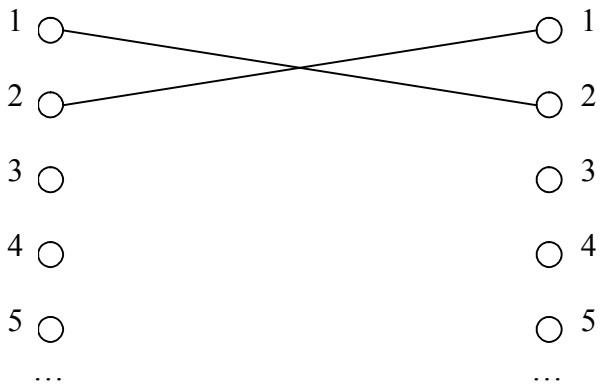
U V

Now, will show that $\Gamma(x)$ outside of $S' > p \rightarrow$ leads to contradiction.

Suppose probability that $\Gamma(x)$ outside of S' exceeds p. Let S be subset of vertices U that do not have neighbors in S'
Since $|S| \geq p|U|$, the set of neighbors of S is greater than $\varepsilon|V|$ elements.

Therefore, contradiction. ■

Given an expander with n vertices, we can construct a bipartite graph:



Note: the presented method for finding random bits is not necessarily the best. For instance, consider performing a random walk on a graph. After log(n) steps, we are at a random node between 1 and n. That initial random walk doesn't buy us anything. However, now starting from a random node, any subsequent random walk will have more randomness than what you would expect.

Constructing expanders—hard.

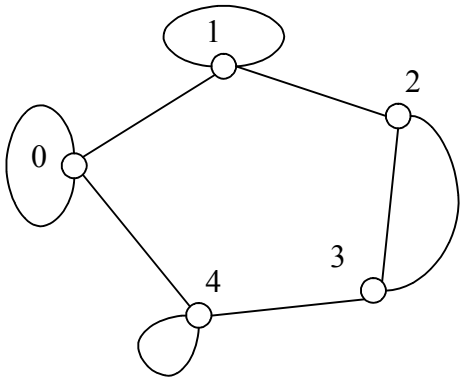
How do we construct expanders?

1) Random graph (via G(n,p), maybe using a value p = d/n)
2) Zig-Zag construction.
   For more information, Google the phrase "zig zag expanders".
3) Margulis construction.
   For more information, Google the phrase "expander graph".

However, there's a problem with these three constructions: they have steep memory requirements because the constructions require storing a random graph.

4) Given (0, 1, 2, … , p-1) where p is prime, we form an expander graph by constructing the following edges:
   a. (i, i+1) mod p
   b. (i, i−1) mod p
   c. $i \rightarrow i^{-1} \bmod p$ where we define $0^{-1} = 0$.

   Below is an example with p = 5—but note it's more effective for large p:



Google the phrase “expander graph prime” for more information.

Eigenvalues of Random Matrices

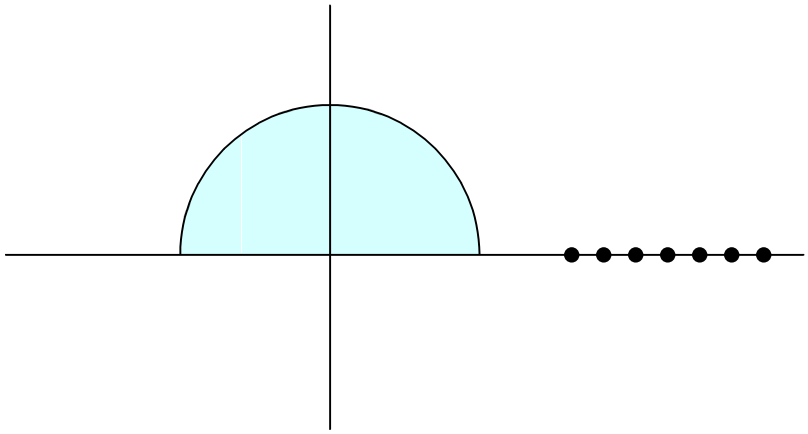
Let A = (a_{ij}), where A is also symmetric.

Also let E[a_{ij}] = 0, elements have finite variance.

(Note that in the previous homework [homework 7 problem 2], the condition E[a_{ij}] = 0 was violated, causing there to be one enormous eigenvalue that represented the expected value sum).

How do we know if the eigenvalues are significant or just variance?

Wigner discovered the eigenvalues of matrices like A have a semicircular probability distribution. Fit the eigenvalues onto the Wigner distribution’s range by dividing the eigenvalues by $n^{1/2}\sigma$. Any scaled eigenvalues outside of the Wigner semicircular distribution are significant.



Suppose we find that in a 10000 dimension data set, that 7 eigenvalues are outside the Wigner distribution. Then we can reduce the problem to its significant dimensions; we project the 10000 dimension data to a 7 dimensional vector.