CS 6850	**Random Graphs are Expanders**
Jon Kleinberg	

We begin by recalling the definition of *expansion.* We use $|S|$ to denote the size of a set of nodes S; we use $\overline{S}$ to denote the complement of a set of nodes S; and we use $e^{out}(S)$ to denote the set of edges with exactly one end in S. We define the *surface-to-volume ratio* of a set S to be

$$\sigma(S) = \frac{|e^{out}(S)|}{\min(|S|, |\overline{S}|)}.$$

The expansion $\alpha(G)$ of a graph G is then defined as the minimum surface-to-volume ratio of any set of nodes S:

$$\alpha(G) = \min_{S \subseteq V} \sigma(S).$$

Since $\sigma(S) = \sigma(\overline{S})$ by definition, it is enough to take this minimum only over the sets S of size at most $n/2$, in which case $|S| \leq |\overline{S}|$:

$$\alpha(G) = \min_{S:|S| \leq n/2} \sigma(S).$$

The most basic non-trivial fact about expander graphs is that they exist at all: there exist fixed constant values of d and α so that for arbitrarily large values of n, there are n-node graphs with maximum degree at most d and expansion at least α. The key point is that neither of the parameters d or α depend on the size n of the graph. To avoid explicitly discussing the underlying parameters all the time, one often speaks informally of a class of graphs having "good expansion properties" if d and α are absolute constants as n goes to infinity.

Constructing large graphs with good expansion properties — and proving these expansion properties — is much more difficult than one might imagine. Trying this oneself is the best way to drive the point home. For example, a $\sqrt{n} \times \sqrt{n}$ grid graph does not maintain a constant expansion parameter of $\alpha > 0$ as n increases: the set S consisting of the leftmost $\sqrt{n}/2$ columns has

$$|e^{out}(S)|/|S| \leq \sqrt{n}/(n/2) = 2\sqrt{n}/n = 2/\sqrt{n}.$$

Or consider an n-node complete binary tree: it may look like it has good expansion properties if one views it from the root downward; but if we think of the subtree S below any given node, it has $|e^{out}(S)|/|S| = 1/|S|$. One can show that much more sophisticated examples than these also fail to serve as good expander graphs. Ultimately, finding an explicit construction of arbitrarily large graphs that could be proved to have good expansion properties required intricate analysis and sophisticated use of some deep results from mathematics; it is only now, three decades after people began studying expanders, that somewhat simpler analyses are emerging.

We will now show that a simple random construction produces good expander graphs with constant probability. In light of our discussion, this is quite surprising: it is extremely difficult to verify that an explicitly constructed graph is a good expander, but it is easy to show that a random graph is likely to be one. The analysis of our random construction will be quite crude, and will not aim for the best possible values of all parameters; rather, its goal is to show how a completely direct use of the Union Bound is enough to verify good expansion.

Neighborhood expansion. To make the analysis a bit cleaner, we first introduce a variation on the definition of expansion that will imply our primary definition. First, if $G = (V, E)$ is a graph, and $S \subseteq V$, we use $N(S)$ to denote the "neighbors" of S — the set of nodes with an edge to some node in S. (Note that $N(S)$ may include some nodes in S but not others.) Now, for any constants $c \leq \frac{1}{2}$ and $\beta > 1$, we say that a graph has *neighborhood expansion* with parameters (β, c) if for every subset S of at most cn nodes, we have $|N(S)| \geq \beta|S|$.

Let us first establish that a graph with good neighborhood expansion also has good expansion in the traditional sense.

(1) *Choose any constants $c \leq \frac{1}{2}$ and $\beta > 1$ for which $\beta c > \frac{1}{2}$. If G has neighborhood expansion with parameters (β, c), then it has expansion at least α, where $\alpha = 2\beta c - 1 > 0$.*

Proof. First, suppose $|S| \leq cn$. Then $N(S) - S$ contains at least $\beta|S| - |S| = (\beta - 1)|S|$ nodes. Since each node in $N(S) - S$ must be the endpoint of a distinct edge in $e^{out}(S)$, we have $|e^{out}(S)| \geq (\beta - 1)|S|$ and hence $|e^{out}(S)|/|S| \geq (\beta - 1)$. Since $2c \leq 1$, we have $\beta - 1 \geq 2\beta c - 1$, and hence $|e^{out}(S)|/|S| \geq (2\beta c - 1)$.

Otherwise, suppose $cn < |S| \leq \frac{1}{2}n$. In this case, choose an arbitrary set $S' \subseteq S$ consisting of exactly cn nodes. Then $|N(S')| \geq \beta|S'| = \beta cn$, and hence $N(S') - S'$ contains at least $\beta cn - \frac{1}{2}n = (\beta c - \frac{1}{2})n$ nodes. Again, each of these nodes must be the endpoint of a distinct edge in $e^{out}(S)$. Thus we have $e^{out}(S) \geq (\beta c - \frac{1}{2})n$ while $|S| \leq \frac{1}{2}$, and so $|e^{out}(S)|/|S| \geq (\beta c - \frac{1}{2})/\frac{1}{2} = 2\beta c - 1$. ∎

The random construction. We start with a set V of n nodes, labeled $1, 2, 3, \ldots, n$, and no edges joining any of them. A *random perfect matching* on V is a set of edges M constructed by randomly ordering the nodes of V, say as $v_1, v_2, \ldots, v_n$, and defining M to be the set of $n/2$ edges (v_{2i-1}, v_{2i}) for $i = 1, 2, \ldots, n/2$.

Here is the full construction of G. We set $d = 90$; we compute d random perfect matchings $M_1, M_2, \ldots, M_d$ on the set V, using orders chosen independently for each; and we define the edge set $E = M_1 \cup M_2 \cup \cdots \cup M_d$. Notice that while G has constant node degree — independent of the number of nodes — it is quite a large constant; this is in keeping with our plan to sacrifice better parameters for the sake of the simplest analysis possible. In fact, random graphs in which each node has degree 3 can be shown to have fairly good expansion properties as well, but the proof of this becomes somewhat more involved.

(2) *With probability at least $3/4$, the graph $G = (V, E)$ has neighborhood expansion with parameters $(1/6, 4)$.*

The proof will consist of an extended but completely direct use of the Union Bound, summing over an exponential number of possible bad events that could prevent G from being a good expander. In order to make the calculations work out, we first need some simple bounds on the growth of the factorial function and the binomial coefficients.

(3) *For every natural number n, we have $n! > \left(\frac{n}{e}\right)^n$.*

Proof. We prove this by induction, the cases $n = 0$ and $n = 1$ being clear. For a larger value of n, we can apply the induction hypothesis together with the fact that $\left(1 + \frac{1}{n}\right)^n < e$ for all natural numbers n. Thus we have

$$(n+1)! = (n+1)n! > (n+1)\left(\frac{n}{e}\right)^n > (n+1)\left(\frac{n}{e}\right)^n \frac{\left(1+\frac{1}{n}\right)^n}{e} = \frac{(n+1)^{n+1}}{e^{n+1}}.$$

■

Using this bound, we now prove

(4) *For every pair of natural numbers n and k, where $n \geq k$, we have $\left(\frac{n}{k}\right)^k \leq \binom{n}{k} < \left(\frac{en}{k}\right)^k$.*

Proof. By (3), we have

$$\binom{n}{k} < \frac{n^k}{k!} < \frac{n^k}{(k/e)^k} = \left(\frac{en}{k}\right)^k.$$

Since $\frac{n}{k} \geq \frac{n-1}{k-1}$ for any natural numbers $n \geq k$, we have

$$\binom{n}{k} \geq \frac{n^k}{k^k} = \left(\frac{n}{k}\right)^k.$$

■

Notice that $\binom{n}{k}$ is not defined when k is not a natural number. However, if k is not a natural number, we can still use (4) to bound $\binom{n}{\lfloor k \rfloor}$ as follows:

$$\binom{n}{\lfloor k \rfloor} < \left(\frac{en}{\lfloor k \rfloor}\right)^{\lfloor k \rfloor} < \left(\frac{en}{k}\right)^k,$$

where the first inequality is just (4), and the second follows from the fact that the function $(en/k)^k$ increases monotonically until $k = n$.

We are now ready for

Proof of (2). If G fails to have the desired property, it means that there is some set S of at most $n/6$ nodes so that $N(S) < 4|S|$. So for every set S of at most $n/6$ nodes, and every set T of size exactly $4|S|$, we define the event $\mathcal{E}_{ST}$ that $N(S) \subseteq T$. We observe that if the union

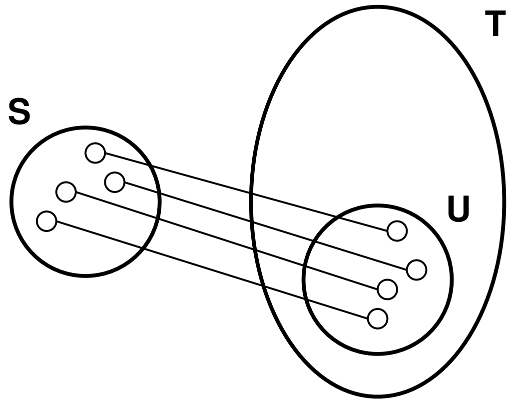

Figure 1: The event $\mathcal{E}_{ST}$ in the analysis of the expander construction.

of all these events $\mathcal{E}_{ST}$ does not occur, then every set S expands by a sufficient amount, and G has the desired neighborhood expansion properties. Thus, it is sufficient to give an upper bound on

$$\Pr\left[\bigcup_{\substack{|S|\leq n/6\\|T|=4|S|}} \mathcal{E}_{ST}\right].$$

To think about this, we first define a related set of events as follows. For every pair of sets S and T with $|T| = 4|S|$, we define the event $\mathcal{E}'_{ST}$ that in a single random perfect matching M, all nodes in S are matched to a node in T.

To bound $\Pr[\mathcal{E}'_{ST}]$, we can imagine constructing the perfect matching M as follows. We define $k = |S|$, and we name the nodes of S as $u_1, u_2, \ldots, u_k$. We first choose a partner for u_1 uniformly at random from the set V. Then (unless u_2 is already matched by this first edge), we choose a partner for u_2 uniformly at random from the remaining unmatched nodes. We continue in this way, always choosing the first node in S that is not yet matched. For at least $k/2$ steps, we will not run out of nodes in S; in each of these steps, there are at least $n-k$ nodes to choose a partner from; and for the process to succeed, we need to choose this partner from the set T of $4k$ nodes. Thus in each step, we succeed in choosing a partner from T with probability at most $4k/(n-k)$; and since $k \leq n/6$, this probability is bounded by

$$\frac{4k}{n-n/6} \leq \frac{4k}{5n/6} = \frac{24k}{5n} = \frac{4.8k}{n}.$$

For the event $\mathcal{E}'_{ST}$ to occur, we must succeed in choosing a partner from T in each of these first $k/2$ steps, and so

$$\Pr[\mathcal{E}'_{ST}] \leq \left(\frac{4.8k}{n}\right)^{(k/2)}.$$

Now, the graph G is built from d random perfect matchings, so if $k = |S| \leq n/6$ and

$|T| = 4k$, then

$$\Pr[\mathcal{E}_{ST}] \;=\; (\Pr[\mathcal{E}'_{ST}])^d \leq \left(\frac{4.8k}{n}\right)^{dk/2}.$$

As promised, we complete the proof with an enormous application of the Union Bound:

$$\Pr\left[\bigcup_{\substack{|S|\leq n/6\\|T|=4|S|}} \mathcal{E}_{ST}\right] \leq \sum_{\substack{|S|\leq n/6\\|T|=4|S|}} \Pr[\mathcal{E}_{ST}].$$

This sum involves exponentially many terms; to unravel it, we consider separately the terms for each possible size of the set S. For sets S of size k, there are $\binom{n}{k}\binom{n}{4k}$ terms, each with probability at most $\left(\frac{4.8k}{n}\right)^{dk/2}$. We then upper-bound the binomial coefficients using (4) and begin canceling as many terms as we can:

$$\begin{aligned}
\sum_{\substack{|S|\leq n/6\\|T|=4|S|}} \Pr[\mathcal{E}_{ST}] \;&\leq\; \sum_{k=1}^{n/6} \binom{n}{k}\binom{n}{4k}\left(\frac{4.8k}{n}\right)^{dk/2} \\
&<\; \sum_{k=1}^{n/6} \left(\frac{en}{k}\right)^{k}\left(\frac{en}{4k}\right)^{4k}\left(\frac{4.8k}{n}\right)^{dk/2} \\
&=\; \sum_{k=1}^{n/6} \left[\frac{e^5\cdot(4.8)^5}{4^4}\left(\frac{4.8k}{n}\right)^{(d/2-5)}\right]^k.
\end{aligned}$$

Now we pause to observe that

$$\frac{e^5\cdot(4.8)^5}{4^4} < 1500;$$

also, since $k \leq n/6$, we have $(4.8k/n) \leq .8$, and with $d = 90$, we have $(.8)^{d/2-5} = (.8)^{40} < 1/(7500)$. Thus we conclude with

$$\begin{aligned}
\sum_{k=1}^{n/6} \left[\frac{e^5\cdot(4.8)^5}{4^4}\left(\frac{4.8k}{n}\right)^{(d/2-5)}\right]^k \;&<\; \sum_{k=1}^{\infty}\left[1500\,(.8)^{40}\right]^k \\
&<\; \sum_{k=1}^{\infty}\left(\frac{1}{5}\right)^k = \frac{1}{4}.
\end{aligned}$$

Thus, with probability at least $3/4$, the bad event does not happen, and the graph G has the desired expansion properties. ■