1 A First Random Graph Model: $G_{n,p}$

In the most heavily-studied model of random graphs, we start with n nodes and join each pair by an undirected edge, independently with probability p. We will call this model $\mathcal{G}_{n,p}$.

Expected degrees. If G is a graph generated using $\mathcal{G}_{n,p}$, we can compute the expected degree of one of its nodes v using linearity of expectation. Let X_v be a random variable denoting the degree of v, and for each other node w, let $X_{v,w}$ be a random variable equal to 1 if there is an edge joining v and w, and equal to 0 otherwise. We have

$$\begin{aligned} X_v &= \sum_w X_{v,w} \\ E[X_v] &= \sum_w E[X_{v,w}] \\ &= \sum_w p = (n-1)p. \end{aligned}$$

So if we think of p as $\frac{c}{n-1}$ for some quantity c, then the expected degree of v is c.

Isolated nodes. Now, if c is a constant, then despite the constant expected degree in G, there will still be many isolated nodes (that is, nodes with no incident edges). To see this, let $\mathcal{E}_v$ denote the event that v is isolated; this requires that each of its $n-1$ potential incident edges not be present, so we have

$$\begin{aligned} \Pr[\mathcal{E}_v] &= (1-p)^{n-1} \\ &= \left(1 - \frac{c}{n-1}\right)^{n-1} \\ &= \left(\left((1 - \frac{c}{n-1}\right)^{\frac{n-1}{c}}\right)^c. \end{aligned}$$

Now, the part inside the outermost parentheses on the last line is between $1/4$ and $1/e$ as $\frac{c}{n-1}$ ranges from $\frac{1}{2}$ down toward 0. Thus, $\Pr[\mathcal{E}_v]$ is between 4^{-c} and e^{-c}, which is a constant when c is constant.

Using this, we can ask how large c needs to be in order for there to be a high probability of no isolated nodes. Let $\mathcal{E}$ be the event that there is any isolated node in G; then by the Union Bound we can write

$$\begin{aligned} \mathcal{E} &= \bigcup_v \mathcal{E}_v \\ \Pr[\mathcal{E}] &\leq \sum_v \Pr[\mathcal{E}_v] \\ &\leq ne^{-c}. \end{aligned}$$

Now we choose c large enough so that e^{-c} is small enough to cancel n. In particular, $c = \ln n$ is not quite enough, but $c = 2 \ln n$ will easily do it:

$$\Pr[\mathcal{E}] \le n e^{-2 \ln n} = n \cdot n^{-2} = n^{-1}.$$

Thus, $\mathcal{G}_{n,p}$ is not an appropriate model for considering random graphs in which all degrees are positive, yet constant — as we've just seen, the average degree in $\mathcal{G}_{n,p}$ needs to become logarithmic before the last isolated node is likely to vanish.

Diameter two. Since we'll soon be considering the question of short paths in graphs, it's interesting to consider one final short calculation in the $\mathcal{G}_{n,p}$ model: determining the value of p at which we're likely to obtain a graph of *diameter two* — that is, a graph in which every pair of nodes is connected by a path of length at most two.

It's useful to ask what we should expect p might be, before trying the calculation. First, if we simply want a graph with very few edges in which every pair of nodes a path of length at most two, we could use the *star graph*: a graph on n nodes $v_1, v_2, \ldots, v_n$, in which the only edges are from v_1 to each other v_i, for $i \ge 2$. Clearly every v_i has a two-step path to every v_j through v_1; and the star graph has only $n - 1$ edges, which is the fewest needed for an n-node graph to even be connected. So if our goal is simply to achieve diameter two with as few edges as possible, the star graph completely answers this question.

However, the star graph contains a node of enormous degree (i.e., degree $n - 1$), and if our goal is instead to achieve diameter two while keeping the maximum degree as small as possible, we end up with a much more subtle question. First note that if the maximum degree is d, then a given node v can reach d other nodes in one step, and at best each of those nodes can reach $d - 1$ other nodes in a second step. So in two steps, v can reach at most $d + d(d - 1) = d^2$ other nodes. In a graph of diameter two, v needs to reach all $n - 1$ other nodes in at most two steps, and so this argument implies that we need $d^2 \ge n - 1$, or $d \ge \sqrt{n-1}$.

A natural question, therefore, is whether there in fact exist graphs of diameter two where d is close to this small. We now show that $G_{n,p}$ comes close to this bound, by showing that a random graph drawn from $G_{n,p}$ will have diameter two with high probability provided that $p \ge \dfrac{c\sqrt{\log n}}{\sqrt{n}}$ for some constant c. Since the expected degree is pn, we can use this to show that $G_{n,p}$ produces graphs of diameter two whose maximum degree is proportional to a constant times $\sqrt{n \log n}$.

The proof is once again an application of the Union Bound. Let $\mathcal{E}_{v,w}$ be the bad event that there is no path of length one or two connecting nodes v and w; so $\mathcal{E} = \cup_{v,w \in V} \mathcal{E}_{v,w}$ is the overall bad event that G does not have diameter two. By symmetry, all of these events $\mathcal{E}_{v,w}$ have the same probability; let's call this probability q (which is of course implicitly a function of the parameters n and p). By Union Bound, we have

$$\begin{aligned} \Pr[\mathcal{E}] &\le \sum_{v,w} \Pr[\mathcal{E}_{v,w}] \\ &\le \binom{n}{2} q. \end{aligned}$$

So for a given n, if we choose p large enough that $q \le n^{-3}$, then the right-hand side of the last inequality will be $\le n^{-1}$, and we'll show that the bad event happens with at most this probability. In other words, G will have diameter two with high probability.

So let's consider an arbitrary pair of nodes v, w and ask what it would take to get $q = \Pr[\mathcal{E}_{v,w}] \le n^{-3}$. For $\mathcal{E}_{v,w}$ to occur, there needs to be no edge from v to w, and also no two-step path through any of the other $n-2$ nodes u. Therefore,

$$\begin{aligned}\Pr[\mathcal{E}_{v,w}] &= (1-p)(1-p^2)^{n-2} \\ &\le (1-p^2)^{n-2}.\end{aligned}$$

Here the first term $(1-p)$ on the right-hand of the first line reflects the probabiity that there is no edge from v to w, and the second term $(1-p^2)^{n-2}$ reflects that probability that we fail to put in the two edges from v to u and u to w for each of the $n-2$ other nodes u.

For simplicity in the calculations to follow, it's useful to parametrize p as $r/(n-2)$ for some number r. Using this parametrization, we write

$$\begin{aligned}\Pr[\mathcal{E}_{v,w}] &\le (1-p^2)^{n-2} \\ &= \left(1 - \frac{r^2}{(n-2)^2}\right)^{n-2} \\ &= \left(\left((1 - \frac{r^2}{(n-2)^2}\right)^{\frac{(n-1)^2}{r^2}}\right)^{\left(\frac{r^2}{n-2}\right)}.\end{aligned}$$

As in our calculations earlier in this section, we recognize the expression inside the outermost parentheses as having the form $(1-\frac{1}{m})^m$, and therefore at most $1/e$. It follows that

$$\Pr[\mathcal{E}_{v,w}] \le \left(\frac{1}{e}\right)^{\left(\frac{r^2}{n-2}\right)}.$$

So to get $\Pr[\mathcal{E}_{v,w}] \le n^{-3}$, which is what we're aiming for, we can simply choose r so that the exponent in this last expression is $3 \ln n$. Therefore we want

$$\frac{r^2}{n-2} = 3 \ln n$$

and so

$$r = \sqrt{3(n-2)\ln n}.$$

Since we've parametrized things so that $p = r/(n-2)$, this means that

$$p = \sqrt{\frac{3 \ln n}{n-2}}$$

is sufficient.

2 Random Graphs with Fixed Degrees

Suppose we want to completely specify the sequence of degrees in our random graph; that is, we want the n nodes to have degrees $d_1, d_2, \ldots, d_n$, and then produce a random graph subject to this constraint. We'll allow graphs that have self-loops (an edge goes from a node to itself) and parallel edges (two edges connect the same pair of nodes), which will make it much easier to construct random graphs with the desired properties. We'll also assume $\sum_i d_i$ is an even number, without which no graph with the desired degrees can exist.

One way to create a random graph with the specified degrees is to first define a set of nodes labeled $1, 2, \ldots, n$ in which node i has d_i "half-edges" sticking out of it. We then simply choose a random pairing on all $\sum_i d_i$ half-edges, glue the paired half-edges together, and declare the resulting graph to be our random graph G. (Note how self-loops and parallel edges may indeed arise from this construction.)

Now, let's consider the special case of this construction in which all d_i are equal to some constant $d > 0$. Graphs produced by this special case of the construction are called *random d-regular graphs*, meaning that all nodes have degree d. When $d = 1$, the only possible graph is a perfect matching (i.e. a collection of disjoint edges), and when $d = 2$, the only possible graphs are collections of disjoint cycles. But things get much more interesting once $d = 3$, and one of the fundamental properties of a random 3-regular is that it has good *expansion*.

3 Expansion

The expansion of a graph is the minimum "surface-to-volume" ratio of any set of nodes. We define this more precisely as follows. 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; while people have discovered more elementary analyses over time, they are still non-trivial.

In contrast, it is more straightforward to show that a random graph with constant degree d has expansion at least α, for some absolute constant $\alpha > 0$, with positive probability. We will do this later in these notes. First, however, we will establish a basic property of expansion: that it implies that all pairs of nodes in the graph are connected by short paths.

4 Expansion implies short paths

Intuitively, expansion implies a strong "robustness" to the graph: to split it into multiple large pieces, one must destroy correspondingly many edges. This property has many other consequences; later in the course, for example, we'll that it implies that a random walk on the graph "mixes" rapidly (approaching its stationary distribution in a small number of steps). For right now, we prove the following "small-world" property of graphs with good expansion:

> *If an n-node graph of maximum degree d has expansion at least α, then every pair of nodes s and t is connected by a path of length at most $O(\frac{d}{\alpha}\log n)$.*

To prove this, we try constructing a path from s to t using breadth-first search (BFS). Let S_j be the set of nodes encountered anywhere in the first j levels of the BFS outward from s. To determine the next, $(j+1)^{\text{st}}$, level of the BFS, we need to follow all the edges out of S_j; the nodes that these edges lead to, together with S_j, will form the set S_{j+1}.

As long as S_j consists of fewer than $n/2$ nodes, the expansion of G implies that it has at least $\alpha|S_j|$ edges leading out of it. Some of these edges may lead to the same nodes, but since no node has degree more than d, we can conclude that at least $\frac{\alpha}{d}|S_j|$ new node are discovered by looking one more BFS level out from S_j. In other words,

$$|S_{j+1}| \geq (1 + \frac{\alpha}{d})|S_j|.$$

This says that the BFS layers out from s grow exponentially, due to the expansion of G, and so as long as S_j has fewer than half the nodes, we have

$$|S_j| \geq (1 + \frac{\alpha}{d})^j.$$

Now, since $\alpha < d$, we get the following by choosing $\ell = \frac{d}{\alpha} \log n$:

$$(1 + \frac{\alpha}{d})^\ell = (1 + \frac{\alpha}{d})^{\frac{d}{\alpha} \log n} > 2^{\log n} = n.$$

Here we go from the second expression to the third using the fact that $(1 + \frac{1}{k})^k$ increases from 2 to e as k ranges from 1 to infinity. So in particular, since $\alpha < d$, the quantity

$$(1 + \frac{\alpha}{d})^{\frac{d}{\alpha}}$$

is greater than 2.

What's the point of this calculation? The point is that the size of S_j can never exceed n, so sometime in the first $\ell = \frac{d}{\alpha} \log n$ steps of the BFS, the inequality

$$|S_{j+1}| \geq (1 + \frac{\alpha}{d})|S_j|$$

must stop holding — and this only happens once S_j contains strictly more than half the nodes.

Let's consider the first $j \leq \frac{d}{\alpha} \log n$ when $|S_j|$ strictly exceeds $n/2$. If t belongs to this set S_j, then we have the short path from s to t that we wanted. If t doesn't belong to this set S_j, then we do the following: we repeat this construction, but starting the BFS outward from t. Again, in at most $\frac{d}{\alpha} \log n$ BFS levels from t, we have set T_i that contains strictly more than half the nodes. Now, S_j and T_i each contain more than half the nodes of the graph, so there must be at least one node that's in both; call this node v. By finding a short s-v path through S_j, and gluing it together with a short t-v path through T_i, we have the desired short path from s to t.

5 Hypercubes are Expanders

Recall that we only know fairly complex proofs that explicitly defined constant-degree graphs have constant expansion. As a result, in the next section, we will prove instead that a random graph of (sufficiently high) constant degree has constant expansion with positive probability. Before doing this, let's establish that if we are willing to slightly increase the degree in the graph — to let it be $\log n$ rather than a constant — then it becomes much easier to verify expansion for certain natural families of graphs. In particular, in this section we will show that the family of *hypercubes* has constant expansion.

Defining hypercubes. We begin by defining this family of graphs. We say that the *hypercube of dimension* d, denoted H_d, is the graph whose nodes are all possible d-bit strings, and where we join two nodes by an edge if their associated strings differ in a single bit.

Thus, the graph H_d has $n = 2^d$ nodes. A given node v of H_d is neighbors with every node that differs from it in one bit position; as a result, v has $d = \log n$ neighbors. (All logarithms in this section will be base-2.) Also, to get from any node v to any other node w in H_d, we can construct a path by changing the bit string for v to the bit string for w one bit at a time. This will produce a path of at most $d = \log n$ edges; and a path of this length is necessary when the bit strings for v and w differ in every position. So the diameter of H_d is exactly $\log n$.

There is a recursive construction of hypercubes that is useful to know, as follows. Suppose we start with two copies of H_{d-1}, and for a node v in the first copy, we say that its *twin* in the second copy is the node w with the same bit string as v; we will denote this node by $t(v)$, and for a set S of nodes in the first copy, we will use $t(S)$ to denote the twins of all nodes in S. Suppose we take the bit strings for the nodes in these two copies of H_{d-1}, append a 0 to the strings for all nodes in the first copy, append a 1 to the strings for all nodes in the second copy, and add edges between each node v in the first copy and its twin $t(v)$ in the second copy. In this way, we've produced a copy of H_d: in other words, H_d consists of two copies of H_{d-1} joined together by a perfect matching between twins that differ only in their final bit.

Building up fro small examples, we see that in particular, H_1 consists of two nodes joined by an edge, H_2 consists of the vertices and edges of a square, and H_3 consists of the vertices and edges of a cube. (These small geometric examples are the reason these types of graphs are called "hypercubes.")

Hypercubes have expansion 1. Now, let's consider the expansion of the hypercube H_d. First, consider its recursive construction from two copies of H_{d-1}, and let A and $t(A)$ denote the nodes in these two copies. Since A has exactly $n/2$ edges with one end in A — i.e., all edges from a node in A to its twin in $t(A)$ — we see that $|A| = n/2$ and $|e^{out}(A)| = n/2$. It follows that A is a set with surface-to-volume ratio 1, and therefore the expansion $\alpha(H_d)$ is at most 1 (since the expansion is the minimum surface-to-volume ratio of any set of size at most $n/2$).

Now let's prove that for every set S of at most $n/2$ nodes in H_d, we have $|e^{out}(S)| \geq |S|$; this would show that every set S of at most $n/2$ nodes in H_d has surface-to-volume ratio at least 1, and so $\alpha(H_d) \geq 1$. Combined with the fact we've already established that $\alpha(H_d) \leq 1$, this will show that $\alpha(H_d)$ is exactly 1.

An inductive proof. We will prove that $|e^{out}(S)| \geq |S|$ for all sets of at most $n/2$ nodes in H_d by induction, using the recursive construction of H_d. First, as a base case, we observe that this fact holds for H_1, since H_1 consists of just two nodes joined by an edge. Next, suppose for the inductive step that this property holds for H_{d-1}, and recall the two copies of H_{d-1} that we use to construct our copy of H_d, with A and $t(A)$ the sets of nodes in these two copies.

Now, consider any set S of at most $n/2$ nodes in H_d. Let $a = |S \cap A|$ and $b = |S \cap t(A)|$ be the number of nodes of S that lie in each of the two copies of H_{d-1} that make up our copy of H_d. Exchanging the roles of A and $t(A)$ if necessary (by symmetry), we can assume that $a \geq b$.

First, notice that we can't have $b > n/4$, since then also $a > n/4$, and since $|S| = a + b$, we wouldn't have $|S| \leq n/2$. So it follows that $b \leq n/4$. As a result, we can apply the induction hypothesis to the set $S \cap t(A)$ in the copy of the hypercube H_{d-1} on $t(A)$: since $S \cap t(A)$ is a set of $b \leq n/4$ nodes in a graph (H_{d-1}) with $n/2$ nodes and expansion 1, there are at least b edges that leave $S \cap t(A)$ and go to nodes in $t(A) - S$. Notice that these b edges all contribute to $e^{out}(S)$, a fact that will be useful in the proof (since we want to establish that $e^{out}(S)$ is large).

We finish the proof by considering two cases.

- **Case 1:** $a \leq n/4$. In this case, we can use the same argument via the induction hypothesis in A that we just used in $t(A)$: the set $S \cap A$ consists of at most $n/4$ nodes in a graph with $n/2$ nodes and expansion 1. Therefore, least a nodes leave $S \cap A$ and go to nodes in $A - S$, all of which contribute to $e^{out}(S)$. Adding these to the $\geq b$ different edges that contribute to $e^{out}(S)$ in $t(A)$, we see that $|e^{out}(S)| \geq a + b = |S|$, completing the proof in this first case.

- **Case2 2:** $a > n/4$. The structure of the argument for this case is depicted in Figure 1. In this case, we can't apply the induction hypothesis to the set $S \cap A$ in H_{d-1}, since $S \cap A$ consists of more than half the nodes of H_{d-1}, and guarantees about e^{out} only apply to sets of at most half the total number of nodes. But instead we can apply it to the complement $A - S$ in A, which has $n/2 - a \leq n/4$ nodes. By the induction hypothesis, this set contributes at least $n/2 - a$ edges in A to $e^{out}(S)$.

  Also, $S \cap A$ has a nodes while $S \cap t(A)$ has only b nodes, so there are at least $a - b$ nodes in $S \cap A$ whose twin does not belong to $S \cap t(A)$. For each of these $a - b$ nodes, their edge to their twin contributes to $e^{out(S)}$, since its end in A belongs to S and its end in $t(A)$ does not.

  So let's count up all the edges in $e^{out(S)}$ that we know about in this case, from all our arguments so far: at least b inside $t(A)$, at least $n/2 - a$ inside A, and at least $a - b$ crossing between A and $t(A)$. This is a total of at least

  $$b + (n/2 - a) + (a - b) = n/2$$

  edges. Thus, $|e^{out}(S)| \geq n/2$ while $|S| \leq n/2$, so we have $|e^{out(S)}| \geq |S|$ in this case as well.

Having established $|e^{out(S)}| \geq |S|$ in both cases, we've shown that $\alpha(H_d) \geq 1$, and this completes the proof.

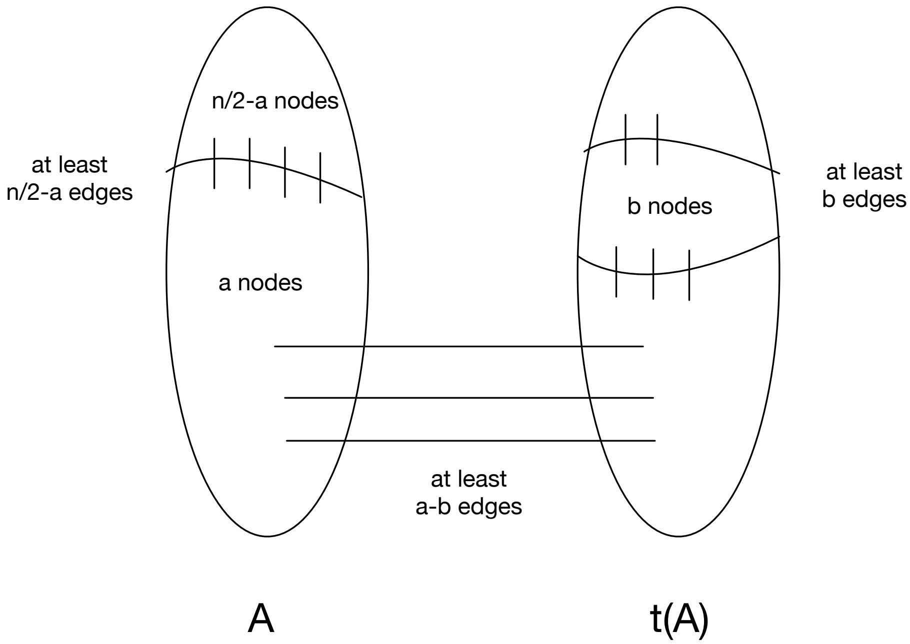

Figure 1: A schematic illustration of Case 2 of the induction proof, in terms of the two copies of H_{d-1} that make up the hypercube H_d.

6 Constant-Degree Random Graphs are Expanders

We will now show that a simple random construction produces good expander graphs with constant probability. In light of our discussion in the previous sections, 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 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

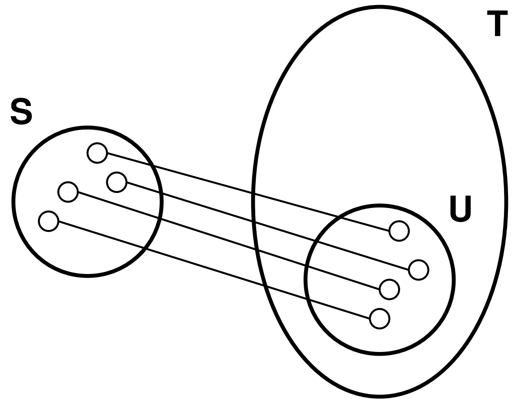

Figure 2: The event $\mathcal{E}_{ST}$ in the analysis of the expander construction.

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

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

Now, the graph G is built from d random perfect matchings, so if $k = |S| \leq n/6$ and $|T| = 4k$, then

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

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

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

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:

$$\sum_{\substack{|S|\leq n/6 \\ |T|=4|S|}} \Pr\left[\mathcal{E}_{ST}\right] \leq \sum_{k=1}^{n/6} \binom{n}{k}\binom{n}{4k}\left(\frac{4.8k}{n}\right)^{dk/2}$$

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

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. ■