Random processes in extremal combinatorics

Peter Allen

PCC 2022

The following section should be read before the tutorial. You will need to know all the 'basic probability' (you probably already do, in which case this is just a reminder). And you will need to know what Theorem 1.5 says and what I mean by 'with high probability'. Most of the section is here for completeness: you don't need to know how to prove Theorem 1.5 to apply it, though I think it helps to have some idea (and how it relates to the Chernoff and Hoeffding bounds you might be more familiar with). And there is finally a template idea of how to apply Theorem 1.5 which might help you get an idea of what the various parts of that theorem are doing.

In the tutorial, I'll build up a bit more slowly and start with applications where some inputs to Theorem 1.5 are set 'trivially' before getting to ones where we need to use the full version.

The intention is to use these tools to prove (simplified versions of) two results in extremal combinatorics: the Blow-up Lemma and a packing statement. The former only needs 'crude' estimates, whereas the latter needs 'accurate' estimates that often go under the name 'differential equations method' in the literature. Both of these—but especially the latter—need quite a bit of calculation. One aim of this tutorial is to show how to do these calculations and that they are really quite routine, even if long.

The other aim of this tutorial is to point out a pitfall with the 'differential equations method' that seems not to be in the literature and which can lead to wasting rather a lot of time the first time you fall into it.

The intention is that once the tutorial starts there will be a version of these notes released with the statements and proofs for these two results, and in addition several exercises.

1 Notation, basic probability and martingales: assumed knowledge

1.1 Notation

We'll use standard graph notation. $N_H(v)$ means the neighbourhood in graph H of vertex v; $N_H(u,v) = \{w \in V(H) : uw, vw \in E(H)\}$ means the common neighbourhood of u and v, and so on.

We will usually use letters like ψ and ϕ for (partial) graph embeddings. A graph embedding ψ of G into H means an injective map from the vertices of G to those of H, such that if $uv \in E(G)$ then $\psi(u)\psi(v) \in E(H)$. We will often talk about *partial* embeddings, which means that ψ maps only a subset of the vertices of G into H (we will always be clear what the subset is). If ψ is clear from the context, we might write $x \hookrightarrow u$ for 'x is embedded to u', where $x \in V(G)$ and $u \in V(H)$, and in particular if ψ is a partial embedding from G to H, and x is not one of the vertices embedded by ψ, we'll write $\psi \cup \{x \hookrightarrow u\}$ to denote the map extending ψ which in addition maps x to u. In addition, I will try to stick to the convention that we always embed the **g**uest graph G into the **h**ost graph H. Letters x, y, z will be in the guest graph; u, v, w will be in the host graph.

We'll rather often want to talk about the *image*

$$\operatorname{im}\psi = \{u : \exists x \in V(G) \text{ such that } \psi(x) = u\}$$

of a (partial) embedding ψ, which is a set of vertices in H.

We will need to do a lot of calculation with error bounds. We'll use interval arithmetic for this: $a \pm b$ means 'some quantity in $[a-b, a+b]$', where we are thinking of a as the 'large' main term and b as the 'small' error term. We will add, multiply, exponentiate and so on these quantities. We will never try to be all that precise with error estimates; we'll routinely do things like rounding up a factor of 42 in the error term to 100, and so on. In particular, that means that any time the symbol $\pm$ appears in an equation, then $=$ is not going to be symmetric. We will often write something like $(a \pm b)^2 = a^2 \pm (2ab + b^2) = a^2 \pm 3ab$. You should read this as meaning that if you take any two numbers in $[a-b, a+b]$ and multiply them, then you will end up with a number within $2ab + b^2$ of a^2, and (assuming $|b| < |a|$, which we will) such a number is also within $3ab$ of a^2. Of course, it's not true that any number within $3ab$ of a^2 can actually be written as a product of two numbers in $[a-b, a+b]$; we can only go left-to-right.

Finally, we will not bother to write floors and ceilings. If something that looks like a real number is supposed to be an integer, you can imagine there are floor or ceiling symbols and the errors this makes will disappear in the errors in our interval arithmetic.

1.2 Basic probability

We will not need a lot of probability theory. In particular, we will always work with finite probability spaces, which means we will not bother talking about σ-algebras; for us all sets are measurable (and we will not need to use this word, which otherwise would appear in most definitions). So a probability space is $(\Omega, \mathbb{P})$, where Ω is a finite set and $\mathbb{P} : 2^\Omega \to [0,1]$ assigns a probability to each subset of Ω, with the conditions $\mathbb{P}[\emptyset] = 0$, $\mathbb{P}[\Omega] = 1$ and $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B]$ for disjoint $A, B \subseteq \Omega$.

An *event* is a subset $E \subseteq \Omega$. We will only be interested in events with positive probability. One way to do this is to insist no element of Ω has zero probability, and formally we could do everything in these notes this way (by removing such elements). It is however convenient not to do this, so in what follows 'event' should always be read as 'event with positive probability'. A collection $A_1, \ldots, A_n$ of events are *independent* if

$$\mathbb{P}\Big[\bigcap_{i \in I} A_i\Big] = \prod_{i \in I} \mathbb{P}[A_i]$$

for every $I \subseteq [n]$. (There are weaker versions of independence, such as pairwise independence in which we restrict to $|I| = 2$; we will not talk about these but remind that this is a strictly weaker condition).

A real-valued random variable X is a function from Ω to $\mathbb{R}$; more generally, a random variable is any function with domain Ω. We will usually be interested in non-negative random variables, and we (as usual) will normally write X where we should really write $X(\omega)$, where $\omega \in \Omega$ is chosen according to $\mathbb{P}$. We'll write $X(\omega)$ in this section in a few places where we want to be clear about definitions, but it will not appear after that.

Two random variables X and Y are independent if for all choices of x and y the events $X = x$ and $Y = y$ are independent (and so on for more random variables).

The *expectation* $\mathbb{E}[X]$ is defined to be $\sum_{x \in \mathbb{R}} x\, \mathbb{P}[X = x]$. Note that this is a finite sum! An important consequence of the definition is *linearity of expectation*:

$$\mathbb{E}[aX + bY] = a\,\mathbb{E}[X] + b\,\mathbb{E}[Y]\,.$$

By expanding the brackets, we can also check that if X and Y are independent, we have $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$. Another immediate consequence of the definition is *Markov's inequality*: if X is non-negative and $x > 0$, then

$$\mathbb{P}[X \geq x] \leq \tfrac{\mathbb{E}[X]}{x}\,.$$

The kth moment of X is $\mathbb{E}[X^k]$. Applying Markov's inequality to $\big(X - \mathbb{E}[X]\big)^2$, and noticing

$$\mathbb{E}\Big[\big(X - \mathbb{E}[X]\big)^2\Big] = \mathbb{E}[X^2] - 2\,\mathbb{E}[X]^2 + \mathbb{E}[X]^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2\,,$$

we have *Chebyshev's inequality*

$$\mathbb{P}\Big[\big|X - \mathbb{E}[X]\big| \geq a\Big] = \mathbb{P}\Big[\big(X - \mathbb{E}[X]\big)^2 \geq a^2\Big] \leq \frac{\mathbb{E}[X^2] - \mathbb{E}[X]^2}{a^2}$$

where $a > 0$.

If E is an event, then the *conditional probability* of $A \subseteq \Omega$ on E is

$$\mathbb{P}[A|E] = \tfrac{\mathbb{P}[A \cap E]}{\mathbb{P}[E]}\,.$$

Similarly, the *conditional expectation* of X on E is

$$\mathbb{E}[X|E] = \sum_{x \in \mathbb{R}} x\, \mathbb{P}[X = x|E]\,.$$

Note that if Y is a random variable and y is a possible value of the random variable, the set $\{Y = y\} = \{\omega \in \Omega : Y(\omega) = y\}$ is an event, so in particular we can write $\mathbb{P}[X|Y = y]$ and $\mathbb{E}[X|Y = y]$, which are real numbers.

We will often want to write $\mathbb{P}[X|Y]$ and $\mathbb{E}[X|Y]$, which are *not* real numbers but random variables. We define $\mathbb{P}[X|Y](\omega)$ as follows: given ω, let $y = Y(\omega)$. We define $\mathbb{P}[X|Y](\omega) = \mathbb{P}[X|Y = y]$. Similarly, we define $\mathbb{E}[X|Y](\omega) = \mathbb{E}[X|Y = y]$ where $y = Y(\omega)$. One should think of this as: if you are told the outcome y of Y (but not the specific ω), then $\mathbb{E}[X|Y]$ is your expectation of X conditional on the current information $Y = y$.

A *Bernoulli* random variable $\mathrm{Ber}(p)$ takes values in $\{0, 1\}$, with probability p of taking 1. A *binomial* random variable $\mathrm{Bin}(n, p)$ is a sum of n independent $\mathrm{Ber}(p)$ random variables.

1.3 Chernoff-type bounds

We need *Jensen's inequality*. Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function, i.e. if $x < y < z$ then $\big(y, f(y)\big)$ lies below the line from $\big(x, f(x)\big)$ to $\big(z, f(z)\big)$. Then we have

$$\tfrac{1}{n} \sum_{i=1}^{n} f(x_i) \leq f\Big(\tfrac{\sum_{i=1}^{n} x_i}{n}\Big)\,,$$

with equality if $x_1 = \cdots = x_n$ (if f is strictly convex, i.e. we can replace 'below' by 'strictly below' then this is if and only if). A related *reverse Jensen* inequality is: if we require $x_i \in [a, b]$ for each i and we fix $\frac{1}{n}\sum_{i=1}^{n} x_i = x \in [a, b]$, then when $\sum_i f(x_i)$ is minimised subject to these two conditions, all but at most one of the x_i is in $\{a, b\}$. Both the Jensen and reverse Jensen inequalities also apply to distributions (one can think of a distribution as being the limit of point masses $x_1, \ldots, x_n$ each of weight $\frac{1}{n}$; and the exceptional value that is not in $\{a, b\}$ disappears in the limit). Of course, the same all holds for (strictly) concave functions swapping minimum and maximum, which we will also refer to as Jensen and reverse Jensen.

Lemma 1.1. *Let $X = Y_1 + \cdots + Y_n$ where the Y_i are independent Bernoulli random variables. Then*

$$\mathbb{E}\Big[e^{tX}\Big] \leq \Big(1 + (e^t - 1)\tfrac{\mathbb{E}[X]}{n}\Big)^n \leq \exp\Big((e^t - 1)\,\mathbb{E}[X]\Big)\,.$$

Proof. Let $p_i = \mathbb{P}[Y_i = 1]$ for each i, and let $p = \frac{1}{n}\,\mathbb{E}[X] = \frac{1}{n}\sum_{i=1}^{n} p_i$. By independence, we have

$$\mathbb{E}\Big[e^{tX}\Big] = \prod_{i=1}^{n} \mathbb{E}\Big[e^{tY_i}\Big] = \prod_{i=1}^{n} \big(1 - p_i + p_i e^t\big) \leq \big(1 - p + pe^t\big)^n\,,$$

where the final inequality is Jensen's inequality (to see this, take the log of the product). □

If $t > 0$, we can write $\mathbb{P}\big[X \geq a\big] = \mathbb{P}\big[e^{tX} \geq e^{ta}\big]$, and we can use Markov's inequality together with optimising $t > 0$ to deduce the following *concentration inequalities*.

Theorem 1.2 (Chernoff bounds)**.** *Let $X = Y_1 + \cdots + Y_n$ where the Y_i are independent Bernoulli random variables. For any $0 < \delta \leq \frac{3}{2}$ and any $s > 0$, we have*

$$\begin{aligned}
\mathbb{P}\big[X \leq (1-\delta)\,\mathbb{E}[X]\big] &\leq \exp\Big(-\tfrac{\delta^2\,\mathbb{E}[X]}{3}\Big)\,,\\
\mathbb{P}\big[X \geq (1+\delta)\,\mathbb{E}[X]\big] &\leq \exp\Big(-\tfrac{\delta^2\,\mathbb{E}[X]}{3}\Big) \quad \textit{and}\\
\mathbb{P}\big[X \geq (1+s)\,\mathbb{E}[X]\big] &\leq \exp\Big(\big(s-(1+s)\ln(1+s)\big)\,\mathbb{E}[X]\Big)\,.
\end{aligned}$$

Proof. We first prove the final inequality. By assumption, for any $t > 0$ we have

$$\mathbb{E}\Big[e^{tX}\Big] \leq e^{(e^t-1)\,\mathbb{E}[X]}\,.$$

We choose $t = \ln(1+s) > 0$ and write

$$\mathbb{P}\big[X \geq (1+s)\,\mathbb{E}[X]\big] = \mathbb{P}\big[e^{tX} \geq e^{t(1+s)\,\mathbb{E}[X]}\big] \leq \mathbb{E}\big[e^{tX}\big]e^{-t(1+s)\,\mathbb{E}[X]}\,,$$

where the final inequality is Markov's inequality. Substituting our upper bound for $\mathbb{E}\big[e^{tX}\big]$ we get

$$\mathbb{P}\big[X \geq (1+s)\,\mathbb{E}[X]\big] \leq e^{(e^t-1)\,\mathbb{E}[X]} \cdot e^{-t(1+s)\,\mathbb{E}[X]}$$

and putting in our choice of t we have

$$\mathbb{P}\big[X \geq (1+s)\,\mathbb{E}[X]\big] \leq \exp\Big(\big(s-(1+s)\ln(1+s)\big)\,\mathbb{E}[X]\Big)\,,$$

as required.

Observe that $s-(1+s)\ln(1+s) \leq -\frac{1}{3}s^2$ for $0 < s \leq \frac{3}{2}$, giving the required second inequality.

For the first inequality, we consider $X' = n - X$ and set δ' such that

$$(1+\delta')\big(n-\mathbb{E}[X]\big) = n-(1-\delta)\,\mathbb{E}[X]\,.$$

For $t > 0$ we use the upper bound

$$\mathbb{E}\Big[e^{tX'}\Big] \leq \Big(1+\big(e^t-1\big)\tfrac{n-\mathbb{E}[X]}{n}\Big)^n = e^{tn}\Big(1-\big(1-e^{-t}\big)\tfrac{\mathbb{E}[X]}{n}\Big)^n \leq e^{tn}\cdot e^{-(1-e^{-t})\,\mathbb{E}[X]}\,,$$

and much as above we get

$$\begin{aligned}
\mathbb{P}\big[X \leq (1-\delta)\,\mathbb{E}[X]\big] &= \mathbb{P}\big[X' \geq (1+\delta')\,\mathbb{E}[X']\big]\\
&= \mathbb{P}\Big[e^{tX'} \geq e^{t(1+\delta')(n-\mathbb{E}[X])}\Big]\\
&\leq \mathbb{E}\big[e^{tX'}\big]\cdot e^{-t(1+\delta')(n-\mathbb{E}[X])}\\
&\leq e^{tn}\cdot e^{-(1-e^{-t})\,\mathbb{E}[X]}\cdot e^{-t(n-(1-\delta)\,\mathbb{E}[X])}\\
&= e^{-(1-e^{-t})\,\mathbb{E}[X]+t(1-\delta)\,\mathbb{E}[X]}\,.
\end{aligned}$$

Optimising, we choose $t = -\ln(1-\delta) > 0$, and get

$$\mathbb{P}\big[X \leq (1-\delta)\,\mathbb{E}[X]\big] \leq \exp\Big(\big(-\delta-(1-\delta)\ln(1-\delta)\big)\,\mathbb{E}[X]\Big) \leq \exp\Big(-\tfrac{1}{3}\delta^2\,\mathbb{E}[X]\Big)\,,$$

by a similar approximation as above. □

We will usually use the first two of these; the last is a sharper bound, useful if s is very large. What is important to observe here is that we only use the assumption on the structure of X in order to apply Lemma 1.1 to get an upper bound on its moment generating function. So the same bounds apply to any X satisfying the conclusion of Lemma 1.1.

In particular, if $\mathbb{E}[X] \leq a$ then we get the two upper bounds from Theorem 1.2 with a replacing $\mathbb{E}[X]$ in the right-hand side formula (so $\exp\big(-\frac{\delta^2 a}{3}\big)$ and so on), and if $\mathbb{E}[X] \geq a$ we get the similar lower bound statement. The same observation applies to the remaining three probabilistic theorems in this section: if we only want to bound the probability of X being large, we only need an upper bound on the expectation (and not a matching or indeed any lower bound).

It is easy to check that the above argument still goes through if each Y_i takes values in $[0,1]$: by the reverse Jensen inequality, $\mathbb{E}[e^{tY_i}]$ is maximised when Y_i is Bernoulli, and so we get the conclusion of Lemma 1.1 if each Y_i is not Bernoulli but just supported on $[0,1]$. By scaling (i.e. considering the random variable $X' = X/R$), we get

Corollary 1.3 (Hoeffding bounds)**.** *Let $R > 0$ and let $X = Y_1 + \cdots + Y_n$ where the Y_i are independent random variables taking values in $[0,R]$. For any $0 < \delta < \frac{3}{2}$, we have*

$$\mathbb{P}\big[X \neq (1 \pm \delta)\,\mathbb{E}[X]\big] \leq 2\exp\Big(-\tfrac{\delta^2\,\mathbb{E}[X]}{3R}\Big).$$

Proof. Let $X' = X/R$, and let $Y_i' = Y_i/R$ for each $1 \leq i \leq n$. Then $\mathbb{E}[X'] = \mathbb{E}[X]/R$, and we have $X \neq (1 \pm \delta)\,\mathbb{E}[X]$ if and only if $X' \neq (1 \pm \delta)\,\mathbb{E}[X']$. Applying Theorem 1.2 to X' we get the claimed inequality. □

1.4 Martingale concentration bounds

When we analyse random processes, we will usually be interested in summing random variables which are *not* independent: for example, if we are embedding, one by one, the vertices of a graph G into some H 'randomly', we might want to know how many, from the first t, of the vertices of G have been embedded to $N_H(v)$ for some given $v \in V(H)$. We can define $Y_i = \mathbb{1}[i \hookrightarrow N_H(v)]$ (that is, the Bernoulli random variable taking value 1 when the condition in brackets, that vertex number i of G is embedded to $N_H(v)$, is True), but these random variables are clearly not independent, so the Chernoff and Hoeffding bounds do not apply. On the other hand, in this example we have some feeling that conditioning on any given event $Y_i = 1$ should not affect the probability of Y_j being 1 for any $j > i$ too much; we should still be able to apply something like the Chernoff bound. The usual way to formalise this is to define a *martingale* and prove martingale concentration inequalities. We are going to avoid this, and instead just show that under reasonable conditions we will still get the conclusion of Lemma 1.1, from which concentration inequalities for $\sum_{i=1}^n Y_i$ follow exactly as above. Since what we are doing here is really a martingale analysis, we will call these bounds martingale concentration inequalities.

To begin with we state something easy to prove, which turns out not to be quite what we want in applications. In order to do this, we need a definition. Those familiar with martingales will recognise that this is (more or less) the concept of a *filtration*.

We will be thinking of random processes in which choices are made at positive integer times (as in the above example, where at time t we embed the tth vertex of G). A *history*

of a random process up to a given time τ will be denoted by the symbol $\mathscr{H}$, and it means a complete record of the random choices up to τ (in the example, a list of all the vertex embeddings made in embedding G). Equivalently, letting Ω be the probability space on which the random process is defined (so a given $\omega \in \Omega$ is a complete record of all the random choices made in a particular outcome of the random process) we can think of $\mathscr{H}$ as a random variable, with $\mathscr{H}(\omega)$ being the set of all elements in Ω which agree with ω up to time τ. Thus the collection of possible values of $\mathscr{H}$ form a partition of Ω. In particular, if Y is a random variable, we have defined the conditional expectation $\mathbb{E}[Y|\mathscr{H}]$. For some random variables Y we may have the property that $Y(\omega') = Y(\omega)$ for each $\omega' \in \mathscr{H}(\omega)$ and all $\omega \in \Omega$, i.e. Y is a constant function on each part of the partition defined by $\mathscr{H}$. Then we will say Y *is defined by* $\mathscr{H}$; intuitively, we only need to observe the process up to time τ in order to know the value of the random variable Y (This is equivalent to saying Y is measurable according to the σ-algebra associated to time τ of the filtration, which is what probabilists write but which I find confusing..). For example, the random variable $\mathbb{1}[i \hookrightarrow N_H(v)]$ is defined by the history up to time i (we have embedded the vertex i; we know if it has been embedded to $N_H(v)$ or not) and for any later time, but in general it isn't defined by the history up to any time before i (we haven't yet embedded it and we in general have some chance strictly between 0 and 1 of embedding it to $N_H(v)$).

Finally, if we are given a collection of increasing times $\tau_0, \tau_1, \ldots, \tau_n$, we can write $\mathscr{H}_0, \mathscr{H}_1, \ldots, \mathscr{H}_n$ to denote the histories up to τ_0 and so on ('a filtration'). We will be interested in summing random variables $Y_1, \ldots, Y_n$ such that Y_i is defined by $\mathscr{H}_i$ for each i. Again in the example, we would take $\tau_i = i$ for each i; $\mathscr{H}_0$ is simply the trivial history (the partition of Ω into one part) and observe that the Bernoulli random variable Y_i is indeed defined by $\mathscr{H}_i$ (the probabilists will say the martingale is adapted to the filtration).

What we get out of this setup is that Chernoff- and Hoeffding-type bounds hold for $X = \sum_{i=1}^n Y_i$; these quantities are likely to be close to the sum $\sum_{i=1}^n \mathbb{E}[Y_i|\mathscr{H}_{i-1}]$. We should notice that this statement is a bit subtle: the sum of conditional expectations is itself a random variable! So what we are really saying is that there are two random variables, and it is unlikely that their values are far apart. For our applications, though, the sum of conditional expectations will deterministically lie in a small interval $[a-b, a+b]$.

Theorem 1.4. *Let $R > 0$. Given a random process with a sequence of histories $\mathscr{H}_0, \ldots, \mathscr{H}_n$ and random variables $Y_1, \ldots, Y_n$ such that $Y_i \in [0, R]$ and Y_i is determined by $\mathscr{H}_i$ for each i, suppose that $\sum_{i=1}^n \mathbb{E}[Y_i|\mathscr{H}_{i-1}] \in [a-b, a+b]$ holds with probability 1, where $a > 0$ and $b < \frac{1}{4}\big(\min(a, n-a)\big)$. Let $X = \sum_{i=1}^n Y_i$, then we have for all $0 < \delta < \frac{3}{2}$ and all $s > 0$*

$$\mathbb{P}[X \geq a + b + \delta a] \leq \exp\Big(-\tfrac{\delta^2 a}{6R}\Big),$$
$$\mathbb{P}[X \leq a - b - \delta a] \leq \exp\Big(-\tfrac{\delta^2 a}{6R}\Big) \quad \textit{and}$$
$$\mathbb{P}[X \geq a + b + sa] \leq \exp\Big(\tfrac{s-(1+s)\ln(1+s)}{2R} a\Big).$$

Proof. By scaling (dividing through by R) it is enough to prove the case $R = 1$. Observe that since $a - b > \frac{a}{2}$ and $n - (a+b) > \frac{1}{2}(n-a)$, it is enough to prove bounds on the moment generating function matching Lemma 1.1, and then the desired probability bounds follow as in the proof of Theorem 1.2. We'll give the proof for the case that $\mathscr{H}_0 = \{\Omega\}$ (i.e. we calculate the expectation of Y_1 before making any random choices, so conditioning on $\mathscr{H}_0$ is redundant); check you can see how to modify the proof to deal with the general case.

We can write

$$\begin{aligned}\mathbb{E}\big[e^{tX}\big] &= \mathbb{E}\big[e^{t(Y_1+Y_2+\cdots+Y_n)}\big] \\ &= \mathbb{E}\big[e^{tY_1}\cdot e^{t(Y_2+\cdots+Y_n)}\big] \\ &= \sum_{y_1} e^{ty_1}\,\mathbb{P}[Y_1=y_1]\,\mathbb{E}\big[e^{t(Y_2+\cdots+Y_n)}\big|Y_1=y_1\big]\end{aligned}$$

where the final equality is immediate from the definition of conditional expectation. Let $X' = Y_2+\cdots+Y_n$. Observe that since $\sum_{i=1}^n \mathbb{E}[Y_i|\mathscr{H}_{i-1}] \le a+b$, we have the deterministic bound $\sum_{i=2}^n \mathbb{E}[Y_i|\mathscr{H}_{i-1}] \le a+b-\mathbb{E}[Y_1]$. What this suggests is to try to prove our upper bound on the moment generating function by induction on n. That is, we want to prove

$$\mathbb{E}\big[e^{tX}\big] \le \Big(1+\big(e^t-1\big)\tfrac{a+b}{n}\Big)^n$$

holds, where $a+b$ is the deterministic upper bound on the sum of conditional expectations and n is the number of summands in X. The $n=1$ base case is immediate: we have

$$\sum_{y_1} e^{ty_1}\,\mathbb{P}[Y_1=y_1] \le 1\cdot(1-\mathbb{E}[Y_1]) + e^t\cdot\mathbb{E}[Y_1] \le 1+(e^t-1)(a+b)$$

where the first inequality is the reverse Jensen inequality. For the induction step, we use essentially the same calculation together with the induction hypothesis which gives

$$\mathbb{E}\big[e^{t(Y_2+\cdots+Y_n)}\big|Y_1=y_1\big] \le \Big(1+\big(e^t-1\big)\tfrac{a+b-\mathbb{E}[Y_1]}{n-1}\Big)^{n-1}.$$

We get

$$\begin{aligned}\mathbb{E}\big[e^{tX}\big] &= \mathbb{E}\big[e^{t(Y_1+Y_2+\cdots+Y_n)}\big] \\ &= \sum_{y_1} e^{ty_1}\,\mathbb{P}[Y_1=y_1]\,\mathbb{E}\big[e^{t(Y_2+\cdots+Y_n)}\big|Y_1=y_1\big] \\ &\le \sum_{y_1} e^{ty_1}\,\mathbb{P}[Y_1=y_1]\Big(1+\big(e^t-1\big)\tfrac{a+b-\mathbb{E}[Y_1]}{n-1}\Big)^{n-1} \\ &= \Big(1+\big(e^t-1\big)\tfrac{a+b-\mathbb{E}[Y_1]}{n-1}\Big)^{n-1}\sum_{y_1} e^{ty_1}\,\mathbb{P}[Y_1=y_1] \\ &\le \Big(1+\big(e^t-1\big)\tfrac{a+b-\mathbb{E}[Y_1]}{n-1}\Big)^{n-1}\cdot\Big(1+\big(e^t-1\big)\,\mathbb{E}[Y_1]\Big) \\ &\le \Big(1+\big(e^t-1\big)\tfrac{a+b}{n}\Big)^n\end{aligned}$$

where the final inequality is Jensen's inequality applied to the log of the product. This upper bound together with the proof of Theorem 1.2 gives the first and third desired probability statements; to obtain the second, we get the corresponding upper bound on $n-X$ from the assumption $\sum_{i=1}^n 1-\mathbb{E}[Y_i|\mathscr{H}_{i-1}] \le n-(a-b)$ and continue as in the proof of Theorem 1.2. □

Finally, we are in a position to state and prove the martingale concentration inequality we actually want. That is, we usually will *not* actually be able to prove deterministic bounds on the sum of conditional expectation; we will only be able to prove such bounds under the

assumption that our random process has not done something unusual. The convenient way to formalise this is that we define a 'good event' $\mathcal{E}$, and assume that if $\mathcal{E}$ occurs then we do have our deterministic bounds on the sum of conditional expectations. Then we will conclude that the probability of $\mathcal{E}$ occurring and X nevertheless being far from its expectation is as in Theorem 1.4. Something which is rather important to note here is that we do not need to assume anything at all about $\mathcal{E}$; it doesn't, in particular, have to be a high probability event (though we will usually use it this way) or have any special structure.

Theorem 1.5. *Let $R > 0$. Given a random process with probability space Ω, and any $\mathcal{E} \subseteq \Omega$, with a sequence of histories $\mathscr{H}_0, \dots, \mathscr{H}_n$ and random variables $Y_1, \dots, Y_n$ such that $Y_i \in [0, R]$ and Y_i is determined by $\mathscr{H}_i$ for each i, suppose that $\sum_{i=1}^n \mathbb{E}[Y_i|\mathscr{H}_{i-1}] \in [a-b, a+b]$ holds whenever $\mathcal{E}$ occurs, where $a > 0$ and $b < \frac{1}{4}\big(\min(a, n-a)\big)$. Let $X = \sum_{i=1}^n Y_i$, then we have for all $0 < \delta < \frac{3}{2}$ and all $s > 0$*

$$\mathbb{P}[X \geq a + b + \delta a \text{ and } \mathcal{E} \text{ occurs}] \leq \exp\Big(-\tfrac{\delta^2 a}{6R}\Big),$$
$$\mathbb{P}[X \leq a - b - \delta a \text{ and } \mathcal{E} \text{ occurs}] \leq \exp\Big(-\tfrac{\delta^2 a}{6R}\Big) \quad \text{and}$$
$$\mathbb{P}[X \geq a + b + sa \text{ and } \mathcal{E} \text{ occurs}] \leq \exp\Big(\tfrac{s-(1+s)\ln(1+s)}{2R} a\Big).$$

Proof. The idea of this proof is simple: we define random variables $Y_1', \dots, Y_n'$ and $X' = Y_1' + \dots + Y_n'$ to which Theorem 1.4 applies, and such that whenever $\mathcal{E}$ occurs we have $Y_i = Y_i'$ for all i (this is a *coupling*).

We do this in the simplest possible way: we run the random process, and observe how the sum $\sum_{i=1}^t \mathbb{E}[Y_i|\mathscr{H}_{i-1}]$ grows. If, for the given run of the random process, at a given time t this sum has exceeded $a + b$, then we set $Y_t' = 0$ (and we will therefore do so for all future times) and thus $\mathbb{E}[Y'|\mathscr{H}i - 1] = 0$, so that $\sum_{i=1}^n \mathbb{E}[Y_i'|\mathscr{H}_{i-1}] \leq a + b$ holds deterministically. Similarly, if we observe $\sum_{i=1}^t \mathbb{E}[Y_i|\mathscr{H}_{i-1}] < a - b - (n - t)$ then we set $Y_t' = 1$ (and we will do so for all future times) and obtain $\sum_{i=1}^n \mathbb{E}[Y_i'|\mathscr{H}_{i-1}] \geq a - b$ deterministically.

The bounds of Theorem 1.4 apply to X', and by definition if $X \neq X'$ then $\mathcal{E}$ has not occurred (if the sum of conditional expectations has already exceeded $a + b$ at some time t it certainly does at time n, similarly if at time t it is below $a - b - (n - t)$ then it will be below $a - b$ at time n), giving the required statement. □

Exercise 1. The above proof does not quite work. Find why and fix it.

1.5 With high probability

We are always going to be talking about things which hold 'with high probability' (*whp*), i.e. with probability tending to 1 as some parameter n tends to infinity. In all cases, we will actually have some fairly rapid convergence—the probability will be $1 - O(n^{-C})$ where C is some large absolute constant, or better. We will usually want a polynomially in n (i.e. n^c for some constant c) large collection of 'good events' to hold simultaneously. The *union bound* says the probability any one of these good events fails is at most the sum of their failure probabilities. For us, the polynomial union bound will always be small enough that the whp 'wins', i.e. $c < C$ and so whp all the good events hold. There will be a few places where we need to be slightly careful about union bounds and summing probabilities, and then we will

do it explicitly, but usually all our union bounds will work with a huge margin of error and we will not do the calculation explicitly.

1.6 Tracking a process

I'm going to repeat this in the tutorial (with a different example) but if the material above was new to you, it might be a good idea to try to get comfortable with this example.

Everything we do is going to fit into a rather simple scheme. We will write down some random process. Then we will write down some events which we expect to occur at time t in this process, for each t. These will be things like 'every vertex has degree roughly ...' and 'the image of our embedding covers about ... fraction of every pair neighbourhood'. That is, we will write down (usually!) polynomially many statements. We will want to prove that it is likely that all our statements hold for every time t, because that will (somehow!) imply the theorem we wanted to prove. Proving it is 'tracking the process'.

To do this, we will always consider the following *failure event*, for a given statement at time t. All our statements hold for all times $t' < t$ (we can track the process to time $t-1$), but the given one fails at time t.

If we can prove that this event is sufficiently unlikely (probability at most n^{-C} for some large enough C) then the union bound will tell us that it is unlikely that any failure event occurs, and in particular that means that all our statements hold for every time t.

Finally, the 'sufficiently unlikely' is always going to come from applying Theorem 1.5. To apply that theorem, we need to have several things available: an event $\mathcal{E}$, a sequence of histories and a sequence of random variables, and some way to calculate sums of conditional expectations.

Let's revisit the example we gave earlier of embedding a graph G into the n-vertex graph H (somehow) randomly. To avoid complicated formulae, I'll pretend that our statements are exact (i.e. error terms are zero). Suppose that H is an n-vertex graph in which every vertex has degree pn (for some $p \in (0,1)$, for example $p = \frac{1}{2}$) and every pair of vertices has $p^2 n$ common neighbours; for example a random graph $G_{n,p}$ would be likely to have (roughly!) these properties. And suppose that G is the $\frac{n}{2}$-vertex path, with vertices labelled $1, \dots, \frac{n}{2}$ along the path. Finally, suppose that at each time t we will embed vertex t of G to a uniform random neighbour of where we embedded $t-1$ which we did not previously use (the first vertex we just embed uniformly at random).

A reasonable guess is that the t vertices embedded by time t look like a uniform random set of t vertices. So we hope that they cover a $\frac{t}{n}$-fraction of every vertex neighbourhood (i.e. pt vertices) and every pair neighbourhood (i.e. $p^2 t$ vertices). This is a collection of $2n + 2\binom{n}{2}$ statements for each time t (upper and lower bounds on the number of vertices used in each vertex or pair neighbourhood), and these are the statements we will track.

Let's fix a vertex v and a time t, and look at the statement that at most tp vertices in $N_H(v)$ are used by time t. The corresponding failure event is that more than $(1+\delta)pt$ vertices were actually used (just once, we'll put in the error term!) and all our statements hold for times $t' < t$.

This failure event is the same as saying that $Y_1 + \cdots + Y_t > (1+\delta)pt$, where Y_i is the

Bernoulli random variable from earlier: Y_i is equal to 1 if the ith vertex of G is embedded to $N_H(v)$. So now we have some idea how Theorem 1.5 and the probability of our failure event relate. We can let $\mathcal{E}$ be the event 'all our statements hold for times $t' < t$', and the probability of the failure event is

$$\mathbb{P}\big[Y_1 + \cdots + Y_t > (1+\delta)pt \text{ and } \mathcal{E} \text{ occurs}\big],$$

which is exactly the kind of probability Theorem 1.5 can be used to bound. It's easy to sort out what $\mathscr{H}_i$ should be: this is the history of all the random choices up to and including the embedding of the ith vertex of G. So now we defined all the terms we need to apply Theorem 1.5. We just need to prove bounds on the sum of conditional expectations which hold whenever $\mathcal{E}$ occurs. We'd like to show this sum is equal to pt, and then Theorem 1.5 will tell us this particular failure event is unlikely.

Why might that occur? The simplest reason is: because each of the t summands is equal to p. How can we estimate $\mathbb{E}[Y_i|\mathscr{H}_{i-1}]$? This is the same as saying: we know how the first $i-1$ vertices of G are embedded, in particular we know the vertex v' to which the $(i-1)$st vertex was embedded, and we want to know how likely we are to embed the ith vertex into $N_H(v)$. Let's suppose that v' is not v (cheating, but not much). We can assume all our statements hold at time $i-1$, because $i-1 < i \leq t$ and if any of our statements before time t failed then $\mathcal{E}$ doesn't occur. (Note, this is *not* conditioning on $\mathcal{E}$..!)

In particular, we will embed the ith vertex of G to a uniform random neighbour of v' which is not used by one of the first $i-1$ neighbours. Our statements tell us there are $p(n-i+1)$ such vertices. Of these, $p^2(n-i+1)$ are in $N_H(v,v')$ so also neighbours of v (the pair neighbourhood statement). So the chance we pick one of those is precisely $\frac{p^2(n-i+1)}{p(n-i+1)} = p$, which is the estimate we wanted: we're done.

It's also worth saying explicitly (one possible choice of) what the probability space Ω could be for this example. We can let it be

$$\big\{(v_1, \ldots, v_{n/2}) : \text{ the } v_i \text{ are distinct vertices of } H\big\}$$

i.e. an element of Ω is a list of $\frac{n}{2}$ distinct vertices of H, the list of vertices in order we chose to embed the $\frac{n}{2}$ vertices of G to. We should notice that some elements of this probability space do have probability zero: for example, those where v_1v_2 is not actually an edge of H. We could remove those elements without changing anything, but it's easier not to bother. A construction like this will work to supply Ω for any (finite) process, in particular all the ones we'll consider. But you do not actually need to know or care what Ω is; we will never need to talk about specific elements of Ω.

On the one hand, I've now said most of the idea of the whole tutorial. On the other hand, this sketch left a lot out, and filling that in will take the tutorial.

We will have error terms in all our statements and we need to figure out how to handle them (which is mainly routine though long calculation).

We need to figure out what the 'right' collection of statements to write down is: in this example, we cheated, because while one can make the calculation above more or less rigorous, at some point we need to write down a failure condition for the statement about pair neighbourhoods, and in order to handle that (at least by generalising the above idea)

we'd need to know something about triple neighbourhoods. And to handle that we'd need to know about 4-vertex neighbourhoods, and so on... which is not going to work. We'll see at the end of the tutorial how to un-cheat this (the statements we wrote do work, just we need to be more clever in calculating sums of conditional expectations). This is something nontrivial.

Finally, we need to know that certain perfectly reasonable looking ways to 'be more clever' will turn out, after doing all the calculations, not to work.

2 Processes we can analyse crudely

When we analyse a random process, errors are unavoidable. We cannot hope that every quantity we are interested in is always exactly its expectation, it will only be close to its expectation—in general these errors can affect our estimates of expectations later in the process, and we end up needing to argue that the errors do not grow too fast.

The simplest way to argue such a thing is to put an absolute bound on the errors: no quantity we care about can go outside a fixed interval (not changing over time) at any point in the process. Sometimes this is good enough to push an analysis through. In this section, we will illustrate the idea by proving (a simplified version of) the Blow-up Lemma. This is originally a result of Komlós, Sárközy and Szemerédi, though there are by now several other proofs. It turns out to be very useful in extremal graph theory, in combination with the Szemerédi Regularity Lemma—but we are not going to explain how to use it here.

2.1 Approximate Blow-up Lemma

The simplified statement we want to prove is roughly the following. Suppose that G is a tripartite graph with parts X_1, X_2, X_3 each of size n and $\Delta(G) \leq \Delta$. Suppose that H is a tripartite graph with parts V_1, V_2, V_3 each of size n (which we call *clusters*), and each pair (V_i, V_j) induces a *quasirandom* bipartite graph with dn^2 edges, where d is a constant (and n is very large). Then G is a subgraph of H.

We will first prove a simplified version of this, where '(A, B) is quasirandom' means (ε, d)-*regular*, that is, for any $A' \subseteq A$ and $B' \subseteq B$ with $|A'|, |B'| \geq \varepsilon n$ we have $e(A', B') \geq (d-\varepsilon)|A'||B'|$. For this statement to be true, we need to assume that there are some isolated vertices in each part of G. We will then explain (briefly!) how to use an extra condition: 'each $u \in V_i$ has at least $(d-\varepsilon)n$ neighbours in V_j' to drop the isolated vertex assumption. To go from this to a full Blow-up Lemma is really just bookkeeping and notation; no new ideas are needed.

Theorem 2.1 (Almost-Blow-up Lemma)**.** *Given $\Delta \in \mathbb{N}$ and $\mu, d > 0$, there exist $\varepsilon > 0$ and n_0 such that the following holds. Let G be any tripartite graph with parts X_1, X_2, X_3 of size n, each of which contains at least $5\mu n$ isolated vertices of G, such that $\Delta(G) \leq \Delta$. Let H be any tripartite graph with parts V_1, V_2, V_3 of size n, such that (V_i, V_j) is (ε, d)-regular for each $i \neq j$. Then G is a subgraph of H.*

The basic idea here is simple. We put an order $V(G) = \{x_1, \dots, x_{3n}\}$ on the vertices of G, such that the isolated vertices come at the end of the order. We embed the vertices of G one-by-one into H; we insist that the vertices of X_i are embedded to V_i for each i. We let ψ_0 be the empty embedding of no vertices, and for each i we let ψ_i be the embedding of the first i vertices we get, so ψ_i differs from ψ_{i-1} in that we choose an image for x_i.

We perform each embedding randomly, subject to maintaining a valid embedding: that is, when we embed $x_i \in V(G)$, and we have already embedded the vertices $N^-(x_i)$ (i.e. the neighbours of x_i that come before i in the order), we need to embed x_i to a vertex of H that is a *candidate for* x_i, i.e. in the common neighbourhood of $\psi_{i-1}\big(N^-(x_i)\big)$. This word 'candidate' just means that if we map x to something in its candidate set, we are not

immediately breaking the 'edges go to edges' condition of an embedding. Critically, it is enough to argue that we can do this for each $1 \leq i \leq (3-15\mu)n$: the last $15\mu n$ vertices are isolated vertices and we can embed them as we like.

As written, this idea will not quite work. There are two problems.

One is the following: suppose x_1, x_2, x_3 forms a triangle in G. If we embed x_1 randomly, and then x_2 randomly, it is possible that we will choose an image for x_2 which has no, or very few, neighbours in $N_H\big(\psi_1(x_1)\big)$. Then we will get stuck when we want to embed x_3 which has to go to this nonexistent or tiny common neighbourhood. In order to avoid this, we will (deterministically) avoid *bad* vertices, i.e. vertices which are candidates for x_i but which make the candidate set for some future x_j too small. We will see that there are always very few bad vertices, compared to the size of the candidate set, which means this doesn't affect the analysis much.

The second problem is that we are not allowed to re-use vertices. We might find that when we come to time t, the vertex x_t has a reasonably large candidate set, but most or all of them have been embedded to already (i.e. they are in $\operatorname{im}\psi_{t-1}$ and so they are not *available*. When we see such a vertex, we call it a *queue vertex*. In order to deal with this, we will set aside, randomly, before beginning the embedding a small *queue reservoir* in each V_i (of size something like μn) which we will use exclusively to embed queue vertices. We will see that whp there are very few (less than ρn, where $\rho \ll \mu$) queue vertices and this means we will not run into the problem of queue vertices having few available candidates in the queue reservoir; we just won't have enough queue vertices to fill up candidate sets in the queue reservoir.

In the following proof, we will put in an extra *buffer reservoir* and say a couple of things about it, which we will not use in this proof; they are needed to prove the stronger version. We will use red for these 'extra' things.

Proof of Theorem 2.1. Given $\Delta \in \mathbb{N}$ and $d > 0$, without loss of generality we can assume μ is not too large. Given μ, we choose $\rho \gg \varepsilon \gg n_0^{-1} > 0$.

Exercise 2. What 'not too large' requirement do we need for the following proof? Find explicit formulae for ρ, ε and n_0 in terms of Δ, d and μ which make the following proof work. *Hint: work through the parameters in order of decreasing size, and do* not *try to optimise your choices!*

Fix G and H as in the theorem statement. We begin by picking in each V_i two disjoint sets of size μn, which we call V_i^{q} (the queue reservoir) and V_i^{buf} (the buffer reservoir), and we let $V_i^{\mathrm{bulk}} = V_i \setminus (V_i^{\mathrm{q}} \cup V_i^{\mathrm{buf}})$ (the bulk).

We set ψ_0 to be the empty embedding. For each i and $x \in X_i$ we set $\mathcal{C}_0(x) = V_i$ (the candidate set). We will now define embeddings $\psi_1, \dots, \psi_{3n-15\mu n}$ sequentially. In each case, ψ_t embeds $x_1, \dots, x_t$. Once we have defined ψ_t, we update the candidate sets: given $y > t$, we set

$$\mathcal{C}_t(y) = \begin{cases} \mathcal{C}_{t-1}(y) & \text{if } x_t y \notin E(G) \\ \mathcal{C}_{t-1}(y) \cap N_H\big(\psi_t(x_t)\big) & \text{if } x_t y \in E(G) \end{cases}.$$

We will never be interested in candidate sets of vertices in $\operatorname{im}\psi_t$ and do not define them. (This definition is identical to the one in the sketch above!)

For any given t and $x \in V(G)$, we define $\mathcal{A}_t(x) = \mathcal{C}_t(x) \setminus \operatorname{im} \psi_t$ (the available candidate set). Superscripts indicate intersection with a reservoir or the bulk, so e.g. $\mathcal{A}_t^{\mathrm{q}}(x) = \mathcal{A}_t(x) \cap V_i^{\mathrm{q}}$ where $x \in X_i$. Finally, we define B_t (the bad set at time t) as follows. Suppose i is such that $X_t \in X_i$, and that $v \in V_i$. If there is a neighbour y of x_t which is not in $\operatorname{im} \psi_t$ and

$$|\mathcal{C}_{t-1}^{\mathrm{bulk}}(y) \cap N_H(v)| < (d-\varepsilon)|\mathcal{C}_{t-1}^{\mathrm{bulk}}(y)|$$

then we put $v \in B_t$. If the same statement holds replacing $\mathcal{C}^{\mathrm{bulk}}$ with $\mathcal{C}^{\mathrm{q}}$ or $\mathcal{C}^{\mathrm{buf}}$, we put $v \in B_t$. Otherwise, we do not put v to B_t.

Our random process is now as follows. At each time $1 \le t \le (3-15\mu)n$ in succession:
If $|\mathcal{A}_{t-1}^{\mathrm{bulk}}(x_t)| \ge \frac{1}{2}d^{\Delta}\mu n$, we choose v uniformly at random from $\mathcal{A}_{t-1}^{\mathrm{bulk}}(x_t) \setminus B_t$.
If $|\mathcal{A}_{t-1}^{\mathrm{bulk}}(x_t)| < \frac{1}{2}d^{\Delta}\mu n$, we choose v uniformly at random from $\mathcal{A}_{t-1}^{\mathrm{q}}(x_t) \setminus B_t$.
Either way, we let $\psi_t = \psi_{t-1} \cup \{x_t \hookrightarrow v\}$. In the latter case, we say x_t is *put in the queue*.

It is possible that we attempt to choose a vertex uniformly at random from $\mathcal{A}_{t-1}^{\mathrm{q}}(x_t) \setminus B_t$ and this set turns out to be small or empty. To avoid this, if $\mathcal{A}_{t-1}^{\mathrm{q}}(x_t)$ has size less than $\frac{1}{2}\mu d^{\Delta} n$ we say the process *fails* and stop immediately.

Observe that we have the following (deterministic) property for any t such that the process has not failed and any unembedded y in $V(G)$ such that y has s embedded neighbours.

$$\left|\mathcal{C}_t^{\mathrm{bulk}}(y)\right| \ge (1-2\mu)(d-2\varepsilon)^s n \quad \text{and} \quad \left|\mathcal{C}_t^{\mathrm{q}}(y)\right|, \left|\mathcal{C}_t^{\mathrm{buf}}(y)\right| \ge \mu(d-2\varepsilon)^s n\,. \tag{1}$$

This is true by induction on t. For $t = 0$ we have $s = 0$ and $\mathcal{C}_t^{\mathrm{q}}(y) = V_i^{\mathrm{q}}$ has size μn, where $y \in X_i$. If $t \ge 1$, then either x_t is not a neighbour of y and neither side of (1) changes, or x_t is a neighbour of y and by definition of B_t we obtain (1).

We are now in a position to bound $|B_t|$. Fix a neighbour y of x_t. Suppose $x_t \in X_i$ and $y \in X_j$. Consider the set S of vertices v in V_i such that $|\mathcal{C}_t^{\mathrm{q}}(y) \cap N_H(v)| < (d-\varepsilon)|\mathcal{C}_t^{\mathrm{q}}(y)|$. By definition, $e\big(S, \mathcal{C}_t^{\mathrm{q}}(y)\big) < (d-\varepsilon)|S||\mathcal{C}_t^{\mathrm{q}}(y)|$. By (1) and choice of ε, we have $|\mathcal{C}_t^{\mathrm{q}}(y)| \ge \varepsilon n$. So if $|S| \ge \varepsilon n$, we have a contradiction to (ε, d)-regularity of (V_i, V_j), and we conclude $|S| < \varepsilon n$. The same argument works replacing $\mathcal{C}^{\mathrm{q}}$ with $\mathcal{C}^{\mathrm{buf}}$ or $\mathcal{C}^{\mathrm{bulk}}$, and there are at most Δ choices for y. We conclude $|B_t| \le 3\Delta\varepsilon n$.

Before performing our probabilistic analysis, we make a final deterministic conclusion. At time t, provided the process has not failed, when we choose v to which we will embed x_t, we do so uniformly at random from a set of size at least $\frac{1}{4}\mu d^{\Delta} n$. If we choose from $\mathcal{A}_{t-1}^{\mathrm{bulk}}(x_t) \setminus B_t$, then $\left|\mathcal{A}_{t-1}^{\mathrm{bulk}}(x_t)\right| \ge \frac{1}{2}\mu d^{\Delta} n$, while if not then since the process has not failed we have $\left|\mathcal{A}_{t-1}^{\mathrm{q}}(x_t)\right| \ge \frac{1}{2}\mu d^{\Delta} n$. Either way, when we exclude B_t, by choice of ε we still have at least $\frac{1}{4}\mu d^{\Delta} n$ vertices to choose from.

We only need one piece of probabilistic analysis to complete the proof now.

Claim 2.1. *With high probability the following holds. For each i and $W \subseteq V_i$ with $|W| \ge \rho n$, there are at most $10\Delta\rho n$ poor vertices for W, i.e. vertices $y \in X_i$ such that there is a time t at which the process has not yet failed, y is not embedded and has s embedded neighbours, and we have $\left|\mathcal{C}_t(y) \cap W\right| < (d-\varepsilon)^s |W|$.*

Proof. Fix i and a set $W \subseteq V_i$ with $|W| \ge \rho n$.

For each time t, let Y_t be a Bernoulli random variable defined as follows. If there exists an unembedded neighbour z of x_t, which has s neighbours in $\{x_1, \dots, x_{t-1}\}$, such that

$$\left|\mathcal{C}_{t-1}(z) \cap W\right| \ge (d-\varepsilon)^s |W| \quad \text{and} \quad \left|\mathcal{C}_t(z) \cap W\right| < (d-\varepsilon)^{s+1} |W|$$

then we set $Y_t = 1$. Otherwise (including if the process has failed at or before time t) we set $Y_t = 0$.

If we run the process and observe that a given y is poor for W, then there is a first time t witnessing y is poor for W. Necessarily we have $t \geq 1$ and x_t is a neighbour of y, and y witnesses that $Y_t = 1$. It follows that the number of poor vertices for W is at most $\Delta \sum_{t=1}^{3n-15\mu n} Y_t$.

In order to prove the claim, it is therefore enough to argue that $\sum_{t=1}^{3n-15\mu n} Y_t > 10\rho n$ is a very unlikely event. The main job here is to upper bound $\mathbb{P}[Y_t = 1|\mathscr{H}_{t-1}]$ where $\mathscr{H}_{t-1}$ is the history up to and including embedding the $t-1$st vertex. If $\mathscr{H}_{t-1}$ is such that the process fails at or before time t (failing at time t is determined by the first $t-1$ random choices!) then the desired probability is zero and there is nothing to prove. So suppose this is not the case.

We select an image v for x_t uniformly at random from a set of size at least $\frac{1}{4}\mu d^{\Delta} n$ within V_i, where $x_t \in X_i$. If this choice of v causes $Y_t = 1$, then by definition there is a neighbour z of x_t such that $Q = \mathcal{C}_{t-1}(z) \cap W$ is has size at least $(d-\varepsilon)^s|W|$ but v has less than $(d-\varepsilon)Q|$ neighbours in Q. Note that $Q \subseteq V_j$ where $z \in X_j$, and $j \neq i$. Let T_z be the set of vertices $v \in V_i$ with less than $(d-\varepsilon)^{s+1}|W|$ neighbours in Q. Then $e(T_z, Q) < (d-\varepsilon)|T_z||Q|$ by definition. Since (V_i, V_j) is (ε, d)-regular and $|Q| \geq \varepsilon n$, we conclude $|T_z| < \varepsilon n$. In other words, if $Y_t = 1$ then we chose v from $\bigcup_{z \in N_G(x_t)} T_z$ which has size at most $\Delta\varepsilon n$. The conditional probability of this event is

$$\mathbb{P}[Y_t = 1|\mathscr{H}_{t-1}] \leq \frac{\Delta\varepsilon n}{\frac{1}{4}\mu d^{\Delta} n} = 4\Delta\varepsilon\mu^{-1}d^{-\Delta}\,,$$

where the final inequality is by choice of ε. Summing up, we have

$$\sum_{t=1}^{3n-15\mu n} \mathbb{P}[Y_t = 1|\mathscr{H}_{t-1}] \leq 4\Delta\varepsilon\mu^{-1}d^{-\Delta} \cdot 3n = 12\Delta\varepsilon\mu^{-1}d^{-\Delta}n\,.$$

By choice of ε, applying the third part of Theorem 1.5, with $\mathcal{E}$ the sure event (i.e. $\mathcal{E} = \Omega$) and $R = 1$, we conclude

$$\mathbb{P}\Big[\sum_{t=1}^{3n-15\mu n} Y_t \geq \rho n\Big] \leq \exp(-n)\,.$$

Note that here we are using the fact that ε is tiny compared to ρ in order to apply Theorem 1.5 with s rather large; the probability we get from Theorem 1.5 is (a bit better than) $\exp\big(-\frac{1}{4}\rho n \log s\big)$ and so we need to choose s large enough (i.e. ε small enough) that $\log s > 4\rho^{-1}$.

Now, taking the union bound over the less than $3 \cdot 2^n$ choices of i and W, we obtain a probability $1 - 3 \cdot 2^n \exp(-n)$, which is the desired whp. ∎

We apply Claim 2.1 to show that the process whp does not fail. Indeed, suppose the process does fail at some time t. Let i be such that $x_t \in X_i$. Since the process fails, we have $|\mathcal{A}^{\mathrm{q}}_t(x_t)| < \frac{1}{2}\mu d^{\Delta} n$. By (1) however we have $|\mathcal{C}^{\mathrm{q}}_t(x_t)| \geq \mu(d-2\varepsilon)^{\Delta} n > \frac{1}{2}\mu d^{\Delta} n + 10\Delta\rho n$. The difference between these two sets is contained in $\operatorname{im}\psi_t$, so in particular more than $10\Delta\rho n$ vertices of X_i were put in the queue.

Now suppose the likely event of Claim 2.1 holds, and apply it with $W = V_i^{\text{bulk}} \setminus \operatorname{im} \psi_{t-1}$. Since $\operatorname{im} \psi_{t-1}$ contains at most $(1-5\mu)n$ vertices of V_i^{bulk} (because this is the total number of vertices our random process embeds to V_i) and V_i^{bulk} has size $(1-2\mu)n$, we see $|W| \geq \mu n \geq \rho n$ so Claim 2.1 applies. Claim 2.1 says that at most $10\Delta\rho n$ vertices of X_i are poor for W, and in particular there is some $x_{t'} \in X_i$ with $t' < t$ such that $x_{t'}$ was put in the queue and $x_{t'}$ is not poor for W.

But $x_{t'}$ being put in the queue means that $\mathcal{A}^{\text{bulk}}_{t'-1}(x_{t'}) = \mathcal{C}_{t'-1}(x_{t'}) \cap (V_i^{\text{bulk}} \setminus \operatorname{im} \psi_{t'-1})$ has size less than $\frac{1}{2}\mu d^\Delta n$. On the other hand, since $\operatorname{im} \psi_{t'-1} \subseteq \operatorname{im} \psi_{t-1}$, we see

$$\mathcal{C}_{t'-1}(x_{t'}) \cap (V_i^{\text{bulk}} \setminus \operatorname{im} \psi_{t'-1}) \supseteq \mathcal{C}_{t'-1}(x_{t'}) \cap W\,.$$

That means that time $t'-1$ witnesses that $x_{t'}$ is poor for W, because $\frac{1}{2}\mu d^\Delta n < (d-\varepsilon)^\Delta |W|$ by choice of n and since $|W| \geq \mu n$. This contradiction shows that the process whp does not fail.

Suppose that the process does not fail, and $\psi_{3n-15\mu n}$ is an embedding of the first $3n-15\mu n$ vertices of G to H. We complete the embedding greedily, choosing for each unembedded vertex of X_i an unused vertex of V_i to embed it to, for each i. We can do this because the unembedded vertices of X_i are isolated. □

2.2 Blow-up Lemma

We obviously cannot simply remove the condition of Theorem 2.1 about isolated vertices in G; it could be that H has (a few) isolated vertices. However if we rule this bad example out by adding a minimum degree condition, we get the following.

Theorem 2.2 (Blow-up Lemma)**.** *Given $\Delta \in \mathbb{N}$ and $\mu, d > 0$, there exist $\varepsilon > 0$ and n_0 such that the following holds. Let G be any tripartite graph with parts X_1, X_2, X_3 of size n such that $\Delta(G) \leq \Delta$. Let H be any tripartite graph with parts V_1, V_2, V_3 of size n, such that (V_i, V_j) is (ε, d)-regular and each $u \in V_i$ has at least $(d-\varepsilon)n$ neighbours in V_j, for each $i \neq j$. Then G is a subgraph of H.*

Maybe the first approach you would think of to prove this is simply to try to argue that we can use the process we used in Theorem 2.1, but (somehow) strengthen the analysis so that it will whp manage to embed all of G into H rather than stopping with some vertices left to cover. Unfortunately this does not work, even if we make unrealistically strong (i.e. that we would not be able to obtain in most applications) assumptions on the quasirandomness ε.

Exercise 3. Suppose $d = \frac{1}{10}$ and $\varepsilon = n^{-1/10}$. Let H be a typical random tripartite graph with edge probability d (which whp satisfies the conditions of Theorem 2.2 with these parameters). Suppose G consists of n disjoint copies of K_3 and the order on $V(G)$ puts these consecutively, with the vertices of X_1 first. Argue that if we run the process in the proof of Theorem 2.1 with any choice of parameters (as opposed to the choices in the proof) then the process *does not* whp succeed. *Hint: First show that whp $B_t = \emptyset$ at every step, and consider which edges of H at V_1 need to be revealed in each step of the process.*

What we instead do is follow the proof of Theorem 2.1 but find a way to guarantee we can deterministically embed the final $15\mu n$ vertices, even though they are not isolated. The

key observation is that we can ensure these last vertices form an independent set, which reduces the problem to finding perfect matchings in auxiliary bipartite graphs. It turns out to be convenient to ask for these final $15n$ vertices to be far apart: at least distance 4 in G. We can find such vertices greedily.

Proof of Theorem 2.2.

Exercise 4. Write down explicit constant choices given Δ and d which make the following proof work.

Suppose G and H satisfying the conditions of Theorem 2.2 are given. We choose reservoirs as in the proof of Theorem 2.1.

We need to be more careful to choose an order for $V(G)$. We do this as follows. For each i, we choose $\pi_i \in \{0, 1, \dots, \Delta\}$ such that there are at least $\frac{n}{\Delta+1}$ vertices in X_i of degree π_i. We choose greedily $5\mu n$ vertices in X_i of degree π_i for each i, which we denote by X_i^{buf} (the *buffer vertices*), such that no two chosen vertices are at distance 3 or less in G. Observe that when we choose a vertex we exclude at most $1 + \Delta + \Delta^2 + \Delta^3 < 4\Delta^3$ vertices from being chosen in future steps; by choice of μ the greedy choice therefore succeeds.

We now order $V(G)$ according to the following restrictions. The final $15\mu n$ vertices are the vertices $\bigcup_i X_i^{\text{buf}}$. The vertices $N_G(x)$ appear consecutively in the order for each i and $x \in X_i^{\text{buf}}$. The neighbours of buffer vertices all come before any other vertices of G in the order. It is easy to see this is possible, and by choice of μ the neighbours of the buffer vertices all appear in the first $\frac{1}{8}d^{\Delta+1}n$ vertices of the order. This has the following nice consequence: the neighbours of buffer vertices (deterministically) do not enter the queue. To see that this is true, observe that by (1) for any vertex x_t of G with s neighbours preceding it in the order we have

$$|\mathcal{C}_{t-1}^{\text{bulk}}(x_t)| \geq (1 - 2\mu)(d - 2\varepsilon)^s n \geq \tfrac{3}{4}(d - 2\varepsilon)^\Delta n$$

and for the first $\frac{1}{8}d^\Delta n$ vertices of the order, we therefore have

$$|\mathcal{A}_{t-1}^{\text{bulk}}(x_t)| \geq \tfrac{3}{4}(d - 2\varepsilon)^\Delta n - \tfrac{1}{8}d^\Delta n > \tfrac{1}{2}d^\Delta \mu n\,.$$

We now run the process described in the proof of Theorem 2.1. All the analysis of that proof remains valid. In particular Claim 2.1 holds.

We need one further piece of probabilistic analysis.

Claim 2.2. *With high probability, either the process fails or for every i and $v \in V_i$, there are at least $2\rho n$ vertices $x \in X_i^{\text{buf}}$ such that $v \in \mathcal{C}_{3n-15\mu n}(x)$.*

Proof. Fix i and $v \in V_i$.

Observe that $v \in \mathcal{C}_{3n-15\mu n}(x)$ occurs if and only if $N_G(x)$ is embedded to $N_H(v)$. We enumerate $X_i^{\text{buf}} = \{y_1, \dots, y_{5\mu n}\}$ according to the order in which their neighbours appear in $V(G)$. Let $\mathscr{H}_t$ denote the random choices made in the process up to and including the embedding of $N_G(y_t)$ for each $1 \leq t \leq 5\mu n$.

The choice to put these neighbourhoods as intervals in the order on $V(G)$ makes this well-defined, and means that in order to apply Theorem 1.5 to estimate the whp number of t such that $v \in \mathcal{C}_{3n-15\mu n}(y_t)$, we just need to find a lower bound on

$$\mathbb{P}\big[\psi_{3n-15\mu n}\big(N_G(y_t)\big) \in N_H(v)\big|\mathscr{H}_{t-1}\big]$$

for each t.

This is the same thing as describing a (large) collection of ways that the process can successively embed the vertices $N_G(y_t)$ into $N_H(v)$; we *don't* need to consider all the ways the process can do this. In particular, if we can show that each vertex of $N_G(y_t)$ has at least $\frac{1}{4}d^{\Delta+1}n$ 'nice' choices for its embedding (i.e. in $N_H(v)$ and allowing for nice choices for later vertices) then we can recall that each vertex of $N_G(y_t)$ is embedded uniformly at random into a set of size at most n to put a lower bound on the desired probability.

Exercise 5. Using the ideas seen before, prove the lower bound

$$\mathbb{P}\big[\psi_{3n-15\mu n}\big(N_G(y_t)\big) \in N_H(v)\big|\mathscr{H}_{t-1}\big] \geq \big(\tfrac{1}{4}d^{\Delta+1}\big)^{\pi_i}.$$

Hint: By choice of μ the vertex v necessarily has close to dn neighbours in V_j^{bulk} for each $j \neq i$.

We can now apply Theorem 1.5, with the sum of conditional probabilities at least $5\mu n \cdot \big(\frac{1}{4}d^{\Delta+1}\big)^{\pi_i} > 4\rho n$, to conclude that with probability at least $1 - \exp\big(-\frac{1}{3}\rho n\big)$ there are at least $2\rho n$ vertices $x \in X_i^{\text{buf}}$ such that $v \in \mathcal{C}_{3n-15\mu n}(x)$. Taking a union bound over the polynomially many choices of v and i the Claim follows. ∎

We now suppose that the high probability events of both Claims 2.1 and 2.2 occur. As in the proof of Theorem 2.1, it follows that the process succeeds in constructing $\psi := \psi_{3n-15\mu n}$, i.e. it embeds all but the buffer vertices of G. For convenience, from this point we drop the subscript $3n - 15\mu n$; $\mathcal{C}$ means $\mathcal{C}_{3n-15\mu n}$.

Given i, we now draw an auxiliary bipartite graph F_i whose parts are X_i^{buf} and $V_i \setminus \operatorname{im}\psi$. By construction, these two sets each have $5\mu n$ vertices. We put an edge xu, where $x \in X_i^{\text{buf}}$ and $u \in V_i \setminus \operatorname{im}\psi$, if and only if $u \in \mathcal{C}(x)$. What this means is: if we extend ψ by embedding x to u where xu is in F_i, then we get a partial embedding (we do not map edges of G to nonedges of H). It follows that a perfect matching in F_i gives us a way to embed all of X_i^{buf}, and so a perfect matching in each F_i gives us a way to extend ψ to an embedding of G into H: we embed each x to its matching partner. What remains is to show these matchings exist.

We check Hall's condition. Let S be a non-empty subset of X_i^{buf}, and let

$$T = \{u \in V(F_i) : xu \in F_i \text{ for some } x \in S\}$$

be the joint neighbourhood of S.

Choosing any one $x \in S$, because no vertices have been embedded to V_i^{buf}, we have $\mathcal{C}^{\text{buf}}(x) \subseteq T$, and by (1) we therefore have $|T| \geq \frac{1}{2}\mu d^{\Delta}n$. This is enough to verify Hall's condition if $|S| < \frac{1}{2}\mu d^{\Delta}n$.

If on the other hand $|S| > 5\mu n - \rho n$, then by Claim 2.2 for every $v \in V_i$ there is a vertex x of S such that $v \in \mathcal{C}(x)$. Thus $T = V_i \setminus \operatorname{im}\psi$ has size $5\mu n$, which verifies Hall's condition in this case also.

What remains is the case $\frac{1}{2}\mu d^{\Delta}n \leq |S| \leq 5\mu n - \rho n$. If Hall's condition fails in this case, then $|T| < 5\mu n - \rho n$. Let $\bar{T} = V_i \setminus (\operatorname{im}\psi \cup T)$. By construction we have $|\bar{T}| \geq \rho n$, and for every $x \in S$ we have $\mathcal{C}(x) \cap \bar{T} = \emptyset$. But in particular this says that every $x \in S$ is poor for $\bar{T}$, and Claim 2.1 says that $|S| \leq \rho n$. By choice of ρ, this is a contradiction.

Thus Hall's condition holds for each F_i, and so there is an embedding of G into H. $\square$

What you are supposed to notice from the above two proofs is that we used quite 'crude' estimates. We did not try to estimate accurately how many buffer vertices will end up having any given $v \in V_i$ as candidate. We did not try to estimate accurately how large $\mathcal{A}_t^{\text{bulk}}(y)$ is at any given time t, rather we found a way to argue that it can only rarely be small and designed a 'failsafe', the queue system, to deal with the consequences if it does get small.

After reading the following section, you will probably realise that we could have done both of these things. In particular, the 'reservoirs' we used aren't really necessary, even though it is necessary (as shown by Exercise 3) to do something 'special' to complete the embedding.

Exercise 6. Modify the process of Theorems 2.1 and 2.2 by removing the queue and buffer reservoir (i.e. set $V_i^{\text{bulk}} = V_i$ for each i and always embed x_t to $\mathcal{A}_{t-1}^{\text{bulk}}(x_t) \setminus B_t$ no matter how big this set is). Prove that the modified process still whp succeeds and deduce both theorems.
Hint: Think carefully about where you will need 'accurate' analysis and where a crude analysis will be good enough, otherwise you will have a lot of unnecessary work to do. Before you start trying to estimate error terms, check your approach is strong enough to obtain the right main terms! This exercise is quite hard.

To be 'useful', what we really want as a Blow-up Lemma is a bit more.

We would like to have 'image restrictions'; for a few (say ρn) vertices x in each X_i we can insist that they are mapped not just to some vertex in V_i, but to a subset $I(x)$ of those vertices. We need to insist on something like $|I(x)| \geq \mu n$. This is not too hard to incorporate to the proof above; the easiest way is to ensure the queue reservoir contains a decent fraction of $I(x)$ for each x (which will occur if it's chosen randomly) and adjust the constant choices.

We would also like to have a setup more compatible with the Szemerédi Regularity Lemma. We can have lots of clusters. These clusters might not all be the same size (though we can bound the ratio of sizes). Some pairs of clusters (V_i, V_j) might not be (ε, d)-regular. This is fine provided we don't have any edges between X_i and X_j in G. We need to be a bit careful about how the number of clusters compares with the regularity parameter ε: the former is generally huge compared to the latter, and we will not be able to get the minimum degree condition $\big|N(v, V_j)\big| \geq (d - \varepsilon)|V_j|$ for all $v \in V_i$, even when we do know (V_i, V_j) is (ε, d)-regular. We will only be able to get such a condition for a 'few' (say Δ-many) clusters V_j for any given V_i. Again, one can formulate some condition which is both useable and provable; it is not really harder to prove than the version above.

3 Processes with growing errors

Sometimes, a crude analysis is not good enough. Looking to Exercise 6, if you were asked to guess how large $\mathcal{A}_t(y)$ 'should' be for a given $y \in X_i$ (which is not embedded at time t but which has s embedded neighbours at time t) you might think along the following lines. First, we might guess that the s embedded vertices will have about $d^s n$ common neighbours, so we can guess $\big|\mathcal{C}_t(y)\big| \approx d^s n$. What about the $\big|\operatorname{im}\psi_t \cap V_i\big|$ vertices that were embedded up to time t to V_i? There isn't any obvious reason why these should be especially likely or unlikely to go to $\mathcal{C}_t(y)$, so we can guess they are equally spread: $\big|\mathcal{A}_t(y)\big| \approx d^s\big(n - |\operatorname{im}\psi_t \cap V_i|\big)$.

Formalising this, we can try to prove that whp for all t and y we have

$$\big|\mathcal{A}_t(y)\big| = d^s\big(n - |\operatorname{im}\psi_t \cap V_i|\big) \pm \alpha_t n \tag{2}$$

where α_t is an 'error function' which should always be fairly small but presumably grows with t. We call the first term in (2) the *main term* and the second the *error term*.

Observe that if the main term is not too small (as it is in Exercise 6; it has size at least $d^\Delta \mu n$) then the error term is indeed a small error. We will always make sure we are in this situation. Let us now justify that α_t will need to be a growing function of t: it cannot be that α_t is constant.

Think about how likely we are to embed x_t to $\mathcal{A}_{t-1}(y)$. We can write down the probability of doing this (conditioned on $\mathscr{H}_{t-1}$, the embedding history up to time $t-1$): it is the ratio $\frac{|\mathcal{A}_{t-1}(x_t)\cap\mathcal{A}_{t-1}(y)\setminus B_t|}{|\mathcal{A}_{t-1}(x_t)\setminus B_t|}$. We don't know exactly how big B_t is nor how it intersects with the available sets, but we do know it is quite tiny and will not affect the ratio much; we can ignore it. But the denominator of $\frac{|\mathcal{A}_{t-1}(x_t)\cap\mathcal{A}_{t-1}(y)|}{|\mathcal{A}_{t-1}(x_t)|}$ already contains an error $\alpha_{t-1}n$, there will presumably be a similar error in the numerator, and so we cannot hope for an estimate of the conditional probability with an error term better than $C\alpha_{t-1}$, where C is a (moderately!) large constant.

But the size of $\mathcal{A}_t(y)$ is (basically) given by summing up these conditional probabilities and applying Theorem 1.5. Since Theorem 1.5 incurs another error, we cannot hope[1] to end up with better error bounds than we have for the sum of conditional probabilities—that is, roughly $C\alpha_t n$. This is not good enough; the constant C is a (moderately) large constant, whereas we would need to get a (moderately) small constant.

What turns out to work for this kind of example is to take α_t to grow exponentially in t/n (assuming the total number of time steps is linear in n). What we are relying on here is the following observation. If $\alpha_t = c\exp\big(C\frac{t}{n}\big)$, where C is some (large) constant and $c > 0$ is small, then (thinking of α_t as a function of t defined on $\mathbb{R}$, which is increasing) we have

$$\sum_{i=0}^{t-1} C\alpha_i \le \int_{x=0}^{t} \alpha_x \,\mathrm{d}x < \int_{x=-\infty}^{t} \alpha_x \,\mathrm{d}x = \alpha_t n\,. \tag{3}$$

This should be read as: if we make a $C\alpha_i$ error at each time $i = 0, 1, \dots, t-1$, summing up we still have only an $\alpha_t n$ error, which will turn out to be (more or less) what we want. And finally, we can guarantee that even α_n (the largest error term) is small simply by choosing c sufficiently small.

[1] At least without showing 'dynamic concentration' or 'self-correction' of the process; we won't go in this direction.

In this section, we will always work with error bounds of this form. That means we are always going to be trying to argue that the errors we make in estimates at time t are linear in α_t; on the other hand, if we have such an argument, then essentially the above calculation will tell us that our errors at time t will whp stay bounded by $\alpha_t n$.

3.1 Packing triangle factors: setup and main theorems

The main aim of this section is to prove that we can *pack* s^* copies of a $(1-2\gamma)n$-vertex triangle factor (i.e. $(1-2\gamma)\frac{n}{6}$ vertex-disjoint K_3s) into a 'quasirandom' n-vertex graph H_0, where $s^* = \frac{|E(H_0)|}{n}$. A packing of *guest graphs* $G_1, \dots, G_{s^*}$ into a *host graph* H_0 means finding embeddings $\psi_i : G_i \to H_0$ for each $1 \le i \le s^*$ which use each edge of H_0 at most once. Another way to view this is: for each $i \ge 1$ sequentially, the map ψ_i is an embedding of G_i into H_{i-1}, and we define H_i by removing the edges $\psi_i(u)\psi_i(v)$ for $uv \in E(G_i)$. Our choice of t^* means we will need to use most of the edges of H_0.

Quite a bit of what we will do will work in more generality than this specific setting, the exercises will point at how to prove something more. This particular theorem isn't especially interesting and can be proved more easily.

We say that an n-vertex graph H is (α, L)-*quasirandom* if the following holds. Let p be such that $|E(H)| = p\binom{n}{2}$. Then for any $1 \le \ell \le L$ and any distinct vertices $v_1, \dots, v_\ell$ of H, we have

$$\big|N_H(v_1, \dots, v_\ell)\big| = (1 \pm \alpha)p^\ell n\,.$$

This is a fairly strong notion of quasirandomness; the (ε, d)-regularity of the previous section only gives lower bounds for vertex, pair etc. neighbourhoods, and it only does that for *most* vertices, pairs etc. not all. But, for example, a typical dense random graph would be (ε, L) quasirandom for any fixed $\varepsilon > 0$ and L if n is large enough; and in particular that's true for K_n.

Theorem 3.1. *Given $\gamma > 0$, there exists $\alpha_0 > 0$ such that if n is sufficiently large the following holds for $L = 2$. Let H_0 be a $\big(\frac{1}{2}\alpha_0, L\big)$-quasirandom graph with at least $\gamma\binom{n}{2}$ edges. Let s^* be an integer at most $\frac{|E(H)|}{n}$, let $t^* = (1-2\gamma)n$, and let $G_1, \dots, G_{s^*}$ be t^*-vertex triangle factors. Then there is a packing of $G_1, \dots, G_{s^*}$ into H_0.*

The idea here is to analyse a very simple random process, which we already hinted at. For each $1 \le s \le s^*$ in succession, we randomly embed G_s into H_{s-1} and then let H_s be obtained from H_{s-1} by removing the edges used to embed G_s (we call this *stage* s).

We randomly embed G_s much as in the previous section. We suppose the vertices of G_s are $1, \dots, t^*$ and the triangles of G_s form consecutive triples in this order. We let ϕ_0 be the trivial partial embedding of no vertices of G_s into H_{s-1}. Then for each $1 \le t \le t^*$ in succession, we embed t uniformly at random to an available candidate vertex, i.e. an unused vertex of H_0 which is adjacent to $\phi_{t-1}(x)$ for any neighbour x of t in G_s with $x < t$. If there is no such vertex we say the embedding fails, otherwise it succeeds.

As in the previous section (and generalising what we just wrote), if $y \ge t$ we define the available candidate set of y at time t to be

$$\mathcal{A}_{t-1}(y) = \big\{u \in V(H_{s-1}) : u \notin \operatorname{im}\phi_{t-1} \text{ and } u\phi_{t-1}(x) \in E(H_{s-1} \text{ for all } x \in N_{G_s}(y) \text{ with } x < y\big\}\,.$$

Note that we don't define bad vertices; our stronger quasirandomness assumption lets us avoid that. We will not need to use the notion of 'candidate set' in this section, so don't define it.

The intuition of why this random process might succeed is as follows. First, we already saw that (something like) the random embedding of G_s works well in the previous section, provided H_{s-1} is sufficiently quasirandom. We can hope that it works so well that the collection of edges we remove in each stage looks something like a random set of edges: because if we remove random edges from H_0 which is quasirandom, then what we end up with in stage s will be a graph H_{s-1} that indeed is quasirandom.

A lot of the work we need to do is therefore going to be to analyse the embedding of G_s into H_{s-1} in detail. Since s is fixed in this, we'll drop the subscripts. The following theorem is what we need.

Theorem 3.2. *Given $\gamma > 0$ there exists a constant $C' > 0$ such that the following holds for any sufficiently small $\alpha > 0$. Let H be a (α, L)-quasirandom graph on n vertices with at least $\gamma^2\binom{n}{2}$ edges, and G a t^*-vertex triangle factor. With high probability, when we randomly embed G into H, the embedding succeeds. For any given $uv \in E(H)$, the probability that uv is used in the embedding is $(1 \pm C'\alpha)\frac{|E(G)|}{|E(H)|}$, and for any given $uv, uv' \in E(H)$ the probability that both uv and uv' are used in the embedding of G is at most $10\gamma^{-9}n^{-2}$.*

The reason for assuming H has at least $\gamma^2\binom{n}{2}$ edges is that under the conditions of Theorem 3.1, we want to use at most $(1-2\gamma)|E(H_0)|$ edges of H_0 in the packing, so even in the final stage s^* the graph H_{s^*-1} does have at least $\gamma^2\binom{n}{2}$ edges. The first probability statement tells us that any given edge uv of H is about equally likely to be used in the embedding, and the second says that we are quite unlikely to use any specified pair of edges at a given vertex. This is our version of 'something like a random set of edges'.

Assuming Theorem 3.2 it is quite easy to prove Theorem 3.1, so we do that first.

3.2 A simple version of 'growing errors'

Proof of Theorem 3.1.

Exercise 7. Given $\gamma > 0$ and C' as supplied by Theorem 3.2, find $c > 0$ and C such that the following proof works.

For each $x \in \mathbb{R}$, we define

$$\alpha_x = c\exp\left(C\tfrac{x}{n}\right).$$

The choice of α_0 from this is the α_0 returned by Theorem 3.1.

We consider the random packing algorithm described above. For each $1 \le s \le s^*$, we would like to prove that whp the stage s packing succeeds and H_s is (α_s, L)-quasirandom. That is, the collection of statements we want to hold after stage s is: for each $1 \le \ell \le L$ and each $v_1, \dots, v_\ell$ distinct vertices of H_s, we have

$$\left|N_{H_s}(v_1, \dots, v_\ell)\right| = (1 \pm \alpha_s)p_s^\ell n\,,$$

where H_s has $p_s\binom{n}{2}$ edges.

This means the failure events we are interested in are of the following form. Fix $1 \le s \le s^*$, fix $1 \le \ell \le L$, and fix $v_1, \dots, v_\ell$ distinct vertices of H_s. Then $H_{s'}$ is $(\alpha_{s'}, L)$-quasirandom for each $s' < s$, and the random embedding succeeds at each of these stages, but

$$\big|N_{H_s}(v_1, \dots, v_\ell)\big| \neq (1 \pm \alpha_s) p_s^\ell n \,.$$

We also care about the failure event that H_{s-1} is (α_{s-1}, L)-quasirandom but the random embedding of G_s fails. However Theorem 3.2 says this event whp does not occur. Note that in total we have at most $s^* L n^L + s^* \le 2Ln^{L+1}$ failure events, which is a polynomial in n. We just need to prove all these events whp do not occur, where 'whp' needs to be a small enough probability to take the union bound; this is going to be true with (a huge amount of) room to spare.

If s is very small (say smaller than $n^{0.9}$), our failure event has probability zero. We removed at most $2s$ edges from each vertex of H_0, and this is not enough to disturb the $\big(\frac{1}{2}\alpha_0, L\big)$-quasirandomness of H_0 enough to break (α_s, L)-quasirandomness of H_s: this is why we asssume H_0 is $\big(\frac{1}{2}\alpha_0, L\big)$-quasirandom not just (α_0, L)-quasirandom.

Exercise 8. Prove the above assertion.

So now we can assume $s \ge n^{0.9}$. We would like to use Theorem 1.5 to prove that whp our failure event does not occur. We use the method outlined in Section 1.6. So, we let for each $1 \le i \le s$ the random variable Y_i be the change in the common neighbourhood in stage i, that is

$$Y_i = \big|N_{H_{i-1}}(v_1, \dots, v_\ell)\big| - \big|N_{H_i}(v_1, \dots, v_\ell)\big| \,.$$

We let $\mathscr{H}_i$ be the history of the random packing up to and including stage i. We let $\mathcal{E}$ be the event that the random packing succeeds up to and including stage $s-1$, and H_i is (α_i, L)-quasirandom for each $1 \le i \le s-1$. We let $R = 2\ell$ (in a given stage i, we can remove at most 2 edges from each v_j and so $0 \le Y_i \le 2\ell$).

Looking to Theorem 1.5, if we can now find some numbers a and b such that $\sum_{i=1}^s \mathbb{E}[Y_i | \mathscr{H}_{i-1}] = a \pm b$ holds within $\mathcal{E}$, then we get (using the first two parts of that theorem)

$$\mathbb{P}\Big[\sum_{i=1}^s Y_i \neq a \pm (\delta a + b) \text{ and } \mathcal{E} \text{ occurs}\Big] \le 2 \exp\Big(-\frac{\delta^2 a}{6R}\Big) \,. \tag{4}$$

The left side is (we'll see) an upper bound on the probability of our failure event. We'll see the a in the right hand side is at least $n^{0.7}$, so we can (for example) take $\delta = n^{-0.1}$ and the right hand side will still tend to zero superpolynomially; with high probability the failure event does not occur. Now the union bound over all the failure events says whp no failure event occurs, and as in Section 1.6 this is just another way of saying 'we proved the theorem'.

The only thing missing is to write down a and b and prove the sum of conditional expectations is $a \pm b$ whenever $\mathcal{E}$ occurs. The first part of this is easy. We let

$$a = \big(p_0^\ell - p_s^\ell\big) n \quad \text{and} \quad b = \tfrac{1}{10}\gamma^{2L}\alpha_s n \,,$$

because then an easy calculation (using $p_s \ge \gamma^2$) gives

$$\big(1 \pm \tfrac{1}{2}\alpha_0\big) p_0^\ell n - a \pm (\delta a + b) = \big(1 \pm \alpha_s\big) p_s^\ell n \,,$$

which is what we want: this justifies the LHS of (4) is an upper bound on the probability of our failure event. The claim that $a \geq n^{0.7}$ follows from our definition of a together with calculating p_s and noting $s \geq n^{0.9}$.

What is 'harder' is to prove $\sum_{i=1}^{s} \mathbb{E}[Y_i|\mathscr{H}_{i-1}] = a \pm b$ holds within $\mathcal{E}$. That is, it is actually not hard at all, it is just a tedious calculation at this point, keeping track of error terms and making a few approximations.

First off, if we are in $\mathcal{E}$ then in particular H_{i-1} is (α_{i-1}, L)-quasirandom, so we get the two probability statements from Theorem 3.2. We can put in explicit numbers for the first: $|E(G_i)| = (1-2\gamma)n$, and $|E(H_{i-1})| = p_{i-1}\binom{n}{2} = |E(H_0)| - (i-1)(1-2\gamma)n$. So for any distinct u, v, v' with $uv, uv' \in E(H_{i-1})$ we have

$$\mathbb{P}[uv \text{ is used}] = \big(1 \pm C'\alpha_{i-1}\big)\tfrac{(1-2\gamma)n}{p_{i-1}\binom{n}{2}} \quad \text{and} \quad \mathbb{P}[uv, uv' \text{ are used}] \leq 10\gamma^{-9}n^{-2}\,.$$

Let's see how we can use this to estimate $\mathbb{E}[Y_i|\mathscr{H}_{i-1}]$. First, we can use linearity of expectation. We can write $Y_i = Z_1 + \cdots + Z_q$ as a sum of Bernoulli random variables, where $q = \big|N_{H_{i-1}}(v_1, \ldots, v_\ell)\big|$ and Z_j is 1 if the jth vertex of $N_{H_{i-1}}(v_1, \ldots, v_\ell)$ (we can enumerate the set arbitrarily) is not in $N_{H_i}(v_1, \ldots, v_\ell)$. In other words, $Z_j = 1$ if we use an edge from at least one of $v_1, \ldots, v_\ell$ to the jth vertex of $N_{H_{i-1}}(v_1, \ldots, v_\ell)$. We'd like to know what this probability is.

That is, we would like to know, given some w adjacent to all of $v_1, \ldots, v_\ell$ in H_{i-1}, the probability that at least one edge $v_k w$ is used in embedding G_i. This probability is

$$\begin{aligned}
&\sum_{1\leq k\leq \ell} \mathbb{P}[v_k w \text{ is used}] - \sum_{1\leq k<k'\leq \ell} \mathbb{P}[v_k w \text{ and } v_{k'} w \text{ are used}] \\
&\qquad = \ell \cdot \big(1 \pm C'\alpha_{i-1}\big)\tfrac{(1-2\gamma)n}{p_{i-1}\binom{n}{2}} \pm \binom{\ell}{2} \cdot 10\gamma^{-9}n^{-2} \\
&\qquad = \ell \cdot \tfrac{2(1-2\gamma)n}{p_{i-1}n^2} \pm C'\alpha_{i-1} \cdot \tfrac{2}{\gamma^2 n} \pm 10\ell^2\gamma^{-9}n^{-2} \pm n^{-1.5} \\
&\qquad = \ell \cdot \tfrac{2(1-2\gamma)}{p_{i-1}n} \pm 4C'\alpha_{i-1}\gamma^{-2}n^{-1}\,,
\end{aligned}$$

where to see the first line is true, observe that we use either zero, one or two but not more edges from w to $\{v_1, \ldots, v_\ell\}$ in embedding G_i, and the second line simply substitutes the known probability estimates. For the third line, we separated the first 'main' term which we wrote accurately from the error terms. We know $\binom{n}{2} = \frac{1}{2}n^2 \pm n$, and substituting this in the main term (to simplify it!) gives us the extra $n^{-1.5}$ error term (this is a lazy but valid estimate). Finally, we notice one error term is much bigger than the others and collect the smaller ones into the bigger one. The point here is: we need to be careful with our main term, which we have to estimate accurately, but we don't care much about getting extra constant factors in the error term.

We are now set up to write our conditional expectation estimate: we know the conditional expectation of each Z_j, we know there are

$$q = |N_{H_{i-1}}(v_1, \ldots, v_\ell)| = \big(1 \pm \alpha_{i-1}\big)p_{i-1}^{\ell}n$$

terms to sum, and by linearity of expectation we can just sum the conditional expectations:

$$\begin{aligned}
\mathbb{E}[Y_i|\mathscr{H}_{i-1}] &= \big(1 \pm \alpha_{i-1}\big)p_{i-1}^{\ell}n \cdot \Big(\ell \cdot \tfrac{2(1-2\gamma)}{p_{i-1}n} \pm 4C'\alpha_{i-1}\gamma^{-2}n^{-1}\Big) \\
&= 2(1-2\gamma)\ell p_{i-1}^{\ell-1} \pm 8LC'\gamma^{-L}\alpha_{i-1}
\end{aligned}$$

where on the second line we again separated the (accurate) main term from the three error terms we get by expanding brackets, then added some extra factors of γ and L to make sure the one term we wrote covers all three error terms.

Next, we need to evaluate

$$\sum_{i=1}^{s} \mathbb{E}[Y_i|\mathscr{H}_{i-1}] = \sum_{i=1}^{s} 2(1-2\gamma)\ell p_{i-1}^{\ell-1} \pm 8LC'\gamma^{-L}\alpha_{i-1}\,. \tag{5}$$

Again, it's simpler to separately deal with the sum of main terms and error terms. The main term we are trying to get to for the sum is $(p_0^\ell - p_s^\ell)n$, so we rather hope that

$$(p_{i-1}^\ell - p_i^\ell)n \approx 2(1-2\gamma)\ell p_{i-1}^{\ell-1}\,.$$

Since we know $p_{i-1}\binom{n}{2} = p_i\binom{n}{2} + (1-2\gamma)n$ (the number of edges of H_{i-1} is that of H_i plus the edges of G_i), we can write

$$p_i = p_{i-1} - (1-2\gamma)\tfrac{2}{n-1}$$

and so

$$p_i^\ell = p_{i-1}^\ell - \ell\cdot p_{i-1}^{\ell-1}\cdot(1-2\gamma)\tfrac{2}{n-1} \pm 6^\ell n^{-2} = p_{i-1}^\ell - \ell\cdot(1-2\gamma)\tfrac{2}{n} \pm 10^\ell n^{-2}\,,$$

where we used the binomial expansion and noticed the less than 2^ℓ terms that take the second factor 2 or more times are all bounded by $3^\ell n^{-2}$.

And this is what we wanted:

$$2(1-2\gamma)\ell p_{i-1}^{\ell-1} = n\Big(p_{i-1}^\ell - p_i^\ell \pm 10^\ell n^{-2}\Big)\,,$$

$$\text{so}\quad \sum_{i=1}^{s} 2(1-2\gamma)\ell p_{i-1}^{\ell-1} = n\sum_{i=1}^{s}\Big(p_{i-1}^\ell - p_i^\ell \pm 10^\ell n^{-2}\Big) = n\big(p_0^\ell - p_s^\ell\big) \pm 10^\ell\,,$$

where for the final equality we use $s \le s^* \le n$.

Finally, we need to evaluate the sum of error terms from (5). Here, we don't need to be accurate. In particular, we can use the observation from (3) to write

$$\sum_{i=1}^{s}\alpha_{i-1} \le C^{-1}\alpha_s n$$

and hence the sum of error terms from (5) is at most

$$\sum_{i=1}^{s} 8LC'\gamma^{-L}\alpha_{i-1} \le 8LC'\gamma^{-L}C^{-1}\alpha_s n \le \tfrac{1}{20}\gamma^{2L}\alpha_s n\,,$$

by choice of C. Finally from (5) and these two calculations we get

$$\sum_{i=1}^{s}\mathbb{E}[Y_i|\mathscr{H}_{i-1}] = n\big(p_0^\ell - p_s^\ell\big) \pm 10^\ell \pm \tfrac{1}{20}\gamma^{2L}\alpha_s n = a \pm b\,,$$

where a and b are as we defined above. This completes the proof. □

Something you should notice in this above proof is that while it is a bit long for something 'easy', there are really rather few ideas involved. Maybe the only real idea there, beyond that we follow the template of Section 1.6, is how to calculate the conditional expectation; and this is also a rather standard strategy (to calculate the expectation of a complicated random variable, see if you can write it as a sum of simpler ones).

Indeed, there are black-box theorems which would do most of the work of this proof for you: this is Wormald's 'Differential Equations Method'. The connection to differential equations here can be seen in (5): the main term here looks rather like the derivative (with respect to p, which is what is changing) of p^ℓ, and what we did was the discrete version of integrating that. However, the black-box theorems don't apply in all situations where one would like to use them.

A final observation—which we'll return to later—is that the choice of α_x works because of the slightly odd way we chose constants in Theorem 3.2, where C' is fixed and independent of α, so that the relative error in the probability estimate is linear in the error in quasirandomness.

3.3 Something a bit more complicated

We still need to prove Theorem 3.2. You should believe the 'embedding succeeds whp' part of this theorem; we already more or less proved it in the previous section. But in the previous section we only gave crude bounds on how big available candidate sets might get; those bounds aren't good enough to get the accurate probability estimate we need. So we will start by trying to prove accurate bounds.

Given a (α, L)-quasirandom n-vertex graph H with $p\binom{n}{2}$ edges, and a subset $X \subseteq V(H)$, we say that (H, X) satisfies the (β, L)-*diet condition* if for all $1 \leq \ell \leq L$ and $v_1, \dots, v_\ell$ distinct vertices of $V(H)$ we have

$$\big|N_H(v_1, \dots, v_\ell) \setminus X\big| = (1 \pm \beta)p^\ell\big(n - |X|\big)\,.$$

We will always use this definition with X being the image of a partial embedding of G; so X 'eats up' about the right fraction (i.e. what would be expected for a random set of size $|X|$) of neighbourhoods, pair neighbourhoods et cetera. Since an available candidate set for a vertex is always something of the LHS form, in particular this tells us how big available candidate sets will be.

As in the previous subsection, we need to define an exponentially increasing error term. The right choice, given $\gamma, \alpha > 0$, is

$$\beta_x = 2\alpha \exp\left(C''\tfrac{x}{n}\right)$$

where $C'' > 0$ depends only on γ, and we set $C' = \beta_n$ to be the number that is returned by Theorem 3.2. We'll fix these constant choices through the section, and choose p such that $|E(H)| = p\binom{n}{2}$ (so $p \geq \gamma^2$).

It turns out to be hard to analyse this process vertex-by-vertex. We will 'cleverly' do it triangle-by-triangle instead; recall that G consists of a triangle-factor and we embed the three vertices of each triangle consecutively. That is, for each $0 \leq t \leq \frac{1}{3}t^*$, we let $\mathscr{H}_t$ denote

the history of the random embedding up to and including embedding the tth *triangle* (i.e. $3t$ vertices in total).

What we would like to prove first is that whp the pair $(H, \operatorname{im}\phi_{3t})$ satisfies the (β_{3t}, L)-diet condition for each $0 \le t \le \frac{1}{3}t^*$. Our proof is going to look rather like the proof in the previous subsection. That is, first off we will observe that if t is small (again, say $t < n^{0.9}$) then the statement holds deterministically. Given ℓ and $v_1, \dots, v_\ell$ we have

$$\big|N_H(v_1,\dots,v_\ell) \setminus \operatorname{im}\phi_{3t}\big| = \big|N_H(v_1,\dots,v_\ell)\big| \pm n^{0.9} = (1\pm\alpha)p^\ell n \pm n^{0.9} = (1\pm\beta_{3t})p^\ell(n-3t)$$

where the second equality is by (α, L)-quasirandomness of H and the third since t is small and by choice of β_{3t}.

To deal with later times, we need to estimate a conditional expectation: the expected number of vertices of the tth triangle that will be embedded to $N_H(v_1,\dots,v_\ell)$, conditioned on $\mathscr{H}_{t-1}$, and assuming that the diet condition holds. As in the previous subsection, we can use linearity of expectation to split this up, and the critical work is the following lemma.

Lemma 3.3. *Under the conditions of Theorem 3.2, and with constants as above, for any $1 \le t \le \frac{3}{t}^*$ the following holds. Fix $v \in V(H)$, and suppose $\mathscr{H}_{t-1}$ satisfies the following: ϕ_{3t-3} is a partial embedding of the first $3t-3$ vertices of G into H such that $(H, \operatorname{im}\phi_{3t-3})$ has the $(\beta_{3t-3}, 2)$-diet condition, and $v \notin \operatorname{im}\phi_{3t-3}$. Then we have*

$$\mathbb{P}\big[v \text{ is embedded to by the } t\text{-th triangle} \big| \mathscr{H}_{t-1}\big] = \big(1 \pm 10\beta_{3t-3}\big)\tfrac{3}{n-3t+3}\,.$$

Proof. Suppose the vertices of the tth triangle are x, y, z in that order. In the rest of this proof we condition on $\mathscr{H}_{t-1}$ and assume ϕ_{3t-3} has the $(\beta_{3t-3}, 2)$-diet condition, and $v \notin \operatorname{im}\phi_{3t-3}$.

The probability that $x \hookrightarrow v$ is exactly $\frac{1}{n-3t+3}$, because v is one of the $n-3t+3$ vertices not in $\operatorname{im}\phi_{3t-3}$.

To embed y to v, we need first to embed x to some $u \in N_H(v)$ and then y to v. The probability of this is

$$\tfrac{|N_H(v)\setminus\operatorname{im}\phi_{3t-3}|}{n-3t+3} \cdot \tfrac{1}{|N_H(u)\setminus\operatorname{im}\phi_{3t-3}|} = (1\pm\beta_{3t-3})p \cdot \tfrac{1}{(1\pm2\beta_{3t-3})p(n-3t+3)} = \tfrac{1\pm4\beta_{3t-3}}{n-3t+3}\,,$$

where the first term of the LHS is the probability of $x \hookrightarrow u \in N_H(v)$ and the second the probability of $y \hookrightarrow v$. Formally we should notice this second term depends on which u is chosen (and so sum over choices of u), but since the (β_{3t-3}, L)-diet condition, which we use to get the first equality, says that the answer is roughly the same for every u we don't bother.

Finally, to embed z to v we need to first embed x to $u \in N_H(v)$, then y to $w \in N_H(u,v)$, then finally z to v. Similarly, the probability of this is

$$\begin{aligned}
&\tfrac{|N_H(v)\setminus\operatorname{im}\phi_{3t-3}|}{n-3t+3} \cdot \tfrac{|N_H(u,v)\setminus\operatorname{im}\phi_{3t-3}|}{|N_H(u)\setminus\operatorname{im}\phi_{3t-3}|} \cdot \tfrac{1}{|N_H(u,w)\setminus\operatorname{im}\phi_{3t-3}|}\\
=&(1\pm\beta_{3t-3})p \cdot \tfrac{(1\pm\beta_{3t-3})p}{1\pm\beta_{3t-3}} \cdot \tfrac{1}{(1\pm\beta_{3t-3})p^2(n-3t+3)}\\
=&\tfrac{1\pm5\beta_{3t-3}}{n-3t+3}\,,
\end{aligned}$$

and summing up these three disjoint events we get the desired probability. □

As we'll now see, this is the critical 'trick' that allows us to avoid the problem of needing triple neighbourhoods to analyse pair neighbourhoods etc. mentioned at the end of Section 1.6: this lemma only needs the $(\beta_{3t-3}, 2)$-diet condition as input even when we are using it to handle the (β, L)-diet condition for larger L. We can now prove that whp the (β_t, L)-diet condition is maintained throughout the process. Formally, we'll only prove this for $3t$ for integer t, which is what we actually need (but it's easy to see that this implies the $(2\beta_t, L)$-diet condition holds for all t).

Lemma 3.4. *Under the conditions of Theorem 3.2, and with constants as above, with probability at least $1 - n^{-10}$ the following holds for all $0 \leq t \leq \frac{1}{3}t^*$. The pair $(H, \operatorname{im}\phi_{3t})$ has the (β_{3t}, L)-diet condition.*

We turn out to need an explicit probability bound here for the rest of the proof of Theorem 3.2.

Exercise 9. Follow the strategy seen in the proof of Theorem 3.1 to prove this lemma.

At this point, we have proved the 'whp succeeds' part of Theorem 3.2. What we still need to do is to prove the two probability estimates. The easier is to upper bound the probability that uv and uv' are both used in embedding G.

Lemma 3.5. *Under the conditions of Theorem 3.2, and with constants as above, for any given $uv, uv' \in E(H)$ the probability that both uv and uv' are used in embedding G is at most $10\gamma^{-9}n^{-3}$.*

Proof. The only way this can happen is if some triangle xyz of G is embedded to u, v, v' in some order. Fix the tth triangle xyz of G, and suppose $x < y < z$.

The probability that x is embedded to some vertex of $\{u, v, v'\}$ is at most $\frac{3}{n-3t+3}$, since x will be embedded to a set of size $n - 3t + 3$ uniformly at random, at most 3 elements of which are in $\{u, v, v'\}$. The probability that then y is embedded to a member of $\{u, v, v'\}$ is at most $\frac{2}{|\mathcal{A}_{y-1}(y)|}$, for a similar reason, and finally we have at most $\frac{1}{|\mathcal{A}_{z-1}(z)|}$ chance of putting z to the final remaining member of $\{u, v, v'\}$.

This would suggest the probability of $xyz \hookrightarrow uvv'$ is at most

$$\tfrac{1}{n-3t+3} \cdot \tfrac{2}{|\mathcal{A}_{y-1}(y)|} \cdot \tfrac{1}{|\mathcal{A}_{z-1}(z)|} \leq \big(1 + 4\beta_{3t-3}\big) \cdot \frac{6}{p^3(n-3t+3)^3}\,.$$

However, we need to be a bit more careful: while we can estimate

$$|\mathcal{A}_{y-1}(y)| = \big(1 \pm \beta_{3t-3}\big)p(n-3t-3)$$

if (H, ϕ_{3t-3}) satisfies the (β_{3t-3}, L)-diet condition, we don't know this event occurs. What we can do is say that the above estimate is an upper bound for the probability that $xyz \hookrightarrow uvv'$ and ϕ_{3t-3} satisfies the (β_{3t-3}, L)-diet condition: so by Lemma 3.4 the probability of $xyz \hookrightarrow uvv'$ is at most

$$\big(1 + 4\beta_{3t-3}\big) \cdot \frac{6}{p^3(n-3t+3)^3} + n^{-10} \leq \tfrac{12}{p^3(2\gamma)^3n^3} \leq 10\gamma^{-9}n^{-3}\,,$$

where here we use $3t \leq t^* = (1-2\gamma)n$ and $p \geq \gamma^2$.

Now we take the union bound over the at most n choices of t to obtain the claimed probability. □

Finally, we need the accurate estimate of the probability of using $uv \in E(H)$ when we embed G. This is more difficult: in the above lemma, we were 'lazy' in that we said that at most 3 of the vertices u, v, v' are unused at the time we come to embed the tth triangle of G. The rest of the estimates we used would be good enough, but clearly this estimate is (especially when t is large) far from the truth: some or all of the vertices are likely to have been used. We need to find out how likely.

Lemma 3.6. *Under the conditions of Theorem 3.2, and with constants as above, let u and v be any two vertices of H. Then we have for each $0 \le t \le \frac{1}{3}t^*$*

$$\mathbb{P}\big[u \notin \operatorname{im}\phi_{3t}\big] = \big(1 \pm 2\beta_{3t}\big)\tfrac{n-3t}{n} \quad \textit{and} \quad \mathbb{P}\big[u, v \notin \operatorname{im}\phi_{3t}\big] = \big(1 \pm 2\beta_{3t}\big)\Big(\tfrac{n-3t}{n}\Big)^2 .$$

Proof. The idea here is simply to apply Lemma 3.3 to estimate the probability of using u given ϕ_{3t-3} (and then work by induction). An annoyance here is that this lemma doesn't apply for all possible ϕ_{3t-3}, but only those where the diet condition holds. Intuitively this shouldn't really matter: Lemma 3.4 says that it's very unlikely we encounter such a ϕ_{3t-3}, and this shouldn't really change the calculation. Here is a trick to formalise that (for the first probability statement).

We define events $Z_1, \dots, Z_t$ as follows. For each $1 \le i \le t$, we run the random embedding up to generating ϕ_{3i-3}. If $(H, \operatorname{im}\phi_{3t-3})$ satisfies the (β_{3t-3}, L)-diet condition, Z_j has not occurred for any $1 \le j < i$, and u is used by the tth triangle, we say Z_i has occurred. If $(H, \operatorname{im}\phi_{3t-3})$ does not satisfy the (β_{3t-3}, L)-diet condition and Z_j has not occurred for any $1 \le j < i$, with probability $\frac{3}{n-3i+3}$ we say Z_i has occurred. Otherwise, Z_i does not occur.

By construction, these events are disjoint (so the probability of their union is the sum of their probabilities). By Lemma 3.3, if $\mathscr{H}_{i-1}$ is such that Z_j does not occur for any $1 \le j \le i-1$ then we have

$$\mathbb{P}\big[Z_i \big| \mathscr{H}_{i-1}\big] = \big(1 \pm 10\beta_{3i}\big)\tfrac{3}{n-3i+3} ;$$

indeed, obtaining this equality for all such $\mathscr{H}_{i-1}$ (not just most of them) is why we made this definition of Z_i. Finally, the event

$$\Big(\bigcup_{i=1}^{t} Z_i\Big) \triangle \big\{u \in \operatorname{im}\phi_{3t}\big\} ,$$

where $\triangle$ is the symmetric difference, is contained in the event that $(H, \operatorname{im}\phi_{3i})$ fails the (β_{3i}, L)-diet condition for some $i \le t$. By Lemma 3.4 this has probability at most n^{-10}.

So it is enough to estimate the probability of the union of the Z_i. We have

$$\mathbb{P}\big[Z_i\big] = \mathbb{P}\big[Z_1 \cup \dots \cup Z_{i-1} \text{ do not occur}\big] \cdot \big(1 \pm 10\beta_{3i}\big)\tfrac{3}{n-3i+3} .$$

Letting $q_i = \mathbb{P}\big[Z_i\big]$ we can rewrite this as a recursion:

$$q_i = \big(1 - q_1 - \dots - q_{i-1}\big) \cdot \big(1 \pm 10\beta_{3i}\big)\tfrac{3}{n-3i+3}$$

and it is not too hard to check that the solution is

$$q_i = \big(1 \pm 20\beta_{3i}\big)\tfrac{3}{n} .$$

Assuming for a moment this is accurate, we have

$$1 - \sum_{i=1}^{t} q_i = \tfrac{n-3t}{n} \pm \sum_{i=1}^{t} 60\beta_{3i} n^{-1} = \tfrac{n-3t}{n} \pm \gamma\beta_{3t} = \big(1 \pm \beta_{3t}\big)\tfrac{n-3t}{n}\,,$$

where the second equality uses the exponential growth of β_x and the final one uses the fact $\frac{n-3t}{n} \geq \frac{n-t^*}{n} \geq 2\gamma$.

Exercise 10. Prove the claimed recursion solution. *Warning: you need to prove both upper and lower bounds together, otherwise something will not work!*

This now proves the claimed first probability statement. Indeed, we already saw

$$\mathbb{P}\big[u \notin \operatorname{im}\phi_{3t}\big] = 1 - \sum_{i=1}^{t} q_i \pm n^{-10}$$

and substituting the calculation of the first term in we are done.

Exercise 11. Perform the similar calculation for the second statement. *Hint: you will need to upper bound the probability that the t-th triangle embedding uses both u and v if you want to apply Lemma 3.3.*

□

Finally, we are now in a position to copy the proof of Lemma 3.5, substituting in the accurate bounds from Lemma 3.6 instead of the lazy estimate used there, to conclude the final part of Theorem 3.2.

Exercise 12. Do this calculation, and check that the various lemmas in this section together indeed prove Theorem 3.2.

4 Pitfalls and further ideas

4.1 A trap

So far, we stuck to a certain way to choose our growing error parameter: it should be exponential. This is a natural choice if we can argue our conditional expectations at time t (when we have an error α_t in the statements we are tracking) come with a relative error linear in α_t.

We also saw that a lot of the difficulty is figuring out how to calculate all the conditional expectations we need; it's natural to feel we ought to throw all the tools at it that we know. But we didn't do that. We did not, for instance, use moment bounds (Chebyshev's inequality and relatives) or anything like the Cauchy-Schwarz inequality. In particular, anyone familiar with the theory of quasirandom graphs will have seen quite a few places where it would be very natural to use a well-known implication: if we know a graph has all vertices of degree roughly pn and pair degrees roughly $p^2 n$, then for any two linear-sized subsets of vertices (not just neighbourhoods!) the edge density between will be roughly p. For example, we

could have used that to avoid a few 'bad' vertices in the proof of Theorem 3.2 (i.e. write a random embedding algorithm looking like the one we used to prove the Blow-up Lemma) and we wouldn't have had to do all the work of proving that whp our available candidate sets are roughly the 'right' size.

Unfortunately, if you do this you can easily end up with a relative error in the conditional expectation at time t which is not linear in α_t. Most likely you instead get a relative error like $\sqrt{\alpha_t}$.

In that case, you will need to choose values for α_i which satisfy, for each $1 \leq t \leq n$ (supposing the process goes over n steps), the inequality

$$\sum_{i=1}^{t} \sqrt{\alpha_{i-1}} \leq C^{-1}\alpha_t n\,.$$

This is almost the same inequality as the one we saw an exponentially growing α_i does satisfy in (3); the only difference is the square root on the LHS.

Exercise 13. Given $\alpha_0 > 0$, find values for the remaining α_i satisfying the above inequalities, which make α_n as small as possible. What is the dependence on α_0?

It's very natural to feel—especially coming from extremal combinatorics—that this process of choosing the α_i is part of the 'choosing the constants' that you do right at the end of writing up your proof, once you're sure nothing else has to change[2]. What this exercise says is that this is not accurate. It is unpleasant to write 40 pages and then discover that 'choosing the constants' turns out to be impossible. Checking the error terms will sum is something you should probably do as soon as you think you found a way to estimate the conditional expectation.

4.2 Further ideas

It's worth trying to modify the proof of Theorem 3.1 to work with graphs G_i that are not triangle factors. The following should not be too hard.

Exercise 14. Modify the deduction of Theorem 3.1 from Theorem 3.2 to work for any bounded degree graphs with a bit less than $|E(H_0)|$ edges (where does the bounded degree come in?).

Quite a bit of the proof of Theorem 3.2 can also be modified, but one thing in particular is difficult. We don't analyse vertex-by-vertex, but rather triangle-by-triangle. This is necessary for our approach: if we embed triangle xyz, we can calculate the probability of $x \hookrightarrow u$, conditioning on the history, precisely because x doesn't have any embedded neighbours. If we condition on the embedding of x and try to calculate the probability of $y \hookrightarrow u$, the answer depends a lot on where we embedded x; we certainly won't be able to write down any deterministic estimate.

[2]If you are from the school of thought that doesn't bother to choose constants at all because 'of course we can do that', think again!

This is a problem for a vertex-by-vertex analysis because we would like to write the number of vertices embedded to $N_H(u)$ as a sum of random variables, one for each vertex embedding (as opposed to one for each triangle embedding), but then Theorem 1.5 demands that we condition on where x is embedded when we calculate the expectation for y. We avoided this by not needing to deal with this (nastily behaved) conditional expectation. But if G is connected, we will not be able to play this trick.

Exercise 15. Prove Theorem 3.2 when G is a $(1-2\gamma)n$-vertex path embedded in order. *Hint: try applying Theorem 1.5 separately for the odd and even vertices along the path.*

If G is more complicated, we might not be able to find a convenient partition of its vertices into a few sets that allow us to play the trick from the above exercise. In particular, if vertices of G have neighbours far away in the embedding order, we cannot do that.

Exercise 16. Prove Theorem 3.2 for the graph G on $\left\{1, 2, \ldots, \frac{n}{2}\right\}$ in which we have an edge from i to $i + \frac{n}{4}$ for each $1 \leq i \leq \frac{n}{4}$ and vertices are embedded in numerical order. *Hint: You will not be able to write a deterministic estimate for each individual conditional expectation. But we only need a deterministic estimate of the* ***sum*** *of the conditional expectations for Theorem 1.5. This exercise is moderately hard.*

In another direction, you might want to pack guest graphs with n vertices into H_0. This is possible provided $\sum_i |E(G_i)|$ is a bit smaller than $|E(H_0)|$. (It's not easy to use the last few edges and get a perfect packing..!) One way to try to do this is to look back at how we proved Theorem 2.2 and try to mimic that: embed the last few vertices by a matching argument.

Exercise 17. Make the above work for guest graphs being n-vertex triangle factors. *Hint: It is not too hard to get a version of Theorem 3.2 which looks reasonable, but you will need to be very, very careful to get enough control of the relative error in estimating the probability that embedding G uses a given uv. It is probably easier to separate $E(H_0)$ into a 'bulk' and a 'reservoir' randomly and use the reservoir for completing spanning embeddings. This however needs rethinking a lot of calculations!*

The next step after this is to prove the main theorem of `https://arxiv.org/abs/1711.04869`. We've seen most of the ideas there at this point (some in exercises to which the solution is not provided). It turns out that dealing with high-degree vertices is not all that hard; it mainly just requires being less lazy with the probabilistic estimates (one should not set $\delta = n^{-0.1}$ really..!). But dealing with general bounded-degree graphs does require one further idea that's not here. Exercise 16 is pushing you in the direction of defining the 'cover condition' from that paper. Proving the cover condition for the special case of Exercise 16 is not so hard, but for more general graphs it does get tricky. The idea you need is contained in the following exercise.

Exercise 18. Revisit Theorem 2.2. Suppose G is a $3n$-vertex triangle factor, and try to go through the proof with one change: do *not* assume that the neighbours of buffer vertices (there will always be two) come consecutively in the embedding order. Prove Claim 2.2.

Hint: You cannot just bound some conditional probability as in Claim 2.2. Before trying this, think about how you would choose the sequence of histories and ensure the conditions

of Theorem 1.5 are met, and given this what one can say about the required conditional probabilities.

Instead, first try to prove a lower bound on the number of buffer vertices whose first neighbour in the embedding order is mapped to $N_H(v)$, using Theorem 1.5 (here the obvious way to choose histories does work). Then use the event that this bound holds as part of the event $\mathcal{E}$ for a second application of Theorem 1.5 to prove a lower bound on the number of buffer vertices whose first and second neighbours are both mapped to $N_H(v)$.

4.3 Conclusion

Of course, there is a lot more one could say in the direction of analysing random processes.

We showed a few ways to estimate conditional expectations; of course there are more things one can try, and some of them will even work.

We did not say anything about trying to analyse processes 'to completion'.

A good example here is the *triangle removal process*, where we start with K_n and repeatedly remove a uniform random triangle, until there is no triangle left to remove. One could analyse this to some extent in the style of Theorem 3.1; indeed, it's easier than that theorem to analyse the first 99% of the process. But our approach won't say much about, for example, exactly how many edges should be left when the process terminates? To answer this kind of question, first one needs to make use of the fact that K_n is much more quasirandom than just (α, L)-quasirandom for some small α. That is, we would want to rethink the choice of constants (which gets tricky).

Even then, a first analysis will suggest that some parameters we want to control are going to start misbehaving earlier than we would like. This is again misleading. For example, if a vertex v has a bit too high degree at some stage of the triangle removal process, it's very likely that v is also in more than the average number of triangles: it's more likely than other vertices to be picked for the next triangle to remove. This means we expect some *self-correction* of the process.

Taking advantage of this intuition is possible, but we didn't discuss at all how to do it. A good place to start here is Tom Bohman's analysis of the triangle-free process, and the two follow-ups by Fiz Pontiveros, Griffiths and Morris; and Bohman and Keevash. The latter two are hard work, though!

We did not say too much about analysing processes where there can be large changes in a single step.

In terms of where this comes into what we did, for example consider the proof of Theorem 3.1. Because G_i has maximum degree 2, the change in $N_H(v_1, \dots, v_\ell)$ is bounded by $2L$ in a single embedding. So we can choose $R = 2L$, and we will get (with much room to spare) enough concentration. If G_i had maximum degree $\sqrt{n}$, we would have needed to take $R = L\sqrt{n}$ instead, but actually this would've worked fine.

The limit for our approach is $R = \Theta\big(\frac{n}{\log n}\big)$ (and to get there we need to be more careful with our choice of δ). This is actually the limit of where this particular process works; for

graphs with larger maximum degree the process in general will *not* succeed (there are more places than just this one where we need this maximum degree!).

In general, if a single step of a process can change some parameter by $\gg \frac{n}{\log n}$ (where we are implying that the parameter supposed to have size on the order of n, and the union bounds we want to take are polynomial in n) then there isn't any good reason to assume that the parameter we're looking at will actually be concentrated. You might need to think of how to do the analysis without considering this parameter. On the other hand, if your 'single step' is really made of a lot of small choices—for example, a 'single step' in the proof of Theorem 3.1 is a lot of small choices) then you may be able to progress by making this explicit.

We certainly did not say anything about something that the probabilists would consider vital: what is the distribution of the outcome? We just establish that some parameters in our processes are likely to stay within certain ranges, we do not say anything much about what the distribution of these parameters is. All our parameters are 'really' much better concentrated than what we have proved; for a probabilist that means we're answering the wrong questions.

This is a fairly significant break between using random processes to prove deterministic stuff (which is what I'm trying to concentrate on) and the probabilist's view on this. It is of course much harder to say anything meaningful about distributions—and this means that probabilists tend not to consider processes which are as complicated as the ones we looked at. This is maybe one reason why looking in the (extensive) probabilistic literature doesn't tend to be very useful for someone who wants to prove combinatorial theorems.