The triangle-free process and lower bounds for $r(3, k)$

Sanjana Das

May 12, 2025

A survey of the papers [Kim95], [Wol11], and [Boh09].

Contents

§1 Introduction

§1.1 The problem and history

The *triangle-free process* is the following random graph process: We start with an empty n-vertex graph G_0. Then at every step, we pick a random edge e. If e can be added to the current graph without creating a triangle, then we add it; otherwise we discard it. We keep on doing this until no more edges can be added without creating a triangle; at this point, we've produced a random maximal triangle-free graph.

One of the main reasons people initially studied the triangle-free process was the question of determining the *Ramsey number* $r(3, k)$, the smallest n for which every n-vertex graph has a triangle or an independent set of size k. In 1981, Ajtai, Kómlos, and Szemerédi [AKS81] proved that every n-vertex triangle-free graph G has independence number $\alpha(G) = \Omega(\sqrt{n \log n})$; this shows

$$r(3, k) = O\left(\frac{k^2}{\log k}\right).$$

As discussed by Spencer [Spe95], there are heuristic reasons to think that the graph produced by the triangle-free process should satisfy $\alpha(G) = O(\sqrt{n \log n})$, which would provide a matching lower bound on $r(3, k)$. (We'll discuss such a heuristic in Subsection 1.2.)

Since the triangle-free process is highly non-independent, it's difficult to analyze. However, in 1995, Kim [Kim95] solved the problem of $r(3, k)$ using a 'semi-random' modification of the triangle-free process — he found a way to introduce enough independence into the process to make it easier to analyze (while preserving these heuristics), and he showed this modified process does produce a graph with $\alpha(G) = O(\sqrt{n \log n})$.

Although this solved the problem of $r(3, k)$, studying the triangle-free process continued to be an interesting problem in its own right. In 2009, Bohman [Boh09] proved that the triangle-free process really does produce a graph with $\Theta(n^{3/2}\sqrt{\log n})$ edges and with $\alpha(G) = O(\sqrt{n \log n})$. In 2011, Wolfovitz [Wol11] considered the question of fixed-size subgraphs of the triangle-free process: He showed that when the number of edges added is $cn^{3/2}\sqrt{\log n}$ for a small constant c, the counts of certain fixed-size subgraphs behave like they would in a random graph of the same edge density. In 2020, Bohman and Keevash [BK20] and Fiz Pontiveros, Griffiths, and Morris [PGM20] found the exact first-order asymptotic for both the number of edges and independence number of the graph produced by the triangle-free process.

In this writeup, we'll describe Kim's semirandom construction and analysis from [Kim95]. We'll then describe the ideas Bohman [Boh09] and Wolfovitz [Wol11] use to analyze the actual triangle-free process (we won't prove their main theorems because of length, but we'll see many of the key lemmas and proof techniques).

§1.2 Heuristics

First, here's a very loose heuristic for how we might expect the triangle-free process to behave, and why we might expect that it produces a graph with $\alpha(G) = O(\sqrt{n \log n})$.

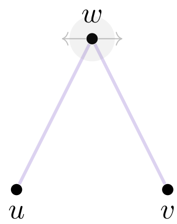
Imagine we embed the triangle-free process into time such that each step takes time $2n^{-3/2}$; this roughly means that in a time-interval of length γ (think of γ as small), we've sampled a $\gamma n^{-1/2}$-fraction of edges (in some random order) and attempted to add them to the graph. We'll refer to edges as:

- *Chosen* if they've already been added to the graph.
- *Open* if they haven't been added, but could be added without creating a triangle.
- *Closed* if they haven't been added, and adding them would create a triangle. (We can essentially forget about such edges — so in a time-interval of length γ, we can imagine that we're really sampling a $\gamma n^{-1/2}$-fraction of *open* edges.)

(In all pictures, we'll draw chosen edges in light purple and open edges in dark purple.)

Imagine that at time t, our chosen edge density is roughly $\Psi(t)n^{-1/2}$; we'd like to come up with an equation that predicts what Ψ should be.

First, what's the *open* edge density? An edge uv is open if it's not in a configuration of the following form (i.e., there is no w such that uw and vw have both been chosen).



There are roughly n choices for w, and the chosen edge density is $\Psi(t)n^{-1/2}$; so if we imagine that the graph has nice 'independence' properties, we might expect uv to be open with probability

$$(1 - \Psi(t)n^{-1})^n \approx \exp(-\Psi(t)^2).$$

Then in a short time-interval of length γ, we're sampling and adding each open edge with probability $\gamma n^{-1/2}$ (this isn't exactly true — adding one open edge could close another — but if we think of γ as small, this won't have much effect), so we'd expect our edge density to increase by roughly $\exp(-\Psi(t)^2) \cdot \gamma n^{-1/2}$. This suggests $\Psi(t + \gamma) \approx \Psi(t) + \exp(-\Psi(t)^2) \cdot \gamma$, giving the differential equation

$$\Psi'(t) = \exp(-\Psi(t)^2) \tag{1.1}$$

(with $\Psi(0) = 0$). This doesn't have an explicit solution, but we can write it as $\int_0^{\Psi(t)} e^{\xi^2}\, d\xi = t$, which means that $\Psi(t) \approx \sqrt{\log t}$ (at least, when t is reasonably large). Throughout the rest of this writeup, we'll use Ψ to refer to the solution to (1.1), and we'll write $\psi(t) = \Psi'(t)$. Since $\Psi(t) \approx \sqrt{\log t}$ (for reasonably large t), we have $\psi(t) = \exp(-\Psi(t)^2) \approx t^{-1}$.

Now let's consider independent sets. For any set of k vertices, we'd expect the number of open edges in the set at time t to be $\psi(t)\binom{k}{2}$. In a time-interval of length γ, we're picking each with probability $\gamma n^{-1/2}$, so the probability we don't pick any of them would be

$$(1 - \gamma n^{-1/2})^{\psi(t)\binom{k}{2}} \approx \exp\left(-\gamma\psi(t) \cdot n^{-1/2}\binom{k}{2}\right).$$

Then the probability this set remains independent up to time t should be

$$\exp\left(-\sum_{i=0}^{t/\gamma} \gamma\psi(\gamma i) \cdot n^{-1/2}\binom{k}{2}\right) \approx \exp\left(-\int_0^t \psi(t)\, dt \cdot n^{-1/2}\binom{k}{2}\right) = \exp\left(-\Psi(t) \cdot n^{-1/2}\binom{k}{2}\right) \tag{1.2}$$

(since we need this event to occur on each time-interval $[i\gamma, (i + 1)\gamma]$ up to t).

Meanwhile, the *number* of sets of size k is roughly $n^k = \exp(k \log n)$. So if $\Psi(t) \cdot n^{-1/2}\binom{k}{2}$ is large compared to $k \log n$, then a union bound should tell us that there are no independent sets of size k.

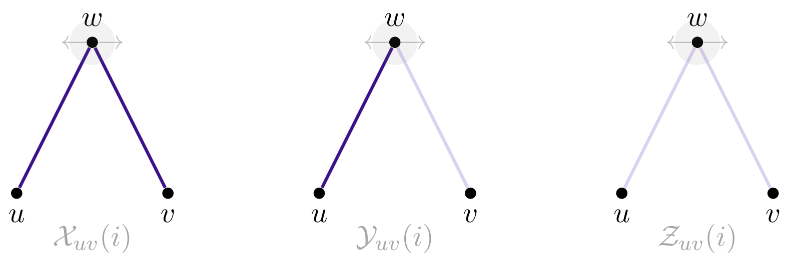
Since $\Psi(t) \approx \sqrt{\log t}$, if we take $t = n^c$ and $k = C\sqrt{n \log n}$ (where c is some small constant, and C is a large constant relative to c), then this inequality holds. So this suggests that if we can show the triangle-free process follows this heuristic up to time $t = n^c$, then we'll be able to prove $\alpha(G) = O(\sqrt{n \log n})$.

§1.3 Tracking configurations and an expected trajectory

The heuristics in Subsection 1.2 give us a guess for how we might expect the triangle-free process to evolve, but it's not clear how to use it to get a proof. We'll now discuss a more refined heuristic which is the starting point for all the proofs.

We'll consider three types of 'configurations' in the current graph G_i:

- We write $\mathcal{X}_{uv}(i)$ for the set of $\{uw, vw\}$ such that both uw and vw are open (so $\mathcal{X}_{uv}(i)$ is a set of pairs of edges). We write $X_{uv}(i) = |\mathcal{X}_{uv}(i)|$.
- We write $\mathcal{Y}_{uv}(i)$ for the set of $\{uw, vw\}$ such that one of uw and vw is open, and the other is chosen. We write $Y_{uv}(i) = |\mathcal{Y}_{uv}(i)|$.
- We write $\mathcal{Z}_{uv}(i)$ for the set of $\{uw, vw\}$ such that both uw and vw are chosen, and $Z_{uv}(i) = |\mathcal{Z}_{uv}(i)|$.



Imagine that at time t, we have that $X_{uv}(i) \approx \varphi_x(t)n$ and $Y_{uv}(i) \approx \varphi_y(t)\sqrt{n}$ for all open edges uv, and that $Z_{uv}(i)$ is small (where φ_x and φ_y are some functions to be determined). The key idea is that this information alone is enough to determine what $\varphi_x(t+\gamma)$ and $\varphi_y(t+\gamma)$ 'should' be (for small γ). First, if uv is currently open, then the probability it gets closed during $[t, t+\gamma]$ should be roughly

$$Y_{uv}(i) \cdot \gamma n^{-1/2} \approx \gamma\varphi_y(t)$$

(since uv gets closed if we add the one open edge in some configuration in $\mathcal{Y}_{uv}(i)$ — it could also get closed if we add both open edges in some configuration in $\mathcal{X}_{uv}(i)$, but if γ is small then this is substantially less likely). This means that for each of the roughly $\varphi_x(t)n$ configurations in $\mathcal{X}_{uv}(i)$, the probability it 'leaves' (meaning that one of its edges becomes closed) should be roughly $2\gamma\varphi_y(t)$; this suggests that we should have

$$\varphi_x'(t) = -2\varphi_x(t)\varphi_y(t). \tag{1.3}$$

(Some configurations will also leave $\mathcal{X}_{uv}(i)$ because we chose one of their edges; but the probability of an edge becoming closed is much bigger than the probability of it being chosen — they're roughly $\gamma\varphi_y(t)$ (which we think of as $n^{-\varepsilon}$ for small ε) and $\gamma n^{-1/2}$, respectively — so this can be ignored.)

Meanwhile, there's two main factors that contribute to changes in $Y_{uv}(i)$. First, some configuration in $\mathcal{Y}_{uv}(i)$ could 'leave' because we closed one of its open edges; there are roughly $\varphi_y(t)\sqrt{n}$ such configurations, and each leaves with probability roughly $\gamma\varphi_y(t)$, so we'd expect to lose $\gamma\varphi_y(t)^2\sqrt{n}$ configurations this way. Meanwhile, configurations in $\mathcal{X}_{uv}(i)$ could 'enter' because we chose one of their two open edges; we'd expect to gain $\varphi_x(t)n \cdot 2\gamma n^{-1/2} = 2\gamma\varphi_x(t)\sqrt{n}$ configurations this way. This suggests

$$\varphi_y'(t) = -\varphi_y(t)^2 + 2\varphi_x(t). \tag{1.4}$$

The heuristics in Subsection 1.2 give the guess $\varphi_x(t) = \psi(t)^2$ and $\varphi_y(t) = 2\Psi(t)\psi(t)$, and these functions indeed satisfy (1.3) and (1.4).

§1.4 Overview

Subsection 1.3 gives an idea for how we might analyze the triangle-free process: We can try splitting time into intervals of length γ, assuming that $X_{uv}(i)$, $Y_{uv}(i)$, and $Z_{uv}(i)$ have roughly followed their expected trajectory up to some time-interval, and showing that they exhibit good concentration around their expectations for the next one (in which case they'll continue to follow the expected trajectory).

However, proving concentration is difficult because of how non-independent the process is. So Kim's approach in [Kim95] is to modify the triangle-free process to a semi-random variant where in each time-interval, edges *are* essentially independent: In each time-interval, instead of running the triangle-free process (where we choose a $\gamma n^{-1/2}$-fraction of the open edges, then go through them in order and only add the ones that wouldn't create triangles), we choose a $\gamma n^{-1/2}$-fraction of the open edges and use *all* of them to update the sets $\mathcal{X}_{uv}(i)$, $\mathcal{Y}_{uv}(i)$, and $\mathcal{Z}_{uv}(i)$, and use the alteration method to get rid of triangles when building G_i. This has much better independence properties, which makes it possible to use tools like the bounded differences inequality to prove concentration.

To prove statements about the triangle-free process itself, though, we need some way of handling this non-independence. Bohman [Boh09] and Wolfovitz [Wol11] take fairly different approaches to this. Wolfovitz's idea is to combine a semi-random approach (where we split time into length-γ intervals and try to prove concentration on each) with *branching processes* that model how different edges affect each other, in a way that lets us say there isn't 'too much' dependence within a round (so that tools like the bounded differences inequality still work).

Meanwhile, instead of splitting time into chunks, Bohman analyzes the process step by step. Of course we can't say the process concentrates around its expectation on a single step (that doesn't really mean anything). Instead, the key idea is to define *martingales* that accumulate how much each step differs from the expected trajectory. We then use martingale concentration inequalities to show that these martingales remain small with high probability, which means the process always remains close to its expected trajectory. (This technique is called the *differential equation method*, and it's really powerful — in particular, [BK20] and [PGM20] both use a much more intricate version of the differential equation method to analyze the triangle-free process all the way to its end.)

In Section 2, we'll explain Kim's construction and its full analysis (in particular, we'll prove it produces a graph with $\alpha(G) = O(\sqrt{n \log n})$). In Section 3, we'll explain the portion of Wolfovitz's argument that proves the triangle-free process follows its expected trajectory (as described in Subsection 1.3), and in Section 4, we'll explain the portion of Bohman's argument that proves this. (Bohman uses a slightly different, and probably more natural, parametrization of the triangle-free process where at each step we choose an *open* edge and add it (rather than choosing an edge and adding it if it's open). This is equivalent in terms of the final graph produced, but the trajectory and differential equations for how it evolves over time are a bit different. I'll describe a version of his analysis for the parametrization described here (so that the process still has the trajectory described in Subsection 1.3) to make the similarities with the other two arguments easier to see.)

§2 Kim 1995: A semi-random construction for $r(3,k)$

In this section, we'll describe Kim's proof of the following theorem from [Kim95] (which uses a semi-random modification of the triangle-free process).

Theorem 2.1 (Kim 1995)
For all (sufficiently large) n, there exists a n-vertex graph G with $\alpha(G) = O(\sqrt{n \log n})$.

§2.1 Setup

Let $\varepsilon > 0$ and $c > 0$ be small constants (we'll assume c is small with respect to ε for convenience, but this isn't necessary), let C be a large constant with respect to c, and let $\gamma = n^{-\varepsilon}$. We're going to construct a sequence of graphs G_0, G_1, ..., where we think of G_i as a semi-random modification of the triangle-free process up to time γi. We'll run up to time n^c, which means i goes up to n^c/γ, and our target bound on $\alpha(G)$ will be $C\sqrt{n \log n}$.

We'll maintain the following objects, for each i:

- A collection of edges E_i (which may contain some triangles), which we'll call the *chosen* edges.
- A triangle-free graph $G_i \subseteq E_i$.
- A collection of edges O_i, which we'll call the *open* edges, such that there are no triangles consisting of one edge in O_i and two in E_i.

We define $\mathcal{X}_{uv}(i)$, $\mathcal{Y}_{uv}(i)$, and $\mathcal{Z}_{uv}(i)$ as in Subsection 1.3 with respect to these chosen and open edges. (The idea is that we'll construct E_i with nice independence properties, and we'll construct G_i by deleting triangles from E_i using the alteration method. This means we want to work with E_i instead of G_i when defining the tracked sets — these nice independence properties will let us prove concentration.)

For convenience, we define $q_i = \psi(\gamma i)$ and $p_i = \sum_{j=0}^{i-1} \gamma q_i$ (so p_i is a discrete approximation to $\Psi(\gamma i)$ — in particular, we have $\Psi(\gamma i) \le p_i \le \Psi(\gamma i) + \gamma$). Intuitively, the way to think about how these scale is that $n^{-\varepsilon} \le q_i \le 1$ and $n^{-\varepsilon} \le p_i \le c\sqrt{\log n}$, and $p_i q_i$ and $p_i^2 q_i$ are bounded (e.g., they're at most $\frac{1}{2}$ — this is because $\psi(t) = \exp(-\Psi(t)^2)$, and xe^{-x^2} is bounded).

We'll want these objects to satisfy the following eight properties. The first three properties correspond to saying that the construction follows the trajectory described in Subsection 1.3 (except that we only state upper bounds), and will be used to prove each other.

Property 2.2. For all edges $uv \notin E_i$, we have $X_{uv}(i) \le q_i^2 n$.

Property 2.3. For all edges $uv \notin E_i$, we have $Y_{uv}(i) \le 2(p_i + 8\gamma) q_i \sqrt{n}$.

Property 2.4. For all edges uv, we have $Z_{uv}(i) \le i(\log n)^2$.

We'll also need two more properties saying that the degrees of individual vertices are roughly what we'd expect, which will be used for the analysis of independent sets.

Property 2.5. For all vertices v, we have $\deg_{O_i}(v) \le q_i n$.

Property 2.6. For all vertices v, we have $\deg_{E_i}(v) \le p_i \sqrt{n} + i n^{1/3}$.

Next, we'll need two properties about the number of open edges in *big* sets.

Property 2.7. Let $k_0 = \gamma^2 q_i^2 \sqrt{n}$. Then the number of open edges between any two disjoint sets of sizes (exactly) k_0 is at most $q_i k_0^2$.

Property 2.8. Let $k = C\sqrt{n \log n}$. Then the number of open edges inside every set of size k is at least $q_i(1 - 64 p_i \gamma)\binom{k}{2} - 16 p_i q_i k \sqrt{n}$.

These bounds make sense because our heuristic (based on Subsection 1.2) is that the density of open edges should be roughly q_i. Note that the error terms in Property 2.8 are 'reasonable' — we always have $p_i \le c\sqrt{\log n}$ and we set $\gamma = n^{-\varepsilon}$, so $p_i \gamma \ll 1$ and $p_i k \sqrt{n}$ goes up to a tiny constant fraction of $\binom{k}{2}$.

Since we care about independent sets, we only really care about a *lower* bound on the number of open edges in big sets; but Property 2.7 will be needed for technical reasons when proving Property 2.8. (The choice of k_0 in Property 2.7 is not super important for the proof — the proof would work for anything on the scale of $\sqrt{n}$ — and it's there just because that's what we'll need for Property 2.8.)

Finally, the last property is about independent sets and roughly corresponds to (1.2), up to constants (this is the only property that looks at G_i).

Property 2.9. The number of independent sets in G_i of size $k = C\sqrt{n \log n}$ is at most

$$n^k \cdot \exp\left(-\frac{1}{64} p_i n^{-1/2} \binom{k}{2}\right).$$

To prove Theorem 2.1, we'll show that given (E_i, G_i, O_i) which satisfy these properties, we can construct $(E_{i+1}, G_{i+1}, O_{i+1})$ which still satisfy these properties (with i replaced by $i+1$), as long as $i \leq n^c/\gamma$. Once we reach $i = n^c/\gamma$, Property 2.9 says that the number of size-k independent sets is at most

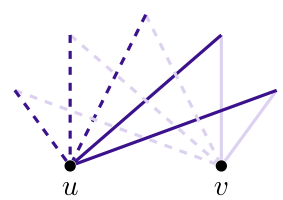
$$\exp\left(C\sqrt{n \log n} \cdot \log n - \frac{1}{64} \cdot c\sqrt{\log n} \cdot n^{-1/2} \cdot \binom{C\sqrt{n \log n}}{2}\right) < 1,$$

which means none exist (proving Theorem 2.1).

§2.2 The construction

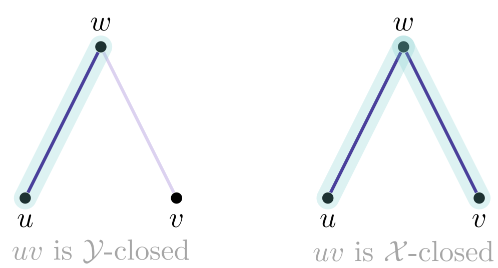
We'll now describe how to construct $(E_{i+1}, G_{i+1}, O_{i+1})$ from (E_i, G_i, O_i).

We first perform a 'regularization' step to make Property 2.3 an equality: For every $uv \notin E_i$, we introduce some dummy vertices w and add dummy edges uw to O_i and vw to E_i so that equality holds in Property 2.3. (We're not going to include these in the output; their purpose is only so that we have two-sided control on the probability that uv gets closed.)

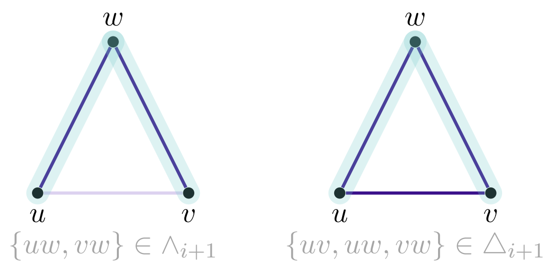


Then we sample a subset of O_i where we include each edge independently with probability $\gamma n^{-1/2}$. We'll refer to these edges as *new*, and we'll depict them by highlighting them in blue. We let F_{i+1} be the set of real (i.e., non-dummy) edges which we picked, and we set $E_{i+1} = E_i \cup F_{i+1}$.

Next, we need to define O_{i+1} by removing edges which got closed this round from O_i. We say that uv gets *$\mathcal{Y}$-closed* if there is some configuration in $\mathcal{Y}_{uv}(i)$ (either real or dummy) for which we picked its open edge this round, and *$\mathcal{X}$-closed* if there is some configuration in $\mathcal{X}_{uv}(i)$ for which we picked both its open edges this round. We define O_{i+1} by removing all such edges from O_i.



Finally, we'll define a subset $F^*_{i+1} \subseteq F_{i+1}$ to add to G_i by using the alteration method to get rid of triangles. First, since $F_{i+1} \subseteq O_i$, there are no edges in F_{i+1} which would form a triangle with two edges in G_i; so we only need to worry about triangles with two or three edges from F_{i+1}. So we define $\wedge_{i+1}$ as the set of pairs of edges $\{uw, vw\} \subseteq F_{i+1}$ for which $uv \in G_i$, and $\triangle_{i+1}$ as the set of triangles in F_{i+1}.



We find a maximal edge-disjoint subset of $\wedge_{i+1} \cup \triangle_{i+1}$ and remove it from F_{i+1} to produce F^*_{i+1}, and set $G_{i+1} = G_i \cup F^*_{i+1}$. (Maximality means that we've removed at least one edge from each configuration in $\wedge_{i+1} \cup \triangle_{i+1}$, so adding F^*_{i+1} to G_i doesn't create any triangles.)

The intuition is that the construction of E_{i+1} and O_{i+1} roughly corresponds to how the triangle-free process works on the time-interval $[\gamma i, \gamma(i+1)]$ as described in Subsections 1.2 and 1.3, so we'd expect the same heuristics to hold (in fact, it corresponds to our description better than the triangle-free process itself does). Meanwhile, we'd expect that the alterations used to produce F^*_{i+1} shouldn't have too much effect — an open edge uv will be included in F_i with probability $\gamma n^{-1/2}$, while it'll be in a configuration in $\wedge_{i+1}$ with probability at most

$$Y_{uv}(i) \cdot (\gamma n^{-1/2})^2 \leq 2(p_i + 8\gamma)q_i\sqrt{n} \cdot \gamma^2 n^{-1} \leq \gamma^2 n^{-1/2}$$

by Property 2.3 (this requires us to choose both uv and the open edge of some configuration in $\mathcal{Y}_{uv}(i)$), and it'll be in a configuration in $\triangle_{i+1}$ with probability at most

$$Z_{uv}(i) \cdot (\gamma n^{-1/2})^3 \leq \gamma^3 n^{-1/2}.$$

Since $\gamma \ll 1$, the fraction of edges in F_{i+1} that we discard should be tiny; this means the same heuristic (1.2) for independent sets in the triangle-free process should apply to G_{i+1}.

Now we'll perform the analysis, meaning we'll prove that Properties 2.2–2.9 all hold with high probability. (Whenever we refer to an edge as open, closed, or chosen, unless otherwise specified we're referring to its status after the ith round, not after the $(i+1)$st round.)

§2.3 Probabilities of closing edges

Nearly all the properties are about the density of open edges, so as a first step, we'll estimate the probability that any given open edge uv gets closed. First note that intuitively, the probability it's $\mathcal{Y}$-closed should be substantially larger than the probability it's $\mathcal{X}$-closed: The probability it's $\mathcal{Y}$-closed is roughly

$$Y_{uv}(i) \cdot \gamma n^{-1/2} = 2\gamma(p_i + 8\gamma)q_i$$

(there are $Y_{uv}(i)$ configurations in $\mathcal{Y}_{uv}(i)$, and uv being $\mathcal{Y}$-closed means that we pick the one open edge in such a configuration; the union-bound approximation is 'reasonable' because $2\gamma(p_i + 8\gamma)q_i \ll 1$), while the probability it's $\mathcal{X}$-closed is roughly

$$Z_{uv}(i) \cdot (\gamma n^{-1/2})^2 \leq \gamma^2 q_i^2$$

(which is significantly smaller because γ is small). So when we're proving upper bounds (i.e., everywhere except Property 2.8), we're only going to focus on decreases that come from edges being $\mathcal{Y}$-closed, and we'll ignore the effect of edges being $\mathcal{X}$-closed.

For this reason, it'll be useful to have the following calculation.

Claim 2.10 — For every open edge uv, we have

$$\frac{q_{i+1}}{q_i} - 24\gamma q_i \le \mathbb{P}[uv \text{ is not } \mathcal{Y}\text{-closed}] \le \frac{q_{i+1}}{q_i} - 8\gamma q_i.$$

(Intuitively, this makes sense because we want to say that our open edge density drops from roughly q_i to roughly q_{i+1}.)

Proof. By construction, there are exactly $2(p_i + 8\gamma)q_i\sqrt{n}$ edges which would $\mathcal{Y}$-close uv, and we pick each with probability $\gamma n^{-1/2}$. So $\mathbb{P}[uv \text{ is not } \mathcal{Y}\text{-closed}] = (1 - \gamma n^{-1/2})^{2(p_i+8\gamma)q_i\sqrt{n}}$, which we can bound by

$$1 - 2\gamma(p_i + 8\gamma)q_i \le \mathbb{P}[uv \text{ is not } \mathcal{Y}\text{-closed}] \le 1 - 2\gamma(p_i + 8\gamma)q_i + 2\gamma^2(p_i + 8\gamma)^2 q_i^2.$$

We can check that $(p_i + 8\gamma)^2 q_i \le 1$, so we can simplify this to

$$1 - 2\gamma p_i q_i - 16\gamma^2 q_i^2 \le \mathbb{P}[uv \text{ is not } \mathcal{Y}\text{-closed}] \le 1 - 2\gamma p_i q_i - 14\gamma^2 q_i^2. \tag{2.1}$$

Now, the intuition is that

$$\frac{q_{i+1} - q_i}{\gamma} = \frac{\psi((i+1)\gamma) - \psi(i\gamma)}{\gamma} \approx -\psi'(i\gamma) = -2\Psi(i\gamma)\psi(i\gamma)^2 \approx -2p_i q_i^2 \tag{2.2}$$

(the second-last equality follows from our differential equation for Ψ), which rearranges to

$$1 - 2\gamma p_i q_i \approx \frac{q_{i+1}}{q_i}.$$

The formal proof follows from tracking the errors in the above approximations. We can bound the error in the first approximation of (2.2) by $\gamma \sup_{t\in[i\gamma,(i+1)\gamma]} |\psi''(t)| \le \gamma \cdot 4\psi(t)^2$. For the second approximation, we can bound $|\Psi(i\gamma) - p_i| \le \gamma$. This gives

$$\left|1 - 2\gamma p_i q_i - \frac{q_{i+1}}{q_i}\right| \le 6\gamma q_i^2;$$

and plugging this into (2.1) gives the desired bounds. □

§2.4 Preliminaries: concentration bounds

We'll use the following two concentration bounds.

Lemma 2.11 (Multiplicative Chernoff)

Let X be a sum of independent Bernoulli random variables with $\mathbb{E}[X] = \mu$. Then for all $\beta > 0$, we have

$$\mathbb{P}[X \ge (1+\beta)\mu] \le e^{-\min\{\beta^2\mu/3, \beta\mu/3\}} \quad \text{and} \quad \mathbb{P}[X \le (1-\beta)\mu] \le e^{-\beta^2\mu/2}.$$

Lemma 2.12

Let $f : \{0,1\}^n \to \mathbb{R}$ be a function with the property that for each j, changing the jth coordinate of x changes $f(x)$ by at most c_j. Then for X sampled according to the p-biased measure on $\{0,1\}^n$ (i.e., with independent $\mathsf{Ber}(p)$ coordinates), for all $\sigma > 0$ we have

$$\mathbb{P}[|f(X) - \mathbb{E}[f(X)]| \ge \sigma] \le 2\exp\left(-\frac{1}{8}\min\left\{\frac{\sigma}{\max_j c_j}, \frac{\sigma^2}{p\sum_j c_j^2}\right\}\right).$$

Lemma 2.12 is essentially a version of the bounded differences inequality for low-probability Bernoullis (and it can be proved by combining the ideas of the proof of the ordinary bounded differences inequality, or Azuma–Hoeffding, with the bounds on moment generating functions used to prove multiplicative Chernoff).

§2.5 Tracking configurations

In this section, we'll prove Properties 2.2, 2.3, and 2.4. We're always going to consider a fixed edge uv and show that the property holds for that edge with probability $1 - n^{-\omega(1)}$, so we can union-bound over edges. (In the proofs of concentration, we'll always have room to spare — we'll end up with concentration probabilities that look like $1 - \exp(-\Omega(n^a))$ for constants a (e.g., $\frac{1}{2}$ or $\frac{1}{4}$) — so factors of p_i, q_i, and γ will not matter, and we'll typically bound them by factors of n^ε to avoid carrying them around.)

§2.5.1 Property 2.2: Tracking $\mathcal{X}_{uv}(i)$

For Property 2.2, the 'main' term driving the change in $X_{uv}(i)$ is configurations leaving $\mathcal{X}_{uv}(i)$ because we $\mathcal{Y}$-closed one of their edges. So we define $\mathcal{X}_{uv}^-(i+1)$ as the set of such configurations, and write $X_{uv}^-(i+1) = |\mathcal{X}_{uv}^-(i+1)|$; then we have

$$X_{uv}(i+1) \le X_{uv}(i) - X_{uv}^-(i+1).$$

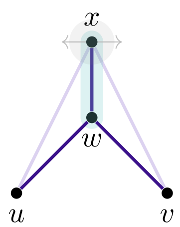
We'll show the right-hand side has expected value a bit under $q_{i+1}n^2$, and $X_{uv}^-(i+1)$ concentrates well. For the expected value, we'll need the following calculation.

Claim 2.13 — For every configuration $\{uw, vw\} \in \mathcal{X}_{uv}(i)$, we have

$$\mathbb{P}[\text{neither } uw \text{ nor } vw \text{ is } \mathcal{Y}\text{-closed}] \le \frac{q_{i+1}^2}{q_i^2} - 8\gamma q_i.$$

This makes intuitive sense, since as we saw in Claim 2.10, the probability that a *single* edge y-survives is a bit less than $\frac{q_{i+1}}{q_i}$. (The precise error term in this claim isn't important.)

Proof. There are exactly $2(p_i + 8\gamma)q_i\sqrt{n}$ edges which would $\mathcal{Y}$-close uw, and the same is true for vw. Furthermore, if an edge $e = wx$ falls into both cases, then we must have $\{ux, vx\} \in \mathcal{Z}_{uv}(i)$. By Property 2.4, this means there are at most $i(\log n)^2 \le n^{2\varepsilon}$ such e.



So there are at least $4(p_i + 8\gamma)q_i\sqrt{n} - n^{2\varepsilon} \ge 4(p_i + 7\gamma)q_i\sqrt{n}$ edges e which would $\mathcal{Y}$-close uw or vw. And we choose each with probability $\gamma n^{-1/2}$, so

$$\mathbb{P}[\text{neither } uw \text{ nor } vw \text{ is } \mathcal{Y}\text{-closed}] \le (1 - \gamma n^{-1/2})^{4(p_i+7\gamma)q_i\sqrt{n}} = \left((1 - \gamma n^{-1/2})^{2(p_i+7\gamma)q_i\sqrt{n}}\right)^2.$$

The same calculation as in the proof of Claim 2.10 shows that $(1 - \gamma n^{-1/2})^{2(p_i+7\gamma)q_i\sqrt{n}} \le \frac{q_{i+1}}{q_i} - 6\gamma q_i$, giving

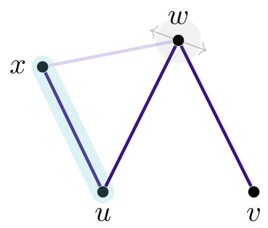
$$\mathbb{P}[\text{neither } uw \text{ nor } vw \text{ is } \mathcal{Y}\text{-closed}] \le \left(\frac{q_{i+1}}{q_i} - 6\gamma q_i\right)^2 \le \frac{q_{i+1}^2}{q_i^2} - 8\gamma q_i. \qquad \square$$

Since $X_{uv}(i) \le q_i^2 n$, this means we have

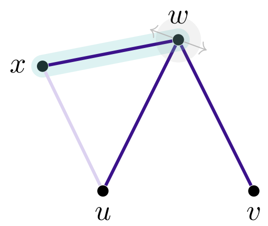
$$\mathbb{E}[X_{uv}(i) - X_{uv}^-(i+1)] \le q_i^2 n \cdot \left(\frac{q_{i+1}^2}{q_i^2} - 8\gamma q_i\right) \le q_{i+1}^2 n - 8\gamma q_i^3 n.$$

We wanted an upper bound of $q_{i+1}^2 n$, so it suffices to prove that $X_{uv}^-(i+1)$ concentrates within a window of length $n^{1-5\varepsilon}$. We'll do so using Lemma 2.12. We're picking each edge e independently with probability $p = \gamma n^{-1/2}$; to bound the c_e's, we want to bound how many configurations in $\mathcal{X}_{uv}(i)$ each edge can affect.

Case 1 (e is incident to u or v). We'll assume without loss of generality that e is incident to u; let $e = ux$. Then for ux to affect a configuration $\{uw, vw\} \in \mathcal{X}_{uv}(i)$ (i.e., to $\mathcal{Y}$-close one of its edges), we must have $\{uw, xw\} \in \mathcal{Y}_{ux}(i)$; by Property 2.3, this means there are at most $\sqrt{n}$ choices for w. Also, there are at most $2n$ edges incident to u or v; so this gives at most $2n$ edges each with $c_e \leq \sqrt{n}$.



Case 2 (e is not incident to u or v). In this case, there's only one configuration in $\mathcal{X}_{uv}(i)$ that e can affect (namely, the one corresponding to one of its endpoints). So this gives at most n^2 edges each with $c_e \leq 1$.



This means when we apply Lemma 2.12 with $\sigma = n^{1-5\varepsilon}$ and $p = \gamma n^{-1/2} \leq n^{-1/2}$, we'll have

$$\frac{\sigma}{\max_e c_e} \geq \frac{n^{1-5\varepsilon}}{\sqrt{n}} \geq n^{1/2-5\varepsilon} \quad \text{and} \quad \frac{\sigma^2}{p\sum_e c_e^2} \geq \frac{n^{2-10\varepsilon}}{n^{-1/2}(2n \cdot n + n^2 \cdot 1)} \geq n^{1/2-10\varepsilon}.$$

This shows $X_{uv}^-(i+1)$ concentrates in a window of length $n^{1-5\varepsilon}$ with probability $1 - \exp(-\Omega(n^{1/2-10\varepsilon}))$ for each edge uv; this is more than good enough to union-bound over all uv (it would suffice to have any positive exponent of n).

§2.5.2 Property 2.3: Tracking $\mathcal{Y}_{uv}(i)$

There are two 'main' terms which should drive the change in $Y_{uv}(i)$, corresponding to the two terms in (1.4):

- We define $\mathcal{Y}_{uv}^-(i+1)$ as the set of configurations in $\mathcal{Y}_{uv}(i)$ that leave because we $\mathcal{Y}$-closed their one open edge, and $Y_{uv}^-(i+1)$ as its size.
- We define $\mathcal{Y}_{uv}^+(i+1)$ as the set of configurations in $\mathcal{Z}_{uv}(i)$ which enter because we picked one of their open edges, and $Y_{uv}^+(i+1)$ as its size.

Then we have

$$Y_{uv}(i+1) \leq Y_{uv}(i) - Y_{uv}^-(i+1) + Y_{uv}^+(i+1).$$

We'll show the expectation of the right-hand side is a bit under $2(p_{i+1} + 8\gamma)q_{i+1}\sqrt{n}$ (our target bound) — the fact that this computation works out essentially corresponds to (1.4) — and that both $Y_{uv}^-(i+1)$ and $Y_{uv}^+(i+1)$ concentrate well.

First, by Claim 2.10 (and Property 2.3) we have

$$\mathbb{E}[Y_{uv}(i) - Y_{uv}^-(i+1)] \le Y_{uv}(i) \cdot \left(\frac{q_{i+1}}{q_i} - 8\gamma q_i\right) \le 2(p_i + 8\gamma)q_{i+1}\sqrt{n} - 128\gamma^2 q_i^2\sqrt{n}.$$

Meanwhile, each configuration in $\mathcal{X}_{uv}(i)$ lands in $\mathcal{Y}_{uv}^+(i+1)$ with probability at most $2\gamma n^{-1/2}$, so

$$\mathbb{E}[Y_{uv}^+(i+1)] \le Z_{uv}(i) \cdot 2\gamma n^{-1/2} \le 2\gamma q_i^2\sqrt{n}.$$

This means in total, we get

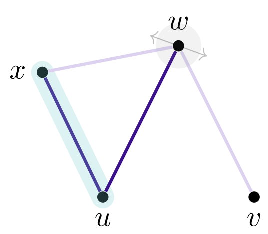
$$\mathbb{E}[Y_{uv}(i) - Y_{uv}^-(i+1) + Y_{uv}^+(i+1)] \le 2((p_i + 8\gamma)q_{i+1} + \gamma q_i^2)\sqrt{n} - 128\gamma^2 q_i^2\sqrt{n}.$$

Our goal was to get a bound of $2(p_{i+1} + 8\gamma)q_{i+1}\sqrt{n}$. This is exactly what the first term would be if it had $\gamma q_i q_{i+1}$ instead of γq_i^2 (since $p_{i+1} = p_i + \gamma q_i$). We can show that $q_i - q_{i+1} \le \gamma q_i$ (since $-\psi(t) \le \psi'(t) \le 0$ for all t), so we get that

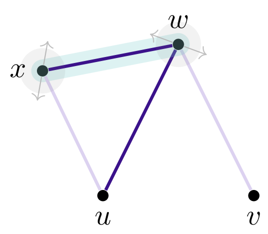
$$\mathbb{E}[Y_{uv}(i) - Y_{uv}^-(i+1) + Y_{uv}^+(i+1)] \le 2(p_{i+1} + 8\gamma)q_{i+1}^2 - 126\gamma^2 q_i^2\sqrt{n}.$$

Now it's enough to show $Y_{uv}^-(i+1)$ and $Y_{uv}^+(i+1)$ concentrate in windows of length $n^{1/2-5\varepsilon}$. For $Y_{uv}^-(i+1)$, we'll again use Lemma 2.12, which means we want to estimate the number of configurations in $\mathcal{Y}_{uv}(i)$ each edge e can affect (meaning that choosing e would $\mathcal{Y}$-close that configuration's open edge).

Case 1 (e is incident to u or v). Without loss of generality let $e = ux$. Then for e to affect some configuration $\{uw, vw\} \in \mathcal{Y}_{uv}(i)$, we must have $\{vw, xw\} \in \mathcal{Z}_{vx}(i)$. By Property 2.4 there are at most $i(\log n)^2 \le n^{2\varepsilon}$ choices for w. So this gives at most $2n$ edges e with $c_e \le n^{2\varepsilon}$.



Case 2 (e is not incident to u or v). Then e affects at most one configuration $\{uw, vw\} \in \mathcal{Y}_{uv}(i)$. Furthermore, for e to affect one configuration, we must have the picture below, where $e = wx$ with $w \in \mathcal{Y}_{uv}(i)$ and $x \in \mathcal{Y}_{uw}(i)$ (or $\mathcal{Y}_{vw}(i)$ — whichever of uw or vw is the open edge in $\{uw, vw\}$). By Property 2.3, there are at most $\sqrt{n}$ choices for w, and then at most $\sqrt{n}$ choices for x. So this gives at most n edges with $c_e \le 1$.

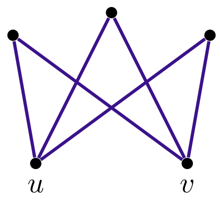


So when we apply Lemma 2.12 with $\sigma = n^{1/2-5\varepsilon}$ and $p = \gamma n^{-1/2} \le n^{-1/2}$, we'll have

$$\frac{\sigma}{\max_e c_e} \ge \frac{n^{1/2-5\varepsilon}}{n^{2\varepsilon}} \ge n^{1/2-7\varepsilon} \quad \text{and} \quad \frac{\sigma^2}{p\sum_e c_e^2} \ge \frac{n^{1-10\varepsilon}}{n^{-1/2}(2n \cdot n^{4\varepsilon} + n \cdot 1)} \ge n^{1/2-15\varepsilon}.$$

This means we get a concentration probability of $1 - \exp(-\Omega(n^{1/2-15\varepsilon}))$, which is certainly good enough.

Now we'll prove concentration for $Y^+_{uv}(i+1)$. Here we're considering all configurations in $\mathcal{Z}_i(uv)$ and seeing whether we choose one of its edges. Since these configurations are edge-disjoint, this means these events for different configurations are *independent*. This means $Y^+_{uv}(i+1)$ is a sum of independent Bernoullis, so it concentrates well by multiplicative Chernoff bounds — we've shown its mean is at most $2\gamma q_i^2\sqrt{n}$, so it concentrates within a window of length $n^{1/2-5\varepsilon}$ with very high probability.



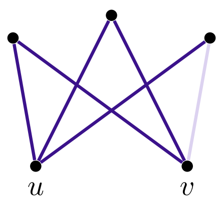
So we've shown that $Y^-_{uv}(i+1)$ and $Y^+_{uv}(i+1)$ both concentrate well, which finishes the proof.

§2.5.3 Property 2.4: Bounding $\mathcal{Z}_{uv}(i)$

We want to show that $\mathcal{Z}_{uv}(i)$ grows by at most $(\log n)^2$ on this round. There are two ways in which a configuration can move into $\mathcal{Z}_{uv}(i+1)$ — either it started in $\mathcal{Y}_{uv}(i)$ and we picked its one open edge, or it started in $\mathcal{X}_{uv}(i)$ and we picked both of its open edges. Then letting $Z^+_{uv}(i+1)$ be the number of such configurations, we have

$$\mathbb{E}[Z^+_{uv}(i+1)] \leq Y_{uv}(i)\cdot\gamma n^{-1/2} + X_{uv}(i)\cdot\gamma^2 n^{-1} \leq \sqrt{n}\cdot\gamma n^{-1/2} + n\cdot\gamma^2 n^{-1/2} \leq 1$$

(we're using Property 2.3 for the first term). And $Z^+_{uv}(i+1)$ is a sum of independent Bernoullis, since the configurations we're considering are edge-disjoint; so by multiplicative Chernoff, it's at most $(\log n)^2$ with probability $1-n^{-\omega(1)}$.



§2.6 Additional bounds on degrees

Next, we'll prove Properties 2.5 and 2.6. (Again we'll imagine fixing v, and we'll show the desired statements hold with probability $1-n^{-\omega(1)}$, so that we can union-bound.)

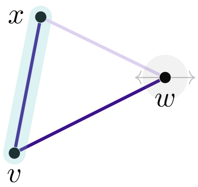
§2.6.1 Property 2.5: Open degrees

We can run a very similar argument to the proof of Property 2.2 — fix some vertex v. We start the round with at most $q_i n$ open edges incident to v, so by Claim 2.10, the expected number of them which are not $\mathcal{Y}$-closed this round (which is an upper bound on $\deg_{O_{i+1}}(v)$) is at most

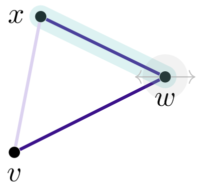
$$q_i n\cdot\left(\frac{q_{i+1}}{q_i} - 8\gamma q_i\right) = q_{i+1}n - 8\gamma q_i^2 n.$$

So to prove that this number is at most $q_{i+1}n$ (with high probability), it suffices to show that the number of edges incident to v which get $\mathcal{Y}$-closed concentrates in a window of length $n^{1-4\varepsilon}$. As usual, we'll do so using Lemma 2.12, which means we want to understand, for every edge e, how many edges incident to v get $\mathcal{Y}$-closed as a result of choosing e.

Case 1 (e is incident to v). Then letting $e = vx$, for e to affect an edge vw we need to have $\{vw, xw\} \in \mathcal{Y}_{vx}(i)$. By Property 2.3 there are at most $\sqrt{n}$ such choices for w, so this gives at most n edges with $c_e \le \sqrt{n}$.



Case 2 (e is not incident to v). Then e affects at most one edge incident to v (since we must have the below picture); so this gives at most n^2 edges e, each with $c_e \le 1$.



Then when we apply Lemma 2.12, we'll have

$$\frac{\sigma}{\max_e c_e} \ge \frac{n^{1-4\varepsilon}}{\sqrt{n}} \ge n^{1/2-4\varepsilon} \quad \text{and} \quad \frac{\sigma^2}{p\sum_e c_e^2} \ge \frac{n^{2-8\varepsilon}}{n^{-1/2}(2n \cdot n + n^2 \cdot 1)} \ge n^{1/2-9\varepsilon}.$$

So we get good concentration in the desired window, which finishes the proof.

§2.6.2 Property 2.6: Chosen degrees

We want to show that on this round, we choose at most $\gamma q_i \sqrt{n} + n^{1/3}$ new edges incident to v (this is how much the bound in Property 2.6 increases by when we go from i to $i+1$, since $p_{i+1} = p_i + \gamma q_i$). First, we're choosing each open edge with probability $\gamma n^{-1/2}$, so by Property 2.5, the *expected* number of edges incident to v we choose is at most $\gamma n^{-1/2} \cdot q_i n = \gamma q_i \sqrt{n}$. And these edges are independent, so we get good concentration in a window of length $n^{1/3}$ by multiplicative Chernoff bounds.

§2.7 Open edge densities in big sets

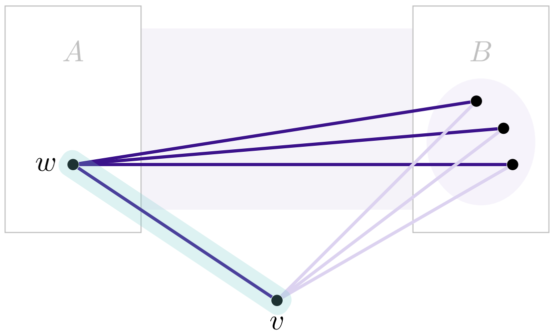
Next, we'll prove Properties 2.7 and 2.8.

§2.7.1 Property 2.7: Upper bound on open edge densities

First we'll illustrate an attempt that *doesn't* work, to motivate the actual proof.

Consider two disjoint sets A and B of size k_0. We begin the round with at most $q_i k_0^2$ open edges between them. As with the previous arguments, Claim 2.10 tells us that each edge is not $\mathcal{Y}$-closed with probability a bit under $\frac{q_{i+1}}{q_i}$, so we *expect* slightly under $q_{i+1}k_0^2$ open edges to remain. Since k_0 is roughly $\sqrt{n}$, we need to prove concentration in a window of size roughly n. Furthermore, we need to union-bound over all possible sets A and B; there are roughly $n^{k_0} = \exp(k_0 \log n)$ sets of size k_0, so for the union bound to work, we want a concentration probability of the form $1 - \exp(n^\alpha)$ for some $\alpha > \frac{1}{2}$.

Let's attempt to use Lemma 2.12 and see what it gives us. To do so, we want to understand, for each edge e, how many edges between A and B picking e this round would $\mathcal{Y}$-close.

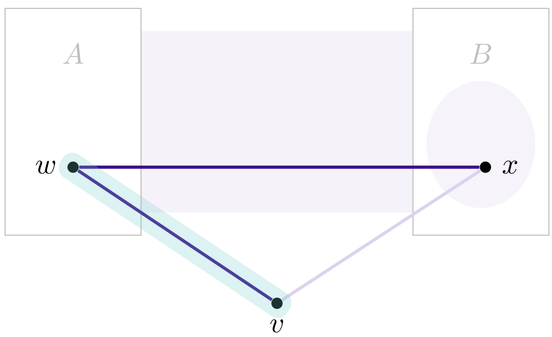


If $e = vw$ (where $v \notin A \cup B$ and $w \in A$), then this is essentially controlled by $\deg_{F_{i+1}}(v, B)$ (the number of chosen edges from v into B). However, this could potentially be as large as $k_0 \approx \sqrt{n}$. And there's potentially $k_0 n \approx n^{3/2}$ relevant edges. So if we try using Lemma 2.12 with $\sigma \approx n$ and $\max_e c_e \approx \sqrt{n}$, we'll have

$$\frac{\sigma}{\max_e c_e} \approx \frac{n}{\sqrt{n}} = \sqrt{n}$$

which means the best concentration probability we can hope for is roughly $1 - \exp(\sqrt{n})$; this isn't good enough to union-bound. (We're suppressing all the 'small' terms in this calculation — e.g., the q_i's and γ's — but they work against us.)

But although this fails, it just *barely* fails. Here we only looked at the first term in Lemma 2.12, but it turns out the second term only barely fails too. To handle that second term, it's hard to deal with $\sum_e c_e^2$, but we can bound this by $\max_e c_e \cdot \sum_e c_e$. And as seen in the above picture, $\sum_e c_e$ essentially counts configurations of the following shape (where e is the edge vw, and c_e is the number of choices for x).



If we instead count such configurations by first choosing w and x, then there are $k_0 \approx \sqrt{n}$ choices for w, $k_0 \approx \sqrt{n}$ choices for x, and at most $Y_{wx}(i) \approx \sqrt{n}$ choices for v; this tells us $\sum_e c_e$ is at most roughly $n^{3/2}$. So we get a bound of $\sum_e c_e^2 \leq \max_e c_e \cdot \sum_e c_e \leq \sqrt{n} \cdot n^{3/2} = n^2$, giving

$$\frac{\sigma}{p \sum_e c_e^2} \approx \frac{n^2}{n^{-1/2} \cdot n^2} = \sqrt{n}.$$

Again, we'd have been happy with any exponent of n strictly greater than $\frac{1}{2}$, but we got $\frac{1}{2}$.

The way we'll fix this is by ignoring vertices v for which $\deg_{E_i}(v, A)$ or $\deg_{E_i}(v, B)$ is too large. We'll show that there aren't too many such vertices (specifically, much fewer than $\sqrt{n}$), so that ignoring them doesn't hurt our probabilities of getting $\mathcal{Y}$-closed (as computed in Claim 2.10) by much; and with this, we'll be able to get good enough concentration.

To prove there aren't too many high-degree vertices, we'll use the following fact.

Fact 2.14 — Suppose that $r \geq 4\sqrt{st}$, and suppose that $A_1, \ldots, A_\ell$ are subsets of a set of size s such that $|A_i| \geq r$ for all i, and $|A_i \cap A_j| \leq t$ for all $i \neq j$. Then we must have $\ell \leq \frac{2s}{r}$ and $\sum_{i=1}^{\ell} |A_i| \leq 2s$.

Intuitively, if we have ℓ *disjoint* subsets of a set of size s, and each has size at least r, then there's at most $\frac{s}{r}$ of them and their sizes sum to at most s. Fact 2.14 essentially says that these statements continue to hold (up to a factor of 2) as long as the sets have small pairwise intersections. (We only need the first statement right now, but we'll use the second in the proof of Property 2.8.)

Proof. To prove that $\ell \leq \frac{2s}{r}$, assume not, and consider the first $\ell' = \frac{2s}{r}$ sets. By the principle of inclusion-exclusion, we have

$$|A_1 \cup \cdots \cup A_{\ell'}| \geq \sum_{i=1}^{\ell'} |A_i| - \sum_{i \neq j} |A_i \cap A_j| \geq \ell' r - \binom{\ell'}{2} t \geq 2s - \frac{2s^2 t}{r^2} > s,$$

which is a contradiction (since these sets live in a universe of size s).

Now to prove the second statement, for each i we have

$$|A_i| - \sum_{j \neq i} |A_i \cap A_j| \geq |A_i| - \frac{2st}{r} \geq |A_i| - \frac{r}{2} \geq \frac{1}{2} |A_i|,$$

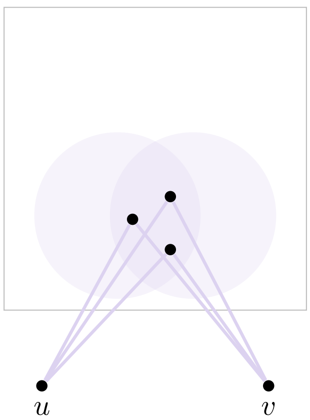
which means we can find a portion consisting of at least half of A_i which is disjoint from all the other sets A_j. Since these portions (over all i) are disjoint and live in a universe of size s, we get that $\frac{1}{2} \sum_{i=1}^{\ell} |A_i| \leq s$. □

Now we say a vertex v is *safe* with respect to A if $\deg_{E_i}(v, A) \leq n^{1/3}$ and *unsafe* otherwise; we define the same notions with respect to B. (The choice of $\frac{1}{3}$ is fairly arbitrary; any exponent strictly between $\frac{1}{4}$ and $\frac{1}{2}$ would work for this argument.)

Claim 2.15 — The number of unsafe vertices with respect to A is at most $n^{1/4}$.

Proof. For every unsafe vertex v, we can define the set $A_v = N_{E_i}(v) \cap A$. These sets A_v have size at least $n^{1/3}$ (by definition) and are subsets of A, which has size $k_0 \leq \sqrt{n}$. Furthermore, for any two unsafe vertices $u \neq v$, vertices in $A_u \cap A_v$ correspond to configurations in $\mathcal{Z}_{uv}(i)$, so by Property 2.4 we have

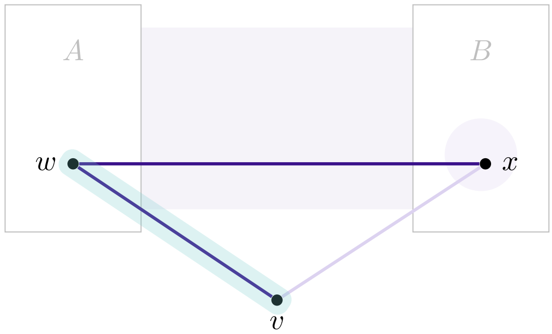
$$|A_u \cap A_v| \leq i(\log n)^2.$$



We have $n^{1/3} \geq 4\sqrt{k_0 \cdot i(\log n)^2}$, so Fact 2.14 applies and gives that the number of sets A_v we have (and therefore the number of unsafe vertices) is at most $\frac{2k_0}{i(\log n)^2} \leq n^{1/4}$. □

Similarly, at most $n^{1/4}$ vertices are unsafe with respect to B.

Now we say an edge wx between A and B gets *safely $\mathcal{Y}$-closed* if it gets $\mathcal{Y}$-closed by a configuration whose tip is safe with respect to both A and B. In other words, this means wx is open at the start of the round, and there is some v such that v is safe with respect to A and B, we have $\{vw, vx\} \in \mathcal{Y}_{wx}(i)$, and the one open edge in $\{vw, vx\}$ gets chosen this round.



Let Y be the number of open edges between A and B which *don't* get safely $\mathcal{Y}$-closed. Then Y is an upper bound on the number of open edges we'll have between A and B after this round; we'll show that it has the right mean and concentrates well.

First, for each open edge wx between A and B, we've seen that the probability that wx doesn't get $\mathcal{Y}$-closed by *any* vertex is

$$(1 - \gamma n^{-1/2})^{2(p_i+8\gamma)q_i\sqrt{n}} \leq \frac{q_{i+1}}{q_i} - 8\gamma q_i$$

(this was the computation we ran in Claim 2.10). At most $2n^{1/4}$ vertices v are unsafe with respect to A or B, so to bound the probability wx doesn't get safely $\mathcal{Y}$-closed, we need to replace $2(p_i + 8\gamma)q_i\sqrt{n}$ with $2(p_i + 8\gamma)q_i\sqrt{n} - 2n^{1/4}$. The effect of this is tiny, so we still get

$$\mathbb{E}[Y] \leq q_i k_0^2 \cdot \left(\frac{q_{i+1}}{q_i} - 7\gamma q_i\right) = q_{i+1}k_0^2 - 7\gamma q_i^2 k_0^2.$$

Now it suffices to show concentration within a window of length $n^{1-16\varepsilon}$ (this is smaller than $7\gamma q_i^2 k_0^2$). And now that we've removed high-degree vertices v from consideration, the failed concentration argument from earlier now works — we have $\max_e c_e \leq n^{1/3}$ and

$$\sum_e c_e^2 \leq \max_e c_e \cdot \sum_e c_e \leq n^{1/3} \cdot k_0^2 \cdot 2(p_i + 8\gamma)q_i\sqrt{n} \leq n^{11/6}$$

(by the same argument as we gave earlier). So when we apply Lemma 2.12, we'll have

$$\frac{\sigma}{\max_e c_e} \geq \frac{n^{1-16\varepsilon}}{n^{1/3}} \geq n^{2/3-16\varepsilon} \quad \text{and} \quad \frac{\sigma^2}{p\sum_e c_e^2} \geq \frac{n^{2-32\varepsilon}}{n^{-1/2} \cdot n^{11/6}} = n^{2/3-32\varepsilon}.$$

So we get a concentration probability of $1 - \exp(-\Omega(n^{2/3-32\varepsilon}))$, which is good enough to union-bound over all $\exp(2k_0 \log n) \leq \exp(n^{1/2+\varepsilon})$ pairs of sets of size k_0.

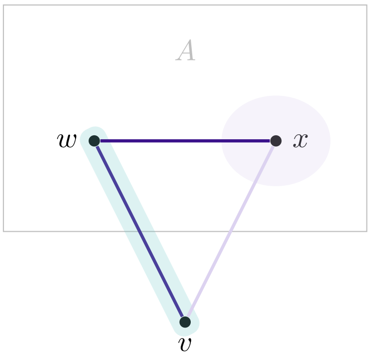
§2.7.2 Property 2.8: Lower bound on open edge densities

Fix some set A of size k; as in the proof of Property 2.7, our goal is to show that it satisfies the desired bound with probability $1 - \exp(n^{\alpha})$ for some $\alpha > \frac{1}{2}$, so that we can union-bound over all such sets.

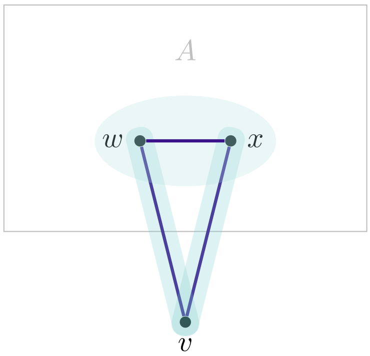
There are two main difficulties in this proof that we haven't encountered in the previous ones. The first difficulty is that up to now, we've only cared about *upper* bounds on open edges, so it sufficed to only consider edges being $\mathcal{Y}$-closed (and not worry about edges being $\mathcal{X}$-closed). Now we want a *lower* bound, so we do need to worry about edges being $\mathcal{X}$-closed. We'd expect this to have substantially smaller contribution (in particular, we'd still expect edges being $\mathcal{Y}$-closed to control the 'main term,' and we'd like to absorb the effect of edges being $\mathcal{X}$-closed into the $64p_i\gamma$ error term) for the reasons discussed in Subsection 2.3; but now we have to figure out how to prove concentration for it. (Some open edges will also leave due to being chosen, but that's a much lower-order term and concentrates well — it's a sum of independent Bernoullis — so can be ignored.)

Furthermore, even with edges being $\mathcal{Y}$-closed, we saw when proving Property 2.7 that concentration breaks down unless we exclude high-degree vertices. When proving Property 2.7, we only wanted an upper bound, so we could just ignore those high-degree vertices. But here we'll have to deal with them.

So the argument will have three main steps. As in the proof of Property 2.7, we'll set $h = n^{1/3}$ and say an edge is *safely $\mathcal{Y}$-closed* if it's $\mathcal{Y}$-closed by a vertex v with $\deg_{E_i}(v, A) \leq h$. We'll also say an edge is *safely $\mathcal{X}$-closed* if it's $\mathcal{X}$-closed by a vertex v such that $\deg_{F_{i+1}}(v, A) \leq h$.



wx safely $\mathcal{Y}$-closed if $\deg_{E_i}(v, A) \leq h$



wx safely $\mathcal{X}$-closed if $\deg_{F_{i+1}}(v, A) \leq h$

(1) First, we'll consider the effect of edges being safely $\mathcal{Y}$-closed. We can deal with this in the same way as we did when proving Property 2.7, and this will give our 'main term.'

(2) Next, we'll consider the number of edges which are safely $\mathcal{X}$-closed (we want to prove this is small). We won't be able to prove concentration for this number itself, but we'll define a proxy which upper-bounds it and whose expectation is small for similar reasons. And this proxy will be a sum of independent random variables, so we *will* be able to prove concentration for it.

(3) Finally, we need to deal with high-degree vertices. We'll actually show that deterministically, there can't be too many edges that are $\mathcal{Y}$-closed or $\mathcal{X}$-closed by high-degree vertices, as long as the chosen edges satisfy certain simple conditions (which will hold with high probability).

Step 1 (The effect of edges being safely $\mathcal{Y}$-closed). Let Y be the number of edges which are not safely $\mathcal{Y}$-closed. Using the lower bound from Claim 2.10 (a lower bound on the probability of not being $\mathcal{Y}$-closed is also a lower bound on the probability of not being safely $\mathcal{Y}$-closed), we get

$$\begin{aligned}\mathbb{E}[Y] &\geq \left(q_i(1 - 64p_i\gamma)\binom{k}{2} - 16p_iq_ik\sqrt{n}\right)\left(\frac{q_{i+1}}{q_i} - 24\gamma q_i\right) \\ &\geq q_{i+1}(1 - 64p_i\gamma)\binom{k}{2} - 24\gamma q_i^2\binom{k}{2} - 16p_iq_{i+1}k\sqrt{n}.\end{aligned}$$

The same argument as in the proof of Property 2.7 shows that Y concentrates within a window of length $n^{1-4\varepsilon}$ (which is smaller than $\gamma q_i^2\binom{k}{2}$) with probability $\exp(-\Omega(n^{2/3-20\varepsilon}))$, which is more than good enough to union-bound.

We're aiming for a final upper bound of

$$q_{i+1}(1-64p_{i+1}\gamma)\binom{k}{2} - 16p_{i+1}q_{i+1}k\sqrt{n}.$$

Since $p_{i+1} = p_i + \gamma q_i$ and $q_{i+1} \geq \frac{1}{2}q_i$, we get to accumulate a total error of

$$(32-25)\gamma^2 q_i^2\binom{k}{2} + 8\gamma q_i^2 k\sqrt{n} = 7\gamma^2 q_i^2\binom{k}{2} + 8\gamma q_i^2 k\sqrt{n} \tag{2.3}$$

in Steps (2) and (3) (we've replaced 24 with 25 to account for the concentration window of Y).

Step 2 (The effect of edges being safely x-closed). For this step, the key insight is that when we argued in Subsection 2.3 that the effect of edges being $\mathcal{X}$-closed should be small, when we said that every edge uv has probability at most

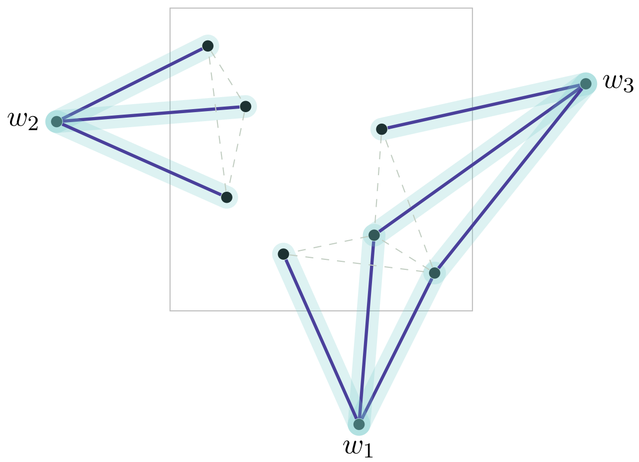
$$X_{uv}(i)\cdot\gamma^2 n^{-1} \leq q_i^2\gamma^2$$

of being $\mathcal{X}$-closed, we just used a union bound over all configurations in $\mathcal{X}_{uv}(i)$; this means the expected *number* of vertices w which $\mathcal{X}$-close uv is also at most $q_i^2\gamma^2$. And if we switch this sum — instead of looking at all open edges uv inside A and then all vertices w which could $\mathcal{X}$-close them, we look at all vertices w and see how many edges inside A they $\mathcal{X}$-close — then we'll essentially get a sum of independent random variables (with one for each w), since for each w, this number only depends on edges incident to w.

More formally, for each vertex w, we define a random variable

$$X_w = \binom{\min\{\deg_{F_{i+1}}(w,A), h\}}{2}.$$

Then $\sum_w X_w$ is an upper bound on the number of edges in A that get safely $\mathcal{X}$-closed — we're essentially going through all vertices w and counting pairs of their neighbors in A among the edges picked this round, and if uv is safely $\mathcal{X}$-closed by w then it'll be counted by X_w. (This overcounts, and it'll also count pairs of neighbors of w where the edge between them had already been closed; but we can afford this.)



If we imagine computing $\mathbb{E}[\sum_w X_w]$ by first summing over pairs uv in A, then each uv has at most $X_{uv}(i) \leq q_i^2 n$ vertices w for which it could *possibly* be counted by X_w (by Property 2.2), and the probability that it *is* counted by X_w is at most $\gamma^2 n^{-1}$; this shows $\mathbb{E}[\sum_w X_w] \leq \gamma^2 q_i^2\binom{k}{2}$. This fits our target error in (2.3), so it suffices to prove that $\sum_w X_w$ concentrates well (e.g., within a window of length $n^{1-4\varepsilon}$).

First, there's one slight issue — the random variables X_w aren't actually independent if $w \in A$. To deal with this, we'll split this sum into two parts — we'll define $X_{\text{in}} = \sum_{w\in A} X_w$ and $X_{\text{out}} = \sum_{w\notin A} X_w$ — and we'll prove concentration for each part separately.

Step 2a (Concentration for X_{in}). To prove concentration for X_{in}, we'll use Lemma 2.12 again. Note that X_{in} only depends on edges inside A, of which there are at most $\binom{k}{2} \leq n^{1+\varepsilon}$. Furthermore, for each edge $e = vw$ inside A, changing whether we pick vw changes X_v and X_w by at most h each (since we've changed $\min\{\deg_{F_{i+1}}(v, A), h\}$ by at most 1), so we have $c_e \leq h$ for every e. This means when we apply Lemma 2.12 with $\sigma = n^{1-4\varepsilon}$, we'll have

$$\frac{\sigma}{\max_e c_e} \geq \frac{n^{1-4\varepsilon}}{n^{1/3}} \geq n^{2/3-4\varepsilon} \quad \text{and} \quad \frac{\sigma^2}{p\sum_e c_e} \geq \frac{n^{2-8\varepsilon}}{n^{-1/2}\cdot n^{1+\varepsilon}\cdot n^{2/3}} \geq n^{5/6-9\varepsilon}.$$

So we get a concentration probability of $1 - \exp(-\Omega(n^{2/3-4\varepsilon}))$, which is good enough to union-bound.

Step 2b (Concentration for X_{out}). The idea is that X_{out} is a sum of *independent* random variables with fairly simple distributions, so we can prove concentration by directly estimating its moment generating function. To do so, for each $w \notin A$, let $f_w(s) = \mathbb{E}[e^{sX_w}]$.

Claim 2.16 — For all vertices w and all $0 \leq s \leq h^{-1}$, we have $0 \leq f''_w(s) \leq 1$.

Proof. Suppose that $\deg_{O_i}(w) = d$. Then we have $\deg_{F_{i+1}}(w) \sim \mathsf{Ber}(p, d)$ where $p = \gamma n^{-1/2}$, which means that $X_w \sim \binom{\min\{i,h\}}{2}$ for $i \sim \mathsf{Ber}(p, d)$. So we can explicitly write out

$$f''_w(s) = \mathbb{E}[X_w^2 e^{sX_w}] = \sum_{i=0}^{d} \binom{d}{i} p^i (1-p)^{d-i} \cdot \binom{\min\{i,h\}}{2} \cdot e^{s\binom{\min\{i,h\}}{2}}.$$

Now since we assumed $s \leq h^{-1}$, we have $s\binom{\min\{i,h\}}{2} \leq i$ for all i. We can also drop the $(1-p)^{d-i}$ terms and bound $\binom{d}{i}\binom{\min\{i,h\}}{2} \leq d^i$ for $i \geq 2$ (and this term is 0 for $i \leq 1$). Then we get

$$f''_w(s) \leq \sum_{i=2}^{d} (epd)^i \leq \sum_{i=2}^{\infty} (epd)^i.$$

But we have $d \leq k = C\sqrt{n \log n}$ and $p = \gamma n^{-1/2}$ (where we chose γ to be $n^{-\varepsilon}$), so $epd \ll 1$. □

Then for each w, by Taylor expansion at 0 we can write

$$\mathbb{E}[e^{sX_w}] \leq 1 + \mathbb{E}[X_w] \cdot s + \frac{1}{2} \cdot 1 \cdot s^2 \leq \exp\left(s\mathbb{E}[X_w] + \frac{s^2}{2}\right)$$

for all $0 \leq s \leq h^{-1}$. Multiplying over all $w \notin A$ (of which there are at most n), we get that

$$\mathbb{E}[e^{sX_{\text{out}}}] = \prod_{w \notin A} \mathbb{E}[e^{sX_w}] \leq \exp\left(s\sum_{w\notin A} \mathbb{E}[X_w] + \frac{ns^2}{2}\right) = e^{s\mathbb{E}[X_{\text{out}}] + ns^2/2}.$$

Then by Markov's inequality, for all $\sigma > 0$ we get that

$$\mathbb{P}[X_{\text{out}} - \mathbb{E}[X_{\text{out}}] \geq \sigma] \leq \frac{\mathbb{E}[e^{sX_{\text{out}}}]}{e^{s\mathbb{E}[X_{\text{out}}]+s\sigma}} \leq e^{ns^2/2 - s\sigma}.$$

We wanted concentration within a window of length $\sigma = n^{1-4\varepsilon}$, so we can set $s = n^{-1/3}$ (which does satisfy $s \leq h^{-1}$) to get a concentration probability of $1 - \exp(-\Omega(n^{2/3-4\varepsilon}))$, which is good enough.

Step 3 (High-degree vertices). Finally, we'll deal with edges closed by high-degree vertices — we'll show that as long as the edges picked this round satisfy certain properties (which hold with high probability), there cannot be too many such edges. Specifically, we'll assume the following conditions:

(i) For all vertices v, we have $\deg_{F_{i+1}}(v) \le \gamma q_i\sqrt{n} + n^{1/3} \le 2\gamma q_i\sqrt{n}$.

(ii) For all $u \neq v$, we have $|N_{E_i}(u) \cap N_{F_{i+1}}(v)| \le (\log n)^2$.

(iii) For all $u \neq v$, we have $|N_{F_{i+1}}(u) \cap N_{F_{i+1}}(v)| \le (\log n)^2$.

These conditions all can be shown to hold with high probability using multiplicative Chernoff bounds, since the quantities of interest are all sums of independent Bernoullis; in fact, we proved them when proving Properties 2.4 and 2.6.

Let S be the set of vertices u such that either $\deg_{E_i}(u, A) \ge h$ or $\deg_{F_{i+1}}(u, A) \ge h$. All the closed edges we haven't accounted for yet go between either $N_{E_i}(u) \cap A$ and $N_{F_{i+1}}(u) \cap A$ (if they were $\mathcal{Y}$-closed) or $N_{F_{i+1}}(u) \cap A$ and $N_{F_{i+1}}(u) \cap A$ (if they were $\mathcal{X}$-closed) for some $u \in S$. Now define

$$A_u = (N_{E_i}(u) \cup N_{F_{i+1}}(u)) \cap A$$

for each $u \in S$. Then the definition of S means that $|A_u| \ge h$ for all u. Meanwhile, for all $u \neq v$, we have

$$|A_u \cap A_v| \le i(\log n)^2 + 3(\log n)^2 \le n^{2\varepsilon}$$

by Property 2.4 (which bounds $N_{E_i}(u, A) \cap N_{E_i}(v, A)$, corresponding to the first term) and the conditions (ii) and (iii). So we can apply Fact 2.14 (since $h \ge 4\sqrt{k \cdot n^{2\varepsilon}}$) to say that $\sum_{u \in S} |A_u| \le 2k$.

Now we claim that for every $u \in S$, the number of open edges between $N_{F_{i+1}}(u) \cap A$ and A_u was at most $2\gamma q_i^2\sqrt{n} \cdot |A_u|$ (at the start of the round).

Case 1 (We have $\deg_{F_{i+1}}(u, A) \le 2\gamma^2 q_i^2\sqrt{n}$). Then this claim is immediate, as the number of edges between $N_{F_{i+1}}(u) \cap A$ and A_u is certainly at most $|N_{F_{i+1}}(u) \cap A| \cdot |A_u| = \deg_{F_{i+1}}(u, A) \cdot |A_u|$.

Case 2 (We have $\deg_{F_{i+1}}(u, A) \ge 2\gamma^2 q_i^2\sqrt{n}$). In this case, we also have $|A_u| \ge 2\gamma^2 q_i^2\sqrt{n}$ (since A_u contains $N_{F_{i+1}}(u) \cap A$), so we can apply Property 2.7 to say that the open edge density between them (at the start of the round) is at most q_i. (Property 2.7 was written for *disjoint* sets of size *exactly* $\gamma^2 q_i^2\sqrt{n}$, but by an averaging argument, the same density bound holds for any two sets of size at least $2\gamma^2 q_i^2\sqrt{n}$.) Furthermore, (i) gives that $\deg_{F_{i+1}}(u) \le 2\gamma q_i\sqrt{n}$. So the number of edges between $N_{F_{i+1}}(u) \cap A$ and A_u is at most

$$q_i \cdot \deg_{F_{i+1}}(u, A) \cdot |A_u| \le q_i \cdot 2\gamma q_i\sqrt{n} \cdot |A_u|.$$

Finally, summing over all $u \in S$, we get that the total number of edges which got closed by some $u \in S$ this round is at most

$$2\gamma q_i^2\sqrt{n} \cdot \sum_u |A_u| \le 4\gamma q_i^2 k\sqrt{n}.$$

This fits the second term in our target error in (2.3), so we're done.

§2.8 Property 2.9: Independent sets

We'll show that for every set A of size k, the probability we don't add any of its edges to G_{i+1} is at most $\exp(-\frac{1}{32}\gamma q_i n^{-1/2}\binom{k}{2})$. This means the *expected* number of independent sets remaining will drop by a factor of $\exp(-\frac{1}{32}\gamma q_i n^{-1/2}\binom{k}{2})$, so the *actual* number will drop by a factor of $\exp(-\frac{1}{64}\gamma q_i n^{-1/2}\binom{k}{2})$ with high probability (which is what we need in order to preserve Property 2.9, since $p_{i+1} = p_i + \gamma q_i$).

First, by Property 2.8, the number of open edges in A at the start of the round is at least $\frac{1}{2}q_i\binom{k}{2}$. We're picking each with probability $\gamma n^{-1/2}$, so by multiplicative Chernoff bounds, the probability that we pick less than $\frac{1}{4}\gamma q_i n^{-1/2}\binom{k}{2}$ of these edges (to place in F_{i+1}) is at most

$$\exp\left(-\frac{1}{12}\gamma q_i n^{-1/2}\binom{k}{2}\right).$$

Meanwhile, we want to show that the probability we *discard* at least this many edges in A during the alteration step (meaning that we chose them but didn't place them in G_{i+1}) is similarly small. For this, as discussed in Subsection 2.2, the intuition is that there's at most $q_i\binom{k}{2}$ open edges inside A by Property 2.7. For such an edge uv to get chosen and discarded, either we must choose both uv and the one open edge of some configuration in $\mathcal{Y}_{uv}(i)$, or we must choose uv and both edges of some configuration in $\mathcal{X}_{uv}(i)$. The probability this occurs is at most

$$2(p_i + 8\gamma)q_i\sqrt{n} \cdot \gamma n^{-1/2} + q_i^2 n \cdot \gamma^2 n^{-1} = 2\gamma(p_i + 8\gamma)q_i + \gamma^2 q_i^2 \ll 1$$

(because γ is small); so the expected number of edges we discard is much less than $\frac{1}{8}\gamma q_i k^2 n^{-1/2}$. To turn this into a statement saying that it's exponentially unlikely that we discard this many edges, we'll use the following fact.

Fact 2.17 — Let $\mathcal{A}_1, \ldots, \mathcal{A}_r$ be events such that $\sum_{i=1}^r \mathbb{P}[\mathcal{A}_i] \leq \eta$. Then the probability that there is some size-ℓ collection of mutually independent $\mathcal{A}_i$'s for which all occur is at most $\frac{\eta^\ell}{\ell!}$.

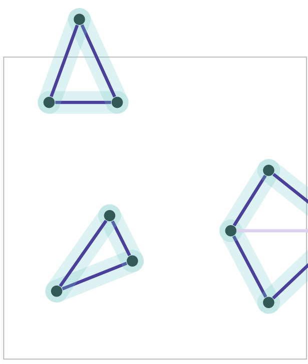
Proof. Imagine we expand out $(\sum_{i=1}^r \mathbb{P}[A_i])^\ell = \eta^\ell$. For every such collection $\{\mathcal{A}_{i_1}, \ldots, \mathcal{A}_{i_\ell}\}$, we'll have $\ell!$ terms $\mathbb{P}[\mathcal{A}_{i_1}] \cdots \mathbb{P}[\mathcal{A}_{i_\ell}]$ in this sum (one for each order of the indices). And because $\mathcal{A}_{i_1}, \ldots, \mathcal{A}_{i_\ell}$ are mutually independent, we have $\mathbb{P}[\mathcal{A}_{i_1}] \cdots \mathbb{P}[\mathcal{A}_{i_\ell}] = \mathbb{P}[\mathcal{A}_{i_1} \wedge \cdots \wedge \mathcal{A}_{i_\ell}]$.

So the expansion of $(\sum_{i=1}^r \mathbb{P}[\mathcal{A}_i])^\ell$ has $\ell!$ copies of the probability corresponding to each such size-ℓ collection (along with possibly some other terms), giving the desired result. □

In our setting, for each open edge uv in A, we define an event $\mathcal{A}_i$ for each configuration in $\mathcal{Y}_{uv}(i)$ (saying that we choose uv and the one open edge in that configuration) and an event $\mathcal{A}_i$ for each configuration in $\mathcal{X}_{uv}(i)$ (saying that we choose uv and both open edges in that configuration). As seen above, we have

$$\sum_i \mathbb{P}[\mathcal{A}_i] \leq q_i\binom{k}{2} \cdot \gamma n^{-1/2} \cdot \left(2\gamma(p_i + 8\gamma)q_i + \gamma^2 q_i^2\right) \leq \frac{1}{960}\gamma q_i n^{-1/2}\binom{k}{2}.$$

And for us to discard m chosen edges from A, at least $\frac{1}{3}m$ events of this form whose associated 2 or 3 open edges are *disjoint* (which means these events are independent) must occur — this is because when we perform alterations, we do so by removing a maximal edge-disjoint subset of $\wedge_{i+1} \cup \triangle_{i+1}$.



So using Fact 2.17 (and bounding $\frac{\eta^\ell}{\ell!} \leq (\frac{4\eta}{\ell})^\ell$) gives that the probability we discard at least $\frac{1}{4}\gamma q_i n^{-1/2}\binom{k}{2}$ edges from A is at most

$$\left(\frac{4 \cdot \frac{1}{960}\gamma q_i n^{-1/2}\binom{k}{2}}{\frac{1}{12}\gamma q_i n^{-1/2}\binom{k}{2}}\right)^{\frac{1}{12}\gamma q_i n^{-1/2}\binom{k}{2}} \leq \exp\left(-\frac{1}{12}\gamma q_i n^{-1/2}\binom{k}{2}\right).$$

Combining these two statements, we get that the probability A remains independent is at most

$$\exp\left(-\frac{1}{12}\gamma q_i n^{-1/2}\binom{k}{2}\right) + \exp\left(-\frac{1}{12}\gamma q_i n^{-1/2}\binom{k}{2}\right) \leq \exp\left(-\frac{1}{32}\gamma q_i n^{-1/2}\binom{k}{2}\right).$$

By Markov's inequality (on the total number of independent sets), this means that the total number of independent sets multiplies by a factor of $\exp(-\frac{1}{64}\gamma q_i n^{-1/2}\binom{k}{2})$ with high probability, as desired.

§3 Wolfovitz 2011: An analysis via branching processes

In this section, we'll explain Wolfovitz's approach to analyzing the triangle-free process from [Wol11], based on the semi-random method and branching processes. In particular, we'll give his proof that the process follows the trajectory described in Subsection 1.3 up to time n^c (for small c).

§3.1 Overview

Let $\varepsilon > 0$ be a very small constant, let $c > 0$ be reasonably small with respect to ε, and let $\gamma = n^{-\varepsilon}$.

Similarly to Kim's approach, we'd like to analyze the triangle-free process in 'rounds' which roughly correspond to time-intervals $[\gamma i, \gamma(i+1)]$. We can reparametrize the process in this framework as follows: We begin with G_0 being the empty graph. On the ith round (where we begin with a graph G_i), we sample a subset F_{i+1} of the currently open edges, including each with probability $\gamma n^{-1/2}$. For each of these edges, we also generate a *birthtime*, which is uniform in $[0, \gamma]$. We then go through these edges in order of birthtime; for each, we add it to G_i if it's still open at the time we attempt to do so, and discard it if it has become closed. We'll run this process up to $i = n^c/\gamma$.

Our goal is to show that $\mathcal{X}_{uv}(i)$, $\mathcal{Y}_{uv}(i)$, and $\mathcal{Z}_{uv}(i)$ follow the trajectory described in Subsection 1.3, up to some slowly deteriorating error terms. To describe these error terms, we write $p_i = \Psi(\gamma i)$ and $q_i = \psi(\gamma i)$, and we define $\delta_i = \gamma p_i q_i + \gamma q_i$ and

$$\Delta_i = n^{-30\varepsilon}\prod_{j=0}^{i-1}(1 + 60\delta_j)$$

(so $\Delta_{i+1} = \Delta_i + 60\Delta_i\delta_i$). Note that these errors remain 'reasonable' up to time n^c — we have

$$\prod_{j=0}^{i-1}(1+60\delta_j) \leq \exp\left(60\sum_{j=0}^{i-1}(p_i q_i + q_i)\gamma\right) \approx \exp\left(60\int_0^{\gamma i}(\Psi(s)\psi(s) + \psi(s))\,ds\right) = \exp(30\Psi(\gamma i)^2 + 60\Psi(\gamma i)),$$

and since $\Psi(t) \approx \sqrt{\log t}$ and we're running up to time $\gamma i = n^c$ (where c is small relative to ε), this will be small compared to $n^{30\varepsilon}$, so Δ_i will remain small (e.g., at most $n^{-20\varepsilon}$).

We'll consider the following properties.

Property 3.1. For all edges $uv \notin G_i$, we have $X_{uv}(i) = q_i^2(1 \pm \Delta_i)n$.

Property 3.2. For all edges $uv \notin G_i$, we have $Y_{uv}(i) = 2q_i(p_i \pm (1+p_i)\Delta_i)\sqrt{n}$.

Property 3.3. For all edges uv, we have $Z_{uv}(i) \leq i(\log n)^2$.

Theorem 3.4

With probability $1 - n^{-\omega(1)}$, Properties 3.1–3.3 hold for all $i \le n^c/\gamma$.

More specifically, we'll show that assuming these properties hold up to the ith round, then they hold after the ith round (with i replaced by $i+1$) with probability $1 - n^{-\omega(1)}$.

First, Property 3.3 follows easily from multiplicative Chernoff, using the same argument as Property 2.4 in Kim's argument (since we only want an upper bound). But there are several challenges that arise when we try to adapt the proofs of Properties 2.2 and 2.3 to the actual triangle-free process.

(1) Most importantly, there can be complicated dependencies between the edges we try to add this round — maybe the first edge e_1 we add closes the second edge e_2 (so we don't actually add e_2); and adding e_2 would have closed e_3, but this didn't happen because we didn't add e_2; and so on. This makes it unclear how to use concentration inequalities, since a single edge could potentially affect lots of others.

(2) There's also the fact that unlike in Kim's semi-random construction, we don't get to perform a regularization step, and this means errors accumulate — errors in $X_{uv}(i)$ and $Y_{uv}(i)$ are going to result in bigger errors in $X_{uv}(i+1)$ and $Y_{uv}(i+1)$. So we need to control these errors carefully enough that they don't blow up — in particular, we need more careful estimates on the *expectations* of the tracked variables (meaning we need careful estimates on the probabilities that edges in F_{i+1} really get added to G_{i+1}, as well as that edges not in F_{i+1} remain open).

The way Wolfovitz deals with both of these challenges is through an approach based on *branching processes.* As a high-level overview, we first sample a collection of 'candidate' edges, including each with probability $\lambda n^{-1/2}$ where $\lambda = n^{2^{20}\varepsilon}$ (then we'll decide which edges to add this round by sampling these candidates with probability $\frac{\gamma}{\lambda}$). Then for every edge, in order to estimate the probability it gets added to G_{i+1} or remains open, we produce a 'dependency tree' that keeps track of how the candidate edges could affect it.

Each individual dependency tree basically corresponds to a branching process. By analyzing this branching process, we can show that it exhibits good 'correlation decay,' so that we can cut it off after a constant number of levels without affecting its outcome by too much. We can also show that it's reasonably resilient to errors, in that a $1 \pm \Delta_i$ error in its input only results in a $1 \pm \Delta_i \delta_i$ error in its outcome. (This is important for dealing with (2) — it's the reason our errors don't blow up.)

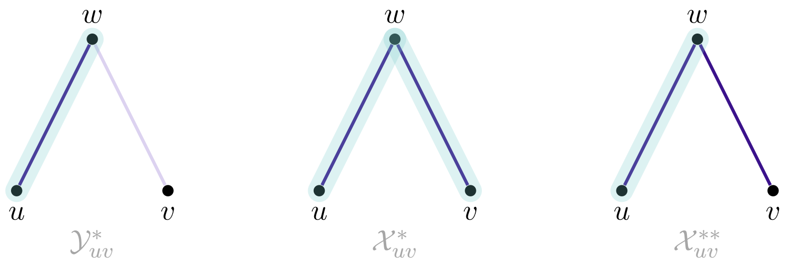
Then when we're considering $X_{uv}(i+1)$ or $Y_{uv}(i+1)$, we can use these truncated dependency trees to get upper and lower bounds which are very close to each other in expectation. And the truncations substantially reduce the amount of dependencies, so that we can actually use the bounded differences inequality to prove concentration; this deals with (1).

§3.2 Sampling candidates

Let $\lambda = n^{2^{20}\varepsilon}$. The first step of the argument is to sample a set of candidates F^*, where we include each open edge with probability $\lambda n^{-1/2}$. (We'll draw candidate edges by highlighting them in blue.)

We'll define the following sets of configurations based on these candidate edges:

- We define $\mathcal{Y}^*_{uv}$ as the set of configurations $\{uw, vw\} \in \mathcal{Y}_{uv}(i)$ whose one open edge was selected as a candidate, and Y^*_{uv} as its size.
- We define $\mathcal{X}^*_{uv}$ as the set of configurations $\{uw, vw\} \in \mathcal{X}_{uv}(i)$ for which *both* their open edges were selected as candidates, and X^*_{uv} as its size.
- We define $\mathcal{X}^{**}_{uv}$ as the set of configurations $\{uw, vw\} \in \mathcal{X}_{uv}(i)$ for which exactly *one* of their open edges was selected as a candidate, and X^*_{uv} as its size.



Claim 3.5 — With probability $1 - n^{-\omega(1)}$, the following statements hold for all $uv \notin G_i$:

(i) $Y^*_{uv} = 2\lambda q_i(p_i \pm (1 + p_i)(\Delta_i + \Delta_i\delta_i))$.

(ii) $X^*_{uv} = \lambda^2 q_i^2(1 \pm (\Delta_i + \Delta_i\delta_i))$.

(iii) $X^{**}_{uv} = 2\lambda q_i^2(1 \pm (\Delta_i + \Delta_i\delta_i))\sqrt{n}$.

Proof. These statements all follow from multiplicative Chernoff.

For (i), there are $2q_i(p_i \pm (1 + p_i)\Delta_i)\sqrt{n}$ configurations in $\mathcal{Y}_{uv}(i)$ by Property 3.2, and for each, we pick its one open edge with probability $\lambda n^{-1/2}$, so the *expected* number of them that will land in $\mathcal{Y}^*_{uv}$ is

$$\mathbb{E}[Y^*_{uv}] = \lambda n^{-1/2} \cdot 2q_i(p_i \pm (1 + p_i)\Delta_i)\sqrt{n} = 2\lambda q_i(p_i \pm (1 + p_i)\Delta_i).$$

And these configurations are all edge-disjoint and therefore independent. So $|Y^\star_i(uv)|$ is a sum of independent Bernoullis, which means it concentrates well. (Note that when we apply multiplicative Chernoff, we'll have mean $\mu \approx 2\lambda p_i q_i$ and error $\beta \approx \frac{1}{2}\Delta_i\delta_i$; the fact that $\lambda = n^{2^{20}\varepsilon}$ while all other quantities are 'reasonable' powers of n^ε means that $\beta^2\mu$ will be large — for example, it's much greater than n^ε — so we'll get a $1 - n^{-\omega(1)}$ concentration probability.)

Similarly, for (ii), there are $q_i^2(1 \pm \Delta_i)n$ configurations in $\mathcal{X}_{uv}(i)$ by Property 3.1, and each lands in $\mathcal{X}^*_{uv}$ with probability $\lambda^2 n^{-1}$ (since we have to pick *both* its edges). So

$$\mathbb{E}[X^*_{uv}] = \lambda^2 n^{-1} \cdot q_i^2(1 \pm \Delta_i)n = \lambda^2 q_i^2(1 \pm \Delta_i).$$

And again, these configurations are edge-disjoint and therefore independent, so multiplicative Chernoff gives good concentration. Finally, for (iii), each configuration in $\mathcal{X}_{uv}(i)$ lands in $\mathcal{X}^{**}_{uv}$ with probability $2\lambda n^{-1/2}$, since we have to pick *one* of its edges (technically, we need to subtract $2\lambda^2 n^{-1}$ to remove the ones where we pick both edges, but this is a much lower-order term and can be absorbed into the error). So

$$\mathbb{E}[X^{**}_{uv}] = 2\lambda q_i^2(1 \pm \Delta_i)\sqrt{n},$$

and again we get good concentration by multiplicative Chernoff. □

We'll also need bounds on certain small structures (we'll use these when proving concentration).

Claim 3.6 — With probability $1 - n^{-\omega(1)}$, for all vertices u, v, and w with $uw, vw \notin G_i$:

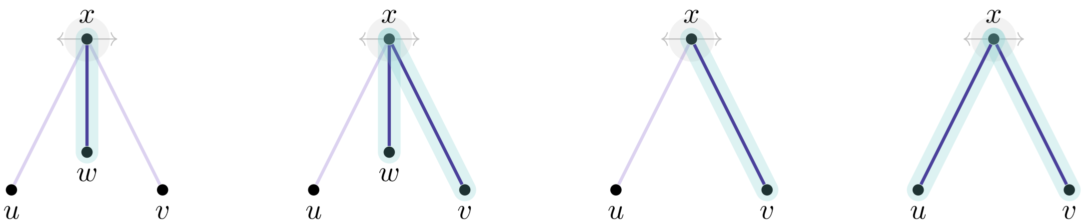
(i) The number of x such that $ux, vx \in G_i$ and $wx \in F^*$ is at most $(\log n)^2$.

(ii) The number of x such that $ux \in G_i$ and $vx, wx \in F^\star_i$ is at most $(\log n)^2$.

Also, for all vertices u and v with $uv \notin G_i$:

(iii) The number of x such that $ux \in G_i$ and $vx \in F^*$ is at most 4λ.

(iv) The number of x such that $ux, vx \in F^*$ is at most $2\lambda^2$.



Proof. For (i), we must have $\{ux, vx\} \in \mathcal{Z}_{uv}(i)$. Property 3.3 says that there are at most $i(\log n)^2$ such vertices x, and for each we include wx in F^* with probability $\lambda n^{-1/2}$. So the *expected* number of valid x is at most

$$i(\log n)^2 \cdot \lambda n^{-1/2} \ll 1.$$

And different vertices x are independent, so multiplicative Chernoff says the actual number of such x is at most $(\log n)^2$ with probability $1 - n^{-\omega(1)}$.

Similarly, for (ii), we must have $\{ux, vx\} \in \mathcal{Y}_{uv}(i)$. Property 3.2 says there are at most $2\sqrt{n}$ such choices of x; and for each, the probability that we include both wx and vx in F^* is $\lambda^2 n^{-1}$. So the *expected* number of valid x is at most

$$2\sqrt{n} \cdot \lambda^2 n^{-1} \ll 1.$$

And different vertices x are independent, so we can again use multiplicative Chernoff to conclude.

For (iii), we must have $\{ux, vx\} \in \mathcal{Y}_{uv}(i)$, and again by Property 3.2 there are at most $2\sqrt{n}$ such choices of x; and for each, the probability we include vx in F^* is $\lambda n^{-1/2}$. So the *expected* number of valid x is at most

$$2\sqrt{n} \cdot \lambda n^{-1/2} = 2\lambda,$$

and since different vertices x are independent (and $\lambda = \omega(\log n)$), multiplicative Chernoff tells us that the actual number is at most 4λ with probability $1 - n^{-\omega(1)}$.

Similarly, for (iv), there are at most n choices for x, and for each, the probability we include both ux and vx in F^* is $\lambda^2 n^{-1}$. So the *expected* number of valid x is at most

$$n \cdot \lambda^2 n^{-1} = \lambda^2,$$

and by multiplicative Chernoff, the actual number is at most $2\lambda^2$ with probability $1 - n^{-\omega(1)}$. □

Claim 3.7 — With probability $1 - n^{-\omega(1)}$, for all v we have $\deg_{F^*}(v) \leq 2\lambda\sqrt{n}$.

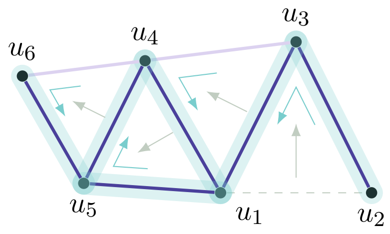
Proof. This again follows from multiplicative Chernoff — for each v, there are at most n open edges incident to v, and we're picking each with probability $\lambda n^{-1/2}$. So the *expected* number of candidate edges incident to v is at most $\lambda\sqrt{n}$, and by multiplicative Chernoff, the *actual* number is at most $2\lambda\sqrt{n}$ with high probability. □

In the rest of the analysis, we'll assume that F^* has already been chosen, and that it satisfies the properties described in Claims 3.5–3.7.

§3.3 Dependency trees

We'll now describe how to construct dependency trees that track how the candidate edges can affect each edge e. We define a *dependency walk* started from e as follows: At each step, if we're currently at an edge uv, we move to a configuration in either $\mathcal{Y}^*_{uv}$ or $\mathcal{X}^*_{uv}$, and then we move to one of the candidate edges of that configuration. (For convenience, we'll sometimes refer to such configurations as $\mathcal{Y}$-type and $\mathcal{X}$-type, respectively; so there's one possible step from a $\mathcal{Y}$-type configuration, and two from an $\mathcal{X}$-type configuration.)

In particular, the starting edge e of the dependency walk may or may not be a candidate edge, but all the other edges on the walk are candidates.



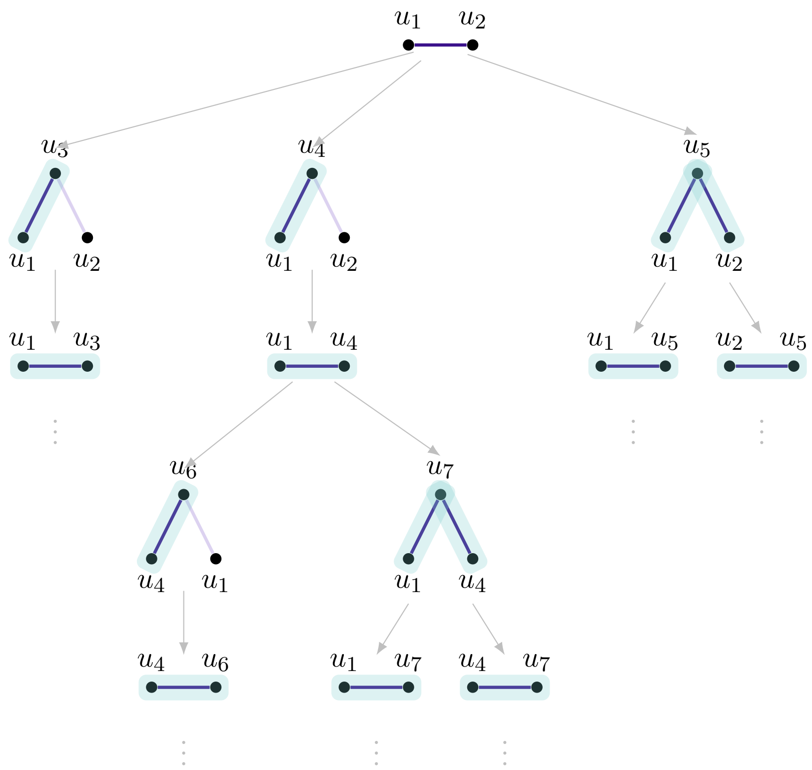
For example, the above picture represents the dependency walk

$$u_1u_2 \to \{u_1u_3, u_2u_3\} \to u_1u_3 \to \{u_1u_4, u_3u_4\} \to u_1u_4 \to \{u_1u_5, u_4u_5\} \to u_4u_5 \to \{u_4u_6, u_5u_6\} \to u_5u_6.$$

We allow a dependency walk to revisit vertices or edges. However, we do *not* allow it to stay on the same triple of vertices for two consecutive configurations — for example, we can't have

$$\{u_1u_5, u_4u_5\} \to u_4u_5 \to \{u_1u_4, u_1u_5\}.$$

For each edge e, we also construct a *dependency tree* $\mathcal{T}(e)$ that records all possible dependency walks started from e, which will look something like the following picture.



(Note that the same edge is allowed to appear in multiple places in the same tree, and the tree may be infinite.) We measure heights in the tree based on only the nodes representing edges: for example, in the above tree we'd say u_1u_2 is at level 0, u_1u_3 and u_2u_3 are at level 1, and so on. We also define truncated versions of $\mathcal{T}(e)$ — we define $\mathcal{T}_h(e)$ to consist of only the nodes up to level h. (In particular, $\mathcal{T}_0(e)$ consists of just e itself; $\mathcal{T}_1(e)$ consists of e at level 0, then the configurations in $\mathcal{Y}_e^*$ and $\mathcal{X}_e^*$ as its children, and then their candidate edges at level 1.)

We can then use these trees to model the triangle-free process as follows: First, for each candidate edge e, we generate a birthtime in $[0, \lambda]$ uniformly at random. We say e is *born* if its birthtime is at most γ; these are the edges we'll be attempting to add during the triangle-free process. We say a configuration is *fully born* if all its candidate edges (one for a $\mathcal{Y}$-type configuration and two for an $\mathcal{X}$-type configuration) are born; we say it is fully born by time x (for $x \in [0, \gamma]$) if all its candidate edges have birthtimes at most x.

Now, given a finite dependency tree $\mathcal{T}$, we compute with it as follows: We say all its leaves (which are nodes representing edges) *survive*. For each intermediate node representing a configuration, we say it *fully survives* if all its candidate edges survive. For each intermediate node representing an edge, if its birthtime is x, we say it *survives* if and only if none of its children configurations has been fully born by time $\min\{x, \gamma\}$ and fully survives.

We say an edge e *survives at depth h* if it survives in $\mathcal{T}_h(e)$. The intuition is that we think of surviving as a proxy for not getting closed; if some child configuration of e is fully born before time x and both its edges are still open at the time we add them to G_i, then this will close e. More precisely, when h is even, this will give an overestimate for whether e remains open (if its birthtime is above γ) or gets added to G_i (if its birthtime is below γ); when h is odd, it'll give an underestimate. (This is because e always survives if $h = 0$, and if we have overestimates at all children of the root, then we have an underestimate at the root.)

We're later going to run a branching process analysis of these dependency trees, which will give the following result — that for any constant-sized collection of edges, working with these truncated dependency trees genuinely give very good estimates for what happens during the triangle-free process.

Lemma 3.8

Let $H \subseteq K_n \setminus G_i$ be a constant-sized collection of edges such that $H \cup G_i$ is triangle-free. Then if we condition on whether or not each edge of H gets born, then for each $h \in \{40, 41\}$, the probability that every $e \in H$ survives at depth h is

$$\left(\frac{p_{i+1} - p_i}{\gamma q_i}\right)^{a_1} \left(\frac{q_{i+1}}{q_i}\right)^{a_2} (1 \pm 12\Delta_i\delta_i)^{a_1+a_2}.$$

(In particular, since $h = 40$ gives an overestimate and $h = 41$ gives an underestimate for the actual process, we get that the same statement is true for the corresponding probability in the actual process.)

The intuition behind why this is the right probability is that for a *single* edge e, if e doesn't get born, then surviving should correspond to remaining open at the end of this round; and we'd expect the density of open edges to drop from q_i to q_{i+1}, which means we'd expect each edge to survive with probability roughly $\frac{q_{i+1}}{q_i}$. Meanwhile, if e does get born at a time $t \in [0, \gamma]$, then surviving should correspond to being open at the time we tried to add it. And this time essentially corresponds to the time $\gamma i + t$ in the full triangle-free process, at which point our heuristic from Subsection 1.2 says we'd expect the open edge density to be $\psi(\gamma t + i)$; so we'd expect e to be open at that time with probability $\frac{\psi(\gamma t+i)}{q_i}$. Averaging over all times, this means if we just condition on e being born, the probability it's open when we try to add it should be

$$\frac{1}{\gamma}\int_0^\gamma \frac{\psi(\gamma t + i)}{q_i}\, dt = \frac{p_{i+1} - p_i}{\gamma q_i}.$$

This explains why Lemma 3.8 makes sense if H consists of a *single* edge. And we'd expect different edges to behave independently, so it makes sense that when H has multiple edges, we just multiply their probabilities.

§3.4 Tracking configurations

We'll prove Lemma 3.8 in Subsection 3.5; for now, we'll see how to use it to prove Properties 3.1 and 3.2. (As usual, we'll fix an edge uv and show that these properties hold with probability $1 - n^{-\omega(1)}$, which is good enough to union-bound.) Our tool for proving concentration will be the (ordinary) bounded differences inequality.

Lemma 3.9 (Bounded differences inequality)

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function with the property that for each j, changing the jth coordinate of x changes $f(x)$ by at most c_j. Then for X sampled according to a distribution on $\mathbb{R}^n$ with independent coordinates, for all $\sigma > 0$ we have

$$\mathbb{P}[|f(X) - \mathbb{E}[f(X)]| \geq \sigma] \leq 2\exp\left(-\frac{1}{2} \cdot \frac{\sigma^2}{\sum_j c_j^2}\right).$$

(One *could* use a low-probability version similar to Lemma 2.12, but here our probabilities will be $\frac{\gamma}{\lambda}$, which is some power of n^ε; and powers of n^ε will not matter in the concentration bounds.)

Note that Claim 3.5 gives control on degrees in our dependency trees — in any dependency walk, if we're currently at uv, we must step to either a configuration in $\mathcal{Y}^*_{uv}$ or $\mathcal{X}^*_{uv}$, so Claim 3.5 means we have at most $2\lambda^2$ choices per step. This is what we'll use to control the c_e's when applying the bounded differences inequality.

§3.4.1 Property 3.1: Tracking $\mathcal{X}_{uv}(i)$

For convenience, let $\mathcal{X}'_{uv}(i) = \mathcal{X}_{uv}(i) \setminus (\mathcal{X}^*_{uv}(i) \cup \mathcal{X}^{**}_{uv}(i))$, and let $X'_{uv}(i)$ be its size. (We can afford to ignore configurations in $\mathcal{X}^*_{uv}(i)$ or $\mathcal{X}^{**}_{uv}(i)$ because they're much smaller — by Claim 3.5 they scale like $\sqrt{n}$, while $\mathcal{X}_{uv}(i)$ scales like n — so whatever happens to them can be absorbed into our error term.)

For each $h \in \{40, 41\}$, let $\mathcal{X}^h_{uv}(i+1)$ be the set of configurations in $\mathcal{X}'_{uv}(i)$ both of whose edges survive at depth h, and let $X^h_{uv}(i+1)$ be its size. Then we have

$$X^{41}_{uv}(i+1) \leq X_{uv}(i+1) \leq X^{40}_{uv}(i+1) + X^*_{uv} + X^{**}_{uv}.$$

So our goal is to show that both $X^{40}_{uv}(i+1)$ and $X^{41}_{uv}(i+1)$ have the correct mean and concentrate well.

Fix $h \in \{40, 41\}$. Then we can estimate $\mathbb{E}[X^h_{uv}(i+1)]$ using Lemma 3.8. We have

$$X'_{uv}(i) = q_i^2(1 \pm (\Delta_i + \Delta_i\delta_i))n$$

by Property 3.1 (the extra $\Delta_i\delta_i$ is there just to account for the removal of configurations in $\mathcal{X}^*_{uv}$ and $\mathcal{X}^{**}_{uv}$), and by Lemma 3.8, for each configuration in $\mathcal{X}'_{uv}(i)$, the probability both its edges survive at depth h is

$$\left(\frac{q_{i+1}}{q_i}\right)^2 \cdot (1 \pm 12\Delta_i\delta_i)^2.$$

(We don't have to worry about the conditioning in Lemma 3.8 because we excluded configurations in $\mathcal{X}^*_{uv}$ and $\mathcal{X}^{**}_{uv}$, so no edges of configurations in $\mathcal{X}'_{uv}(i)$ can possibly be born this round.) So we get that

$$\mathbb{E}[X^h_{uv}(i+1)] = q_i^2(1 \pm (\Delta_i + \Delta_i\delta_i))n \cdot \left(\frac{q_{i+1}}{q_i}\right)^2 \cdot (1 \pm 12\Delta_i\delta_i)^2 = q_{i+1}^2(1 \pm (\Delta_i + 26\Delta_i\delta_i))n.$$

We wanted a bound of

$$q_{i+1}^2(1 \pm \Delta_{i+1})n = q_{i+1}^2(1 \pm (\Delta_i + 60\Delta_i\delta_i))n,$$

so it just remains to prove that $X^h_{uv}(i+1)$ concentrates well (e.g., within a window of length $n^{1-80\varepsilon}$). We'll do this by the bounded differences inequality. So we want to understand, for each candidate edge e, how many configurations $\{uw, vw\} \in \mathcal{X}'_{uv}(i)$ can be affected by e (meaning that changing the birthtime of e could change whether or not uw or vw survives at depth h).

First, if e affects $\{uw, vw\}$, then there must be some dependency walk from uw (or vw, but we'll work with just uw for convenience) to e of length at most h. Let f be the edge immediately after uw on this walk. We'll split into cases based on whether f is incident to u or not.

Case 1 ($f = ux$ for some x). This means we have one of the two following pictures.

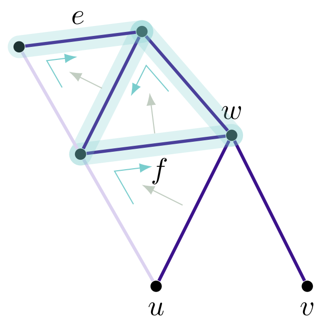
First, to bound the *number* of edges e which fall into this case (over all possible w), note that by Claim 3.7 there are at most $2\lambda\sqrt{n}$ choices for f (since it must be a candidate edge incident to u). Then at each step of the dependency walk, we have at most $2\lambda^2$ choices; we're taking h steps, so in total there are at most

$$2\lambda\sqrt{n} \cdot (2\lambda^2)^h \leq \lambda^{3h}\sqrt{n}$$

edges e which fall into this case.

Then to bound the number of configurations $\{uw, vw\}$ that a fixed edge e in this case can affect, we can first imagine running the dependency walk backwards from e to f; this shows there's at most $(2\lambda^2)^h$ choices for f. Then given $f = ux$, we have that $\{uw, xw\}$ must be in either $\mathcal{Y}_{ux}(i)$ (as in the first picture) or $\mathcal{X}^{**}_{ux}$ (as in the second). So by Property 3.2 and Claim 3.5, there are at most $2\lambda^2\sqrt{n}$ choices for w. So this case gives $\lambda^{3h}\sqrt{n}$ edges with $c_e \leq (2\lambda^2)^h \cdot 2\lambda^2\sqrt{n} \leq \lambda^{3h}\sqrt{n}$.

Case 2 (f is not incident to u). This means we have a picture like the following.



First, to estimate the number of edges e which fall into this case, there are at most n choices for w. Then there's at most $2\lambda^2$ choices at each step of the dependency walk, and we're taking h steps. So in total, there are at most $(2\lambda^2)^h n \leq \lambda^{3h} n$ choices for e.

To bound the number of configurations $\{uw, vw\}$ each such e can affect, we can again imagine walking backwards from e to f, so there's at most $(2\lambda^2)^h$ choices for f. Then given f, there's at most 2 choices for w (it has to be one of the two endpoints of f).

So this case gives $\lambda^{3h} n$ edges with $c_e \leq (2\lambda^2)^h \cdot 2 \leq \lambda^{3h}$.

So when we apply the bounded differences inequality with $\sigma = n^{1-80\varepsilon}$, we'll have

$$\frac{\sigma^2}{\sum_e c_e^2} \geq \frac{n^{2-160\varepsilon}}{\lambda^{3h}\sqrt{n} \cdot \lambda^{6h} n + \lambda^{3h} n \cdot \lambda^{6h}} \geq n^{1/2 - 2^{30}\varepsilon}.$$

So we get a concentration probability of $1 - \exp(-\Omega(n^{1/2-2^{30}\varepsilon}))$, which is good enough (all we needed was for the exponent of n to be positive).

§3.4.2 Property 3.2: Tracking $\mathcal{Y}_{uv}(i)$

Based on (1.4), we'd expect the primary driving forces behind the change in $\mathcal{Y}_{uv}(i)$ to be configurations leaving through their open edge becoming closed, and configurations coming in from $\mathcal{X}_{uv}^{**}$ (which means their one candidate edge has to be born and survive, meaning that it gets added to the graph, and their one non-candidate edge has to survive, meaning that it remains open). So we'll define the following sets:

- Let $\mathcal{Y}'_{uv}(i) = \mathcal{Y}_{uv}(i) \setminus \mathcal{Y}^*_{uv}$, and for $h \in \{40, 41\}$, let $\mathcal{Y}^h_{uv}(i+1)$ be the set of configurations in $\mathcal{Y}'_{uv}(i)$ whose open edge survives at depth h. (We write $Y'_{uv}(i)$ and $Y^h_{uv}(i+1)$ for their sizes.)
- For $h \in \{40, 41\}$, let $\mathcal{Y}^{h+}_{uv}(i+1)$ be the set of configurations in $\mathcal{X}^{**}_{uv}$ such that their one candidate edge is born and survives at depth h, and their one non-candidate edge survives at depth h. We write $Y^{h+}_{uv}(i+1)$ for its size.

(We're ignoring $\mathcal{Y}^*_{uv}$ and $\mathcal{X}^*_{uv}$ because they're tiny by Claim 3.5, so can be absorbed into our error term.) Then we have

$$Y^{41}_{uv}(i+1) + Y^{41+}_{uv}(i+1) \leq Y_{uv}(i+1) \leq Y^{40}_{uv}(i+1) + Y^{40+}_{uv}(i+1) + Y^*_{uv} + X^*_{uv},$$

so it suffices to show that both $Y^h_{uv}(i+1)$ and $Y^{h+}_{uv}(i+1)$ have the right means and concentrate well.

To compute $\mathbb{E}[Y^h_{uv}(i+1)]$, by Property 3.2 we have $Y'_{uv}(i+1) = 2q_i(p_i \pm (1+p_i)(\Delta_i + \Delta_i\delta_i))\sqrt{n}$, and by Lemma 3.8, for each configuration in $\mathcal{Y}'_{uv}(i+1)$, the probability that its one open edge survives is $\frac{q_{i+1}}{q_i} \cdot (1 \pm 12\Delta_i\delta_i)$. This means we have

$$\begin{aligned}\mathbb{E}[Y^h_{uv}(i+1)] &= 2q_i(p_i \pm (1+p_i)(\Delta_i + \Delta_i\delta_i))\sqrt{n} \cdot \frac{q_{i+1}}{q_i} \cdot (1 \pm 12\Delta_i\delta_i) \\ &= 2q_{i+1}(p_i \pm (1+p_i)(\Delta_i + 14\Delta_i\delta_i))\sqrt{n}.\end{aligned}$$

To compute $\mathbb{E}[Y^{h+}_{uv}(i+1)]$, by Claim 3.5 we have

$$X^{**}_{uv} = 2\lambda q_i^2(1 \pm (\Delta_i + \Delta_i\delta_i))\sqrt{n}.$$

And Lemma 3.8 says that for each configuration in $\mathcal{X}^{**}_{uv}$, the probability it lands in $\mathcal{Y}^{h+}_{uv}(i+1)$ is

$$\frac{\gamma}{\lambda} \cdot \left(\frac{p_{i+1} - p_i}{q_i\gamma}\right)\left(\frac{q_{i+1}}{q_i}\right) \cdot (1 \pm 12\Delta_i\delta_i)^2 = \frac{1}{\lambda} \cdot \frac{(p_{i+1} - p_i)q_{i+1}}{q_i^2} \cdot (1 \pm 12\Delta_i\delta_i)^2$$

(the $\frac{\gamma}{\lambda}$ corresponds to the probability that the one candidate edge gets born). So we get

$$\begin{aligned}\mathbb{E}[Y^{h+}_{uv}(i+1)] &= 2\lambda q_i^2(1 \pm (\Delta_i + \Delta_i\delta_i))\sqrt{n} \cdot \frac{1}{\lambda} \cdot \frac{(p_{i+1} - p_i)q_{i+1}}{q_i^2} \cdot (1 \pm 12\Delta_i\delta_i)^2 \\ &= 2(p_{i+1} - p_i)q_{i+1}(1 \pm (\Delta_i + 26\Delta_i\delta_i))\sqrt{n}.\end{aligned}$$

When we add these together, we'll get that the main term in $\mathbb{E}[Y^h_{uv}(i+1) + Y^{h+}_{uv}(i+1)]$ is

$$2p_iq_{i+1}\sqrt{n} + 2(p_{i+1} - p_i)q_{i+1}\sqrt{n} = 2p_{i+1}q_{i+1}\sqrt{n},$$

which is exactly what we wanted (the fact that the calculation works out corresponds to (1.4)). We do have to be a bit careful about how the errors interact, but they do work out: our new error becomes

$$\frac{2(1+p_i)q_{i+1}(\Delta_i + 14\Delta_i\delta_i) + 2(p_{i+1} - p_i)q_{i+1}(\Delta_i + 26\Delta_i\delta_i)}{2q_{i+1}},$$

and using the fact that $p_{i+1} - p_i \leq \gamma q_i \leq \delta_i$, we get that this is at most

$$(1+p_i)(\Delta_i + 14\Delta_i\delta_i) + 27\Delta_i\delta_i \leq (1+p_{i+1})(\Delta_i + 42\Delta_i\delta_i).$$

So we've shown that $Y_{uv}^h(i+1)$ and $Y_{uv}^{h+}(i+1)$ have the right expectations, i.e.,

$$\mathbb{E}[Y_{uv}^h(i+1) + Y_{uv}^{h+}(i+1)] = 2q_{i+1}(p_{i+1} \pm (1 + p_{i+1})(\Delta_i + 42\Delta_i\delta_i))\sqrt{n}.$$

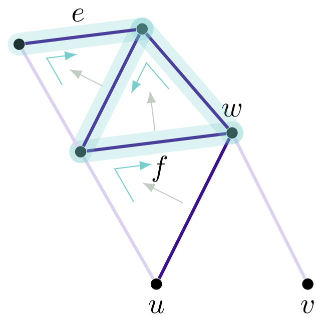
And now it remains to prove that they concentrate well (e.g., within a window of length $n^{1/2-80\varepsilon}$).

We'll start by proving concentration for $Y_{uv}^h(i+1)$. We'll again use the bounded differences inequality; this means we want to understand the number of configurations $\{uw, vw\}$ in $\mathcal{Y}_{uv}(i)$ that each edge e affects (meaning that the birthtime of e could change whether or not the configuration survives at depth h).

First, if an edge e affects a configuration $\{uw, vw\} \in \mathcal{Y}_{uv}(i)$, then there must be a dependency walk from the open edge in this configuration (which we'll assume is uw) to e of length at most h.

First, to bound the *number* of relevant edges e, there are at most $2\sqrt{n}$ choices for w by Property 3.2; then there's at most $2\lambda^2$ choices for each step of the dependency walk, so there's at most $(2\lambda^2)^h \cdot 2\sqrt{n} \leq \lambda^{3h}\sqrt{n}$ relevant edges e.

To bound the number of configurations that each affects, we can imagine taking a dependency walk backwards from e to the edge f which originally came right after uw on the dependency walk; there are again $2\lambda^2$ choices at each step. If f is not incident to u, then there are at most 2 choices for w (it has to be one of the endpoints of f).



Meanwhile, if $f = ux$ is incident to u, then we have one of the following two pictures.

In the first picture, Property 3.3 tells us there are at most $i(\log n)^2$ choices for w, since we must have $\{wx, vw\} \in \mathcal{Z}_{vx}(i)$. In the second, Claim 3.6(iii) tells us there are at most 4λ choices for w.

So in total, there are at most $\lambda^{3h}\sqrt{n}$ edges, each with $c_e \leq (2\lambda^2)^h(i(\log n)^2 + 4\lambda) \leq \lambda^{3h}$. We wanted a concentration window of length $\sigma = n^{1/2-80\varepsilon}$; when we apply the bounded differences inequality, we'll have

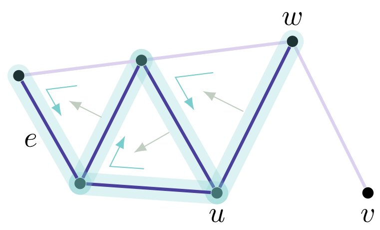
$$\frac{\sigma^2}{\sum_e c_e^2} \geq \frac{n^{1-160\varepsilon}}{\lambda^{3h}\sqrt{n} \cdot \lambda^{6h}} \geq n^{1/2-2^{30}\varepsilon},$$

giving a concentration probability of $1 - \exp(-\Omega(n^{1/2-2^{30}\varepsilon}))$.

The proof of concentration for $Y_{uv}^{h+}(i+1)$ is very similar. To bound the number of relevant edges, this time we use Claim 3.5(iii) to say there are at most $2\lambda\sqrt{n}$ choices for w, and then there's again $2\lambda^2$ choices for each step of the dependency walk.

To bound the number of configurations that each edge e affects, this time there's two cases — uw (which we're assuming is the edge of the configuration that e affects) could either be the candidate edge or the non-candidate edge of $\{uw, vw\}$.

First, if uw is the candidate edge, then we can run the dependency walk backwards from e all the way to uw; this shows there are at most $(2\lambda^2)^h$ choices for w.



Meanwhile, if uw is not the candidate edge, then we again run the dependency walk backwards from e to the edge f that originally came right after uw. If f is not incident to u, there's at most 2 choices for w (namely, the two endpoints of f). If $f = ux$, then we have one of the following two pictures.

Then Claim 3.6(iii) (in the first case) or Claim 3.6(iv) (in the second) tells us that there are at most $2\lambda^2$ choices for w.

So we again get that there are at most $\lambda^{3h}\sqrt{n}$ relevant edges e, each with $c_e \le \lambda^{3h}$; this means we get the same concentration bound for $Y_{uv}^{h+}(i+1)$.

§3.5 Lemma 3.8: Survival probabilities in dependency trees

In this section, we'll prove Lemma 3.8 by using branching processes to analyze our dependency trees. (Throughout this section, we think of H and $h \in \{40, 41\}$ as fixed.)

§3.5.1 Making the trees independent

The first step of the argument is that it's not great that the dependency trees $\mathcal{T}_h(e)$ are allowed to contain repeated edges — we'd really like to analyze these trees as branching processes, and to do so we'd want their nodes to have *independent* birthtimes (which isn't the case if an edge appears in multiple places). Fortunately, it turns out that we can fix this by running a second round of sampling.

Set $\nu = n^{2^{10}\varepsilon}$. We define a subset $F' \subseteq F^*$ by including each edge with probability $\frac{\nu}{\lambda}$; we call the edges in F' *refined candidates*. When generating the birthtimes for edges, we'll only work with the edges in F' — so for each $e \in F'$ we generate a birthtime in $[0, \nu]$ uniformly at random (and it gets born if its birthtime is at most γ), while we ignore edges $e \notin F'$.

Once we've fixed F', only the edges in F' matter for the analysis of our dependency trees. So we define sets $\mathcal{Y}'_{uv}$ and $\mathcal{X}'_{uv}$ analogously to $\mathcal{Y}^*_{uv}$ and $\mathcal{X}^*_{uv}$, but with F' in place of F^* (so $\mathcal{Y}'_{uv}$ consists of configurations

$\{uw, vw\} \in \mathcal{Y}^*_{uv}$ whose one open edge is a refined candidate, and $\mathcal{X}'_{uv}$ consists of configurations in $\mathcal{X}^*_{uv}$ whose two open edges are refined candidates). As usual, we write Y'_{uv} and X'_{uv} for the sizes of these sets. We also define refined dependency trees $\mathcal{T}'_h(e)$ analogously to $\mathcal{T}_h(e)$, by restricting to the edges in F' and the configurations in $\mathcal{Y}'_{uv}$ and $\mathcal{X}'_{uv}$.

Claim 3.10 — With probability $1 - n^{-\omega(1)}$, the following statements hold for all $uv \notin G_i$:

(i) $Y'_{uv} = 2\nu q_i(p_i \pm 2(1 + p_i)\Delta_i)$.

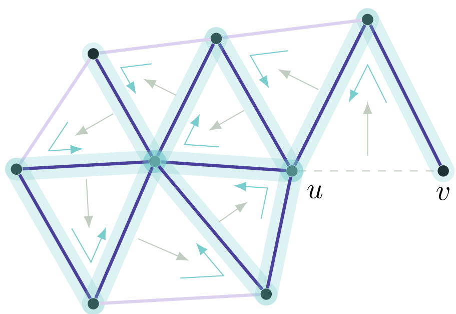
(ii) $X'_{uv} = \nu^2 q_i^2 (1 \pm 2\Delta_i)$.

These statements both follow from Claim 3.5 (which means the *expected* values of Y'_{uv} and X'_{uv} are roughly $2\lambda q_i p_i \cdot \frac{\nu}{\lambda}$ and $\lambda^2 q_i^2 \cdot \frac{\nu^2}{\lambda^2}$, respectively) and multiplicative Chernoff (which gives good concentration).

Claim 3.11 — With probability at least $1 - n^{-64\varepsilon}$, none of the trees $\mathcal{T}'_h(e)$ for $e \in H$ contain repeated edges, and no two of these trees share edges.

Proof sketch. The intuition is as follows: Imagine we consider the expected number of dependency walks (of length at most $2h$) that start at an edge in H and either return to an edge they've already visited, or reach some other edge in H. At every step, there are roughly λ or λ^2 configurations we could *possibly* move to (among the original candidate edges) of $\mathcal{Y}$-type and $\mathcal{X}$-type, respectively, and each appears among the refined candidate edges with probability $\frac{\nu}{\lambda}$ or $\frac{\nu^2}{\lambda^2}$, respectively; this means the *expected* number of steps we could take is roughly $\nu + 2\nu^2 \leq 3\nu^2$. And on the last step, in order to return to an already visited edge or an edge in H, we'll no longer have the factor of λ or λ^2 for the next step (now there's a *constant* number of steps), but we'll still have at least one factor of $\frac{\nu}{\lambda}$ (corresponding to some edge needing to be picked). So we'll get that the expected number of such walks is at most roughly a constant times $(3\nu^2)^{2h} \cdot \frac{\nu}{\lambda}$; and since λ is a much bigger power of n^ε than ν is, this is at most $n^{-64\varepsilon}$.

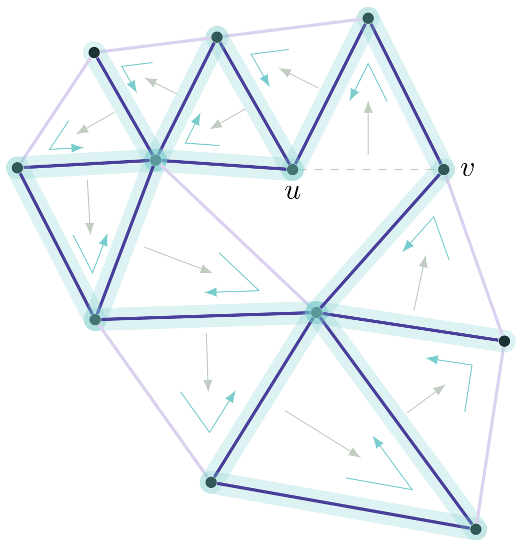
Formalizing this intuition takes a bit of care, because we need to worry about our dependency walk possibly repeating edges (in which case we wouldn't get a $\frac{\nu}{\lambda}$ factor for that edge the second time we used it). One way to do so is as follows: Consider a *shortest* dependency walk that starts from an edge in H, and returns to a *vertex* that either had been previously visited by the walk or is an endpoint of an edge in H. (When we take a dependency walk, every edge shares exactly one vertex with the previous one. We don't count that shared vertex — the 'previously visited' case means there's some vertex u that the walk visits, then leaves, and then returns to.)



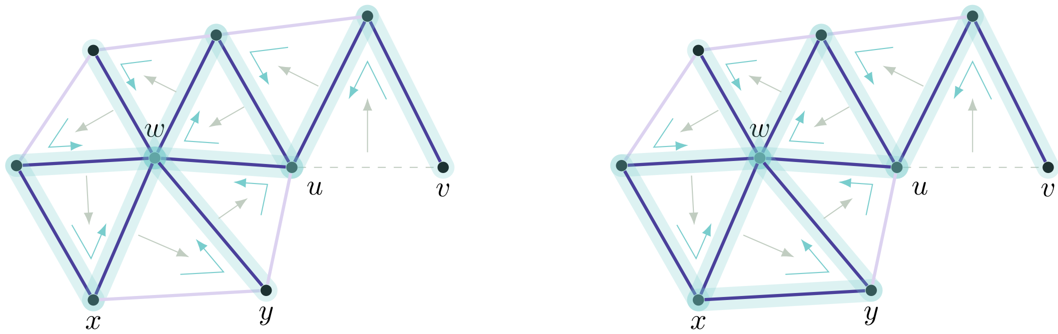
(For example, the above picture shows a walk that starts at an edge uv, takes 8 steps, and revisits u.)

First, apart from the final step, the configurations we see on this walk can't share any edges (if they did, we'd get a shorter walk that revisited a vertex). If the final step is a $\mathcal{X}$-type configuration (as in the above picture), then it can't be the case that we've used *both* of its edges already (again, this would give a shorter walk that revisited a vertex). So the above argument does still work — we do have an expected number of at most $3\nu^2$ choices for all but the last step, and an expected number of at most $\frac{\nu}{\lambda}$ choices for the last step (up to constants), since there's a constant number of choices for the revisited vertex and some new edge has to be picked (which occurs with probability $\frac{\nu}{\lambda}$).

If the final step is a $\mathcal{Y}$-type configuration and we hadn't used its one candidate edge already, then the above argument still works (for the same reason).



The only 'bad' case (where the above argument doesn't work) is if the final step is a $\mathcal{Y}$-type configuration and we have used its one candidate edge already, as in the below pictures.

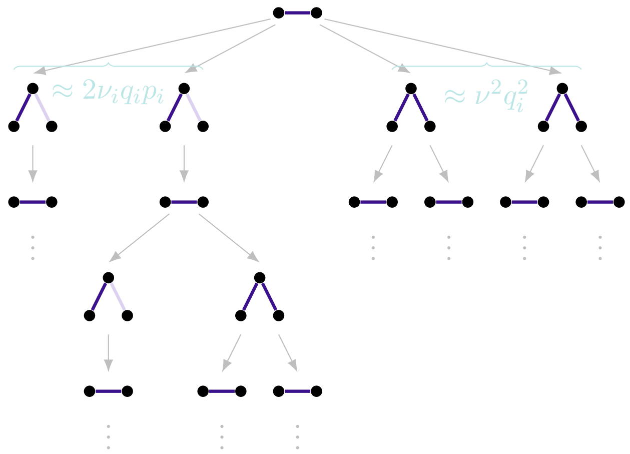


Let the reused edge be uw (where u is the revisited vertex), the third vertex of that configuration be y, and the third vertex of our previous configuration be x (as shown above). Then given x, w, and u, by either Claim 3.6(i) or Claim 3.6(ii) there's at most $(\log n)^2$ choices for y, as opposed to the bound of λ or λ^2 we used earlier. So this again lets us get an extra factor of λ in the denominator. □

Since $n^{-64\varepsilon}$ is much smaller than our target error in Lemma 3.8 (which is roughly $\Delta_i\delta_i$), we can afford to ignore the case where Claim 3.11 fails. So from now on, we'll assume that we've fixed F' such that Claims 3.10 and 3.11 hold. Then the dependency trees $\mathcal{T}'_h(e)$ across different edges $e \in H$ are independent, so to prove Lemma 3.8, it suffices to consider just a *single* tree (with no repeated edges).

§3.5.2 A branching process problem

Now we've reduced Lemma 3.8 to the following branching process problem: We're given a tree $\mathcal{T}$ of height h (whose nodes represent edges in our graph). Each node has $2\nu q_i(p_i \pm 2(1+p_i)\Delta_i)$ $\mathcal{Y}$-type children, each leading to one node on the next level, and $\nu^2 q_i^2(1 \pm 2\Delta_i)$ $\mathcal{X}$-type children, each leading to two nodes on the next level. (These quantities come from Claim 3.10.)



Every node receives a birthtime in $[0, \nu]$ uniformly at random; it's born if its birthtime is in $[0, \gamma]$. A node with birthtime t survives if it does not have any child configuration which is fully born by time $\min\{t, \gamma\}$ and fully survives. Our goal is to understand the probability that the root survives. So we'll let $p_t(\mathcal{T})$ be the probability that the root survives, given that its birthtime is t.

Claim 3.12 — For all $t \in [0, \gamma]$, we have

$$p_t(\mathcal{T}) = \frac{\psi(\gamma i + t)}{\psi(\gamma i)} \pm 10\Delta_i \delta_i.$$

As discussed after the statement of Lemma 3.8, the reason we'd intuitively expect this to be the right answer is that $p_t(\mathcal{T})$ is supposed to model the probability that in the triangle-free process, an edge e which was open at time γi remains open at time $\gamma i + t$, and we'd heuristically expect the density of open edges to drop from $\psi(\gamma i)$ to $\psi(\gamma i + t)$ between these two times (based on Subsection 1.2).

Claim 3.12 implies Lemma 3.8 for one edge e — the case where we condition on e not being born corresponds to taking $p_\gamma(\mathcal{T})$ (note that $p_t(\mathcal{T}) = p_\gamma(\mathcal{T})$ if $t \geq \gamma$), where we get

$$\frac{\psi(\gamma i + \gamma)}{\psi(\gamma i)} = \frac{q_{i+1}}{q_i},$$

and the case where we condition on e being born corresponds to averaging over $x \in [0, \gamma]$, where we get

$$\frac{1}{\gamma} \int_0^\gamma \frac{\psi(\gamma i + t)}{\psi(\gamma i)}\, dt = \frac{p_{i+1} - p_i}{\gamma q_i}.$$

(Lemma 3.8 is written with multiplicative rather than additive error, but these probabilities are very close to 1, so this doesn't matter.) And since Claim 3.11 means that different edges $e \in H$ are independent, this also implies Lemma 3.8 for multiple edges. So now it just remains to prove this claim, which we'll do in the next sections (we'll be a bit sketchy with the proof, but it can be made rigorous).

§3.5.3 An infinite idealized branching process

Consider an infinite tree $\mathcal{T}_\infty$ of the same form as in our problem, where every node has *exactly* $2\nu q_i p_i$ $\mathcal{Y}$-type children and $\nu^2 q_i^2$ $\mathcal{X}$-type children. We're first going to show that Claim 3.12 holds for this infinite tree.

First, it's not immediately clear that our process is even well-defined for such an infinite tree. However, it turns out that the process has good correlation decay, so that it *is* well-defined.

Claim 3.13 — Let $\mathcal{T}_\infty^{\ell\to\tau}$ be the tree where we cut $\mathcal{T}_\infty$ off at height ℓ, sample birthtimes for all nodes as normal, and declare whether the nodes at level h survive based on some assignment τ. Then

$$\left|p_t(\mathcal{T}_\infty^{\ell\to\tau_1}) - p_t(\mathcal{T}_\infty^{\ell\to\tau_2})\right| \leq (6\gamma)^\ell$$

for all t and all assignments τ_1 and τ_2.

Proof. Imagine we take $\mathcal{T}_\infty^{\ell\to\tau_1}$ and $\mathcal{T}_\infty^{\ell\to\tau_2}$, and we couple them by assigning the same birthtimes to all nodes. Then the only way that they could possibly have different outcomes for the root is if there's a path from the root to level ℓ along which all ℓ configurations are fully born. (If there's a configuration somewhere in the tree that isn't fully born, then it doesn't affect whether its parent survives, so its subtree is irrelevant.)

But we claim that the *expected* number of such paths is at most $(6\gamma)^\ell$. To see this, every node in the tree representing an edge has at most 2ν $\mathcal{Y}$-type children, each of which is born with probability $\frac{\gamma}{\nu}$ (and offers one edge to step to), and at most $2\nu^2$ $\mathcal{X}$-type children, each of which is fully born with probability $(\frac{\gamma}{\nu})^2$ (and offers two edges to step to). This means if we're trying to walk down such a path from the root, the expected number of choices we have at each step is at most

$$2\nu \cdot \frac{\gamma}{\nu} \cdot 1 + 2\nu^2 \cdot \frac{\gamma^2}{\nu^2} \cdot 2 \leq 6\gamma.$$

And we're doing this ℓ times, so the expected number of paths we end up with is at most $(6\gamma)^\ell$. Then by Markov's inequality, the probability there exists some such path is at most $(6\gamma)^\ell$. □

Now, we say a node follows the *target survival distribution* if conditioned on its birthtime being $t \in [0,\gamma]$, the probability it survives is $\frac{\psi(\gamma i+t)}{\psi(\gamma i)}$.

Claim 3.14 — Suppose that for some node e, all edges in its children follow the target survival distribution. Then e follows the target survival distribution as well.

Remark 3.15. Technically, this is only true up to an error on the order of $\frac{1}{\nu} = n^{-2^{10}\varepsilon}$, but this won't matter for our purposes — this is tiny compared to our target error, and we can deal with it in the same way as we'll deal with the error terms in the number of children when proving Claim 3.16.

Proof. Suppose we condition on e having birthtime $t \in [0,\gamma]$.

Then for e to survive, we need it to *not* be the case that any of its children configurations gets fully born at some time $0 \leq s \leq t$ and fully survives. For each of its $\mathcal{Y}$-type children, the probability that its one edge is born at some time $0 \leq s \leq t$ and survives is

$$\frac{1}{\nu}\int_0^s \frac{\psi(i\gamma+s)}{\psi(i\gamma)}\,ds = \frac{1}{\nu}\cdot\frac{\Psi(i\gamma+t)-\Psi(i\gamma)}{\psi(i\gamma)}$$

(since we assumed it follows the target survival distribution). For each of its $\mathcal{X}$-type children, we're asking this to happen for *two* edges (which are independent); so the corresponding probability is

$$\frac{1}{\nu^2}\cdot\frac{(\Psi(i\gamma+t)-\Psi(i\gamma))^2}{\psi(i\gamma)^2}.$$

Since our node e has $2\nu\Psi(i\gamma)\psi(i\gamma)$ $\mathcal{Y}$-type children and $\nu^2\psi(i\gamma)^2$ $\mathcal{X}$-type children, this means the probability it survives — meaning that none of these events occur across all its children — is

$$\mathbb{P}[e \text{ survives}] = \left(1 - \frac{1}{\nu}\cdot\frac{\Psi(i\gamma+t)-\Psi(i\gamma)}{\psi(i\gamma)}\right)^{2\nu\Psi(i\gamma)\psi(i\gamma)}\left(1-\frac{1}{\nu^2}\cdot\frac{(\Psi(i\gamma+t)-\Psi(i\gamma))^2}{\psi(i\gamma)^2}\right)^{\nu^2\psi(i\gamma)^2}.$$

Using the approximation $1 - x \approx e^{-x}$ (we can afford to do so because $\nu = n^{2^{10}\varepsilon}$ much larger than the error we want in Lemma 3.8; see Remark 3.15), we get that

$$\begin{aligned}\mathbb{P}[e \text{ survives}] &= \exp\left(\frac{1}{\nu} \cdot \frac{\Psi(i\gamma + t) - \Psi(i\gamma)}{\psi(i\gamma)} \cdot 2\nu\Psi(i\gamma)\psi(i\gamma) - \frac{1}{\nu^2} \cdot \frac{(\Psi(i\gamma + t) - \Psi(i\gamma))^2}{\psi(i\gamma)^2} \cdot \nu^2\psi(i\gamma)^2\right) \\ &= \exp\left(\Psi(i\gamma)^2 - \Psi(i\gamma + t)^2\right).\end{aligned}$$

And because $\exp(-\Psi(t)^2) = \psi(t)$, this is precisely $\frac{\psi(i\gamma+t)}{\psi(i\gamma)}$. □

This essentially shows that Claim 3.12 holds for $\mathcal{T}_\infty$ (with a much smaller error term) — we can imagine cutting $\mathcal{T}_\infty$ off at a sufficiently low level and initializing the leaves with the target survival distribution (which we're allowed to do by Claim 3.13), and Claim 3.14 shows that all intermediate nodes will also have this distribution.

§3.5.4 Comparison to the idealized process

Finally, it remains to show that moving from $\mathcal{T}_\infty$ to our given tree $\mathcal{T}$ — which has errors in the number of children — doesn't introduce too much error.

> **Claim 3.16** — For all t, we have $|p_t(\mathcal{T}_\infty) - p_t(\mathcal{T})| \leq 8\Delta_i\delta_i$.

Proof. Imagine we create a sequence of trees $\mathcal{T}_0$, $\mathcal{T}_1$, ..., $\mathcal{T}_h$ as follows: We take $\mathcal{T}_\ell$ to match our tree $\mathcal{T}$ up to height ℓ, and at each node of height ℓ, we place an infinite tree $\mathcal{T}_\infty$ rooted at that node. In particular, we have $\mathcal{T}_0 = \mathcal{T}_\infty$, and $\mathcal{T}_h$ is the same as $\mathcal{T}$ except that we've added infinite trees rooted at each leaf of $\mathcal{T}_h$.

We already saw in Claim 3.13 that $|p_t(\mathcal{T}_h) - p_t(\mathcal{T})| \leq (6\gamma)^h$ (which is much smaller than $\Delta_i\delta_i$), so it suffices to show that $p_t(\mathcal{T}_\ell)$ and $p_t(\mathcal{T}_{\ell+1})$ are similar for all ℓ.

To do so, we'll use coupling. Note that $\mathcal{T}_\ell$ and $\mathcal{T}_{\ell+1}$ look the same up to level ℓ, and only differ in the number of children that each node at level ℓ has (for $\mathcal{T}_\ell$ these numbers are exactly $\nu q_i p_i$, and for $\mathcal{T}_{\ell+1}$ they're given by $\mathcal{T}$). These numbers can differ by up to $4\nu(1 + p_i)q_i\Delta_i$ for $\mathcal{Y}$-type children, and $2\nu^2 q_i^2\Delta_i$ for $\mathcal{X}$-type children. We can imagine matching up the children of each level-ℓ node in $\mathcal{T}_\ell$ and $\mathcal{T}_{\ell+1}$, so each node has at most this many unmatched children of each type.

Then we can couple $\mathcal{T}_\ell$ and $\mathcal{T}_{\ell+1}$ by assigning the same birthtimes to all nodes up to level ℓ, and also coupling the infinite trees rooted at matched children of the nodes at level ℓ.

In this coupling, the only way the outcomes at the root can be different in the two trees is if there is a path from the root to one of the unmatched children along which all configurations, including that unmatched child, are fully born — this is because otherwise we can cut off the unmatched children without affecting the survival of the root (and the two trees are identical apart from these unmatched children).

And we can show the expected number of such paths is small, by the same argument as in Claim 3.13 — imagine we compute the expected number of such paths by seeing the expected number of steps we can take from every node (starting with the root). For each step but the last, the current node has at most 2ν possible steps to $\mathcal{Y}$-type children, each of which is born with probability $\frac{\gamma}{\nu}$ (and gives one edge to step to) and at most $2\nu^2$ possible steps to $\mathcal{X}$-type children, each of which is born with probability $\frac{\gamma^2}{\nu^2}$ (and gives two edges to step to); so the expected number of steps we can take is at most 6γ. For the last step, since we need to step to an unmatched child, we instead have at most $4\nu(1 + p_i)q_i\Delta_i$ possible steps to $\mathcal{Y}$-type children and $2\nu^2 q_i^2\Delta_i$ possible steps to $\mathcal{X}$-type children (which again come with probabilities $\frac{\gamma}{\nu}$ and $\frac{\gamma^2}{\nu^2}$), so the expected number of steps we can take is at most

$$4\gamma(1 + p_i)q_i\Delta_i + 4\gamma^2 q_i^2\Delta_i \leq 8\Delta_i\delta_i$$

(by how we defined δ_i). So in total, the expected number of paths — and therefore the probability there exists one — is at most $(6\gamma)^\ell \cdot 8\Delta_i\delta_i$.

Finally, we've shown that $|p_t(\mathcal{T}_\ell) - p_t(\mathcal{T}_{\ell+1})| \leq (6\gamma)^\ell \cdot 8\Delta_i\delta_i$ for all ℓ; this shows that

$$|p_t(\mathcal{T}_0) - p_t(\mathcal{T}_h)| \leq 8\Delta_i\delta_i \sum_{\ell=0}^{h-1} (6\gamma)^\ell \leq 9\Delta_i\delta_i$$

(since 6γ is much smaller than 1). And $\mathcal{T}_0 = \mathcal{T}_\infty$, while we've seen $|p_t(\mathcal{T}_h) - p_t(\mathcal{T})| \leq (6\gamma)^h$ (which is much smaller than $\Delta_i\delta_i$), which proves the claim. □

§4 Bohman 2009: An analysis via the differential equation method

Finally, in this section we'll explain Bohman's analysis of the triangle-free process from [Boh09], based on the differential equation method. We'll give his proof that the process follows the trajectory from Subsection 1.3 up to time n^c, which illustrates a different approach to analyzing the process than the one in Section 3.

§4.1 Overview

In this analysis, we'll work with the triangle-free process step by step. So we begin with the graph G_0. Given G_i, we choose a random edge e_{i+1}; if it's open, we add it to G_i to produce G_{i+1} (and otherwise we discard it). We'll think of step i as corresponding to time $t = 2n^{-3/2}i$.

Let $\varepsilon > 0$ be a somewhat small constant, and let c be small with respect to ε. As in Subsection 1.3, we write

$$\varphi_x(t) = \psi(t)^2 \quad \text{and} \quad \varphi_y(t) = 2\Psi(t)\psi(t).$$

We also define slowly deteriorating error terms

$$g_x(t) = n^{-\varepsilon} \cdot \psi(t)e^{40\Psi(t)^2+40\Psi(t)} \quad \text{and} \quad g_y(t) = n^{-\varepsilon} \cdot e^{40\Psi(t)^2+40\Psi(t)}.$$

Note that $\Psi(t) \approx \sqrt{\log t}$, so as long as c is small relative to ε, we'll have that $e^{40\Psi(t)^2+40\Psi(t)} \approx e^{40c\log n} \ll n^\varepsilon$, which means these error terms remain small. (The place these error terms come from is that when we try to run the argument, we'll see that we need precisely the conditions in Fact 4.6 (up to constants), and then we try to engineer error terms satisfying them.)

The statement we'll prove is the following (where $t = 2n^{-3/2}i$).

Property 4.1. For all edges $uv \notin G_i$, we have $X_{uv}(i) = (\varphi_x(t) \pm g_x(t))n$.

Property 4.2. For all edges $uv \notin G_i$, we have $Y_{uv}(i) = (\varphi_y(t) \pm g_y(t))\sqrt{n}$.

Property 4.3. For all edges uv, we have $Z_{uv}(i) \leq n^\varepsilon$.

Theorem 4.4
With probability $1 - n^{-\omega(1)}$, Properties 4.1–4.2 hold for all $i \leq n^{3/2+c}$ (where $t = 2n^{-3/2}i$).

In the analysis we saw in Sections 2 and 3, we split the process into large chunks of time and showed that on each chunk, the changes in these random variables concentrated around their expectations (where the expected changes correspond to the differential equations (1.3) and (1.4)); we used this to show that they remain on the expected trajectory from one chunk to the next.

In contrast, here we'll look at how these random variables change on each individual step of the process. Of course we can't say that these one-step changes concentrate around their expectations. Instead, we define martingales that *accumulate* how much these one-step changes differ from their expectations over the entire process. We can use martingale concentration inequalities to show that these martingales remain small throughout the process with high probability; and if they do, this means our random variables really follow the expected trajectory. This turns out to result in a very neat proof.

§4.2 Preliminaries

We'll use the following martingale concentration inequality.

Lemma 4.5

Let $a \le b$. Then for any martingale $Z_0, Z_1, \ldots$ with $Z_0 = 0$ such that $-a \le Z_{i+1} - Z_i \le b$ for all i,

$$\mathbb{P}[|Z_i| \ge \sigma] \le 2\exp\left(-\frac{1}{8}\min\left\{\frac{\sigma}{b}, \frac{\sigma^2}{iab}\right\}\right).$$

(This is essentially a martingale version of Lemma 2.12, with uniform interval sizes, and it can be proved in the same way.)

We'll also need the following fact about the error functions g_x and g_y.

Fact 4.6 — For all $t \ge 0$, we have

- $g_x(t) \ge 30\int_0^s g_y(s)\varphi_x(s)\,ds$ and $g_x(t) \ge 30\int_0^s g_x(s)\varphi_y(s)\,ds$;
- $g_y(t) \ge 30\int_0^s g_y(s)\varphi_y(s)\,ds$ and $g_y(t) \ge 30\int_0^s g_x(s)\,ds$.

Proof idea. These can be verified by a straightforward computation — in all, we either pull out the $e^{40\Psi(t)}$ terms (noting that $e^{40\Psi(t)} \ge e^{40\Psi(s)}$ for all $s \le t$, so if the inequality is true without these terms, it's also true with them) and substitute $u = \Psi(s)^2$, or pull out the $e^{40\Psi(t)^2}$ terms and substitute $u = \Psi(s)$. □

§4.3 Proof of Theorem 4.4

We'll now describe the structure of the argument (since this is a bit intricate). For each i, we define a 'bad' event $\mathcal{B}_i$ that one of Properties 4.1–4.3 fails to hold for some $j \le i$. We'll define a bunch of martingales (which are supposed to follow the process unless $\mathcal{B}_i$ occurs, at which point we stop them). We'll show that these martingales concentrate well with probability $1 - n^{-\omega(1)}$ for each i (regardless of whether $\mathcal{B}_i$ occurs or not). Let $\mathcal{M}_i$ be the event that any one of these martingales goes outside its concentration window at some time $j \le i$; so the probability that any $\mathcal{M}_i$ occurs is tiny. We'll also show that if neither $\mathcal{B}_{i-1}$ nor $\mathcal{M}_i$ occurs, then $\mathcal{B}_i$ does not occur either (i.e., Properties 4.1–4.3 all hold for i). Since of course $\mathcal{B}_0$ does not occur, this means that as long as no $\mathcal{M}_i$ occurs (which is true with very high probability), no $\mathcal{B}_i$ occurs either (up to $i = n^{3/2+c}$).

§4.3.1 Property 4.1: Tracking $\mathcal{X}_{uv}(i)$

In this section, we'll run the portion of the argument corresponding to the sets $\mathcal{X}_{uv}(i)$. Specifically, we'll define the collection of associated martingales, show that these martingales concentrate (in windows of length $n^{1-4\varepsilon}$) with high probability, and show that if $\mathcal{B}_{i-1}$ does not occur and the martingale associated with uv concentrates, then Property 4.1 holds for uv for step i.

For each step of the process, we define $\mathcal{X}_{uv}^-(i+1)$ as the set of configurations in $\mathcal{X}_{uv}(i)$ that leave because we either chose or closed one of their open edges on step $i+1$, and $X_{uv}^-(i+1)$ as its size; then we have

$$\mathcal{X}_{uv}(i+1) = \mathcal{X}_{uv}(i) \setminus \mathcal{X}_{uv}^-(i+1) \quad \text{and} \quad X_{uv}(i+1) = X_{uv}(i) - X_{uv}^-(i+1).$$

We define a martingale $M_{uv}(i)$ by $M_{uv}(0) = 0$ and

$$M_{uv}(i+1) - M_{uv}(i) = \begin{cases} X_{uv}^-(i+1) - \mathbb{E}[X_{uv}^-(i+1) \mid G_i] & \text{off the event } \mathcal{B}_i \\ 0 & \text{on the event } \mathcal{B}_i. \end{cases}$$

In words, $M_{uv}(i)$ accumulates how much the one-step changes in $X_{uv}(i)$ differ from their expectations until $\mathcal{B}_i$ occurs, at which point it stops. In particular, off $\mathcal{B}_{i-1}$ we have

$$X_{uv}(i) = (n-2) - \sum_{j=0}^{i-1} \mathbb{E}[X_{uv}^-(j+1) \mid G_j] - M_{uv}(i)$$

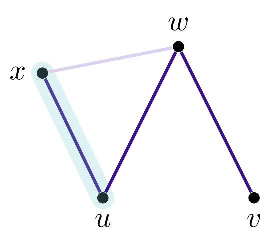
(where $n-2$ corresponds to $X_{uv}(0)$).

Claim 4.7 — Off the event $\mathcal{B}_i$, we have

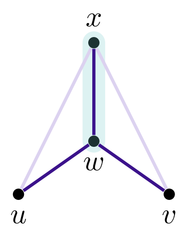
$$\mathbb{E}[X_{uv}^-(i+1) \mid G_i] = 4(\varphi_x(t)\varphi_y(t) \pm 2g_y(t)\varphi_x(t) \pm 2g_x(t)\varphi_y(t))n^{-1/2}.$$

(We always write t to mean $2n^{-3/2}i$.)

Proof. Consider some configuration $\{uw, vw\} \in \mathcal{X}_{uv}(i)$. For it to land in $\mathcal{X}_{uv}^-(i+1)$ because we closed one of its edges, the edge e_{i+1} we picked must be the open edge of some configuration in $\mathcal{Y}_{uw}(i)$ or $\mathcal{Y}_{vw}(i)$.



By Property 4.2, there are $(\varphi_y(t) \pm g_y(t))\sqrt{n}$ choices for e_{i+1} that would close uw, and the same is true for vw. We have to be a bit careful because some edges coud close both. But if choosing an edge wx would close both, then we must have $\{ux, vx\} \in \mathcal{Z}_{uv}(i)$; and by Property 4.3 there are very few such edges. So we can account for them by slightly enlarging the error terms.



There's also two ways for $\{uw, vw\}$ to land in $\mathcal{X}_{uv}(i)$ because we *chose* one of its edges to be e_{i+1}, but we can account for this by slightly enlarging the error terms as well.

So for each $\{uw, vw\} \in \mathcal{X}_{uv}(i)$, the probability it lands in $\mathcal{X}^-_{uv}(i+1)$ is roughly

$$\frac{2 \cdot (\varphi_y(t) \pm g_y(t))\sqrt{n}}{\frac{1}{2}n^2} = 4(\varphi_y(t) \pm g_y(t))n^{-3/2}$$

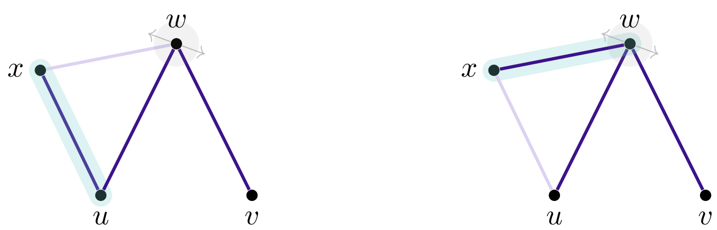
(since there are roughly $\frac{1}{2}n^2$ possible edges we could choose). And there are $(\varphi_x(t) \pm g_x(t))n$ configurations in $\mathcal{X}_{uv}(i)$ by Property 4.1 for i, so multiplying these, we get

$$\mathbb{E}[X^-_{uv}(i+1) \mid G_i] = 4(\varphi_x(t)\varphi_y(t) \pm 2g_y(t)\varphi_x(t) \pm 2g_x(t)\varphi_y(t))n^{-1/2}$$

(where multiplying the error terms by 2 accounts for the lower-order contributions discussed above). □

Claim 4.8 — Off the event $\mathcal{B}_i$, we always have $X^-_{uv}(i+1) \le 2\sqrt{n}$.

Proof. We want to bound the maximum possible number of configurations in $\mathcal{X}_{uv}(i)$ that choosing a single edge e could affect.



If e is incident to u, then letting $e = ux$, choosing e can only affect a configuration $\{uw, vw\}$ if $\{uw, xw\} \in \mathcal{Y}_{ux}(i)$; by Property 4.1 there are at most $2\sqrt{n}$ such configurations. If e is not incident to u or v, then it affects at most one configuration. □

Claim 4.9 — For all uv and all $i \le n^{3/2+c}$, we have $|M_{uv}(i)| \le n^{1-4\varepsilon}$ with probability $1 - n^{-\omega(1)}$.

(This means we'll define $\mathcal{M}_i$ such that it includes the events that $|M_{uv}(j)| > n^{1-4\varepsilon}$ for any $j \le i$.)

Proof. We'll use Lemma 4.5 — we have

$$-2n^{-1/2} \le M_{uv}(j+1) - M_{uv}(j) \le 2\sqrt{n}$$

for all j (where the lower bound comes from Claim 4.7 and the upper bound from Claim 4.8), so applying Lemma 4.5 with $\sigma = n^{1-4\varepsilon}$, we have

$$\frac{\sigma}{b} \ge \frac{n^{1-4\varepsilon}}{2\sqrt{n}} \ge n^{1/2-5\varepsilon} \quad \text{and} \quad \frac{\sigma^2}{iab} \ge \frac{n^{2-8\varepsilon}}{n^{3/2+c} \cdot 2n^{-1/2} \cdot 2\sqrt{n}} \ge n^{1/2-10\varepsilon},$$

which means we get concentration with probability $1 - \exp(-\Omega(n^{1/2-10\varepsilon}))$. □

Claim 4.10 — Suppose that $\mathcal{B}_{i-1}$ does not occur and $|M_{uv}(i)| \leq n^{1-4\varepsilon}$ (and $i \leq n^{3/2+c}$). Then

$$X_{uv}(i) = (\varphi_x(t) \pm g_x(t))n.$$

Proof. Off the event $\mathcal{B}_{i-1}$, by Claim 4.7 we have

$$\begin{aligned} X_{uv}(i) &= X_{uv}(0) - \sum_{j=0}^{i-1} \mathbb{E}[X_{uv}^-(j+1) \mid G_j] - M_{uv}(i) \\ &= (n-2) - \sum_{j=0}^{i-1} 4(\varphi_x(s)\varphi_y(s) \pm 2g_y(s)\varphi_x(s) \pm 2g_x(s)\varphi_y(s))n^{-1/2} - M_{uv}(i). \end{aligned}$$

Plugging in $|M_{uv}(i)| \leq n^{1-4\varepsilon}$ and normalizing by n, we get

$$\frac{X_{uv}(i)}{n} = 1 - \sum_{j=0}^{i-1} 2(\varphi_x(s)\varphi_y(s) \pm 2g_y(s)\varphi_x(s) \pm 2g_x(s)\varphi_y(s)) \cdot 2n^{-3/2} \pm 2n^{-4\varepsilon}$$

(where $s = 2n^{-3/2}j$ is the time corresponding to j). We can convert this sum into an integral over s (recall that one step of the process corresponds to a time-interval of length $2n^{-3/2}$); this introduces an error of at most $2n^{-3/2}$, which can be absorbed into the $\pm 2n^{-4\varepsilon}$ error term. So we get

$$\frac{X_{uv}(i)}{n} = 1 - \int_0^t 2(\varphi_x(s)\varphi_y(s) \pm 2g_y(s)\varphi_x(s) \pm 2g_y(x)\varphi_y(s))\, ds \pm 3n^{-4\varepsilon}.$$

For the main term, (1.3) means that we have

$$\varphi_x(t) = 1 - \int_0^t 2\varphi_x(s)\varphi_y(s)\, ds.$$

To deal with the error terms, Fact 4.6 means their total contribution is at most

$$4\int_0^t g_y(s)\varphi_x(s)\, ds + 4\int_0^t g_y(s)\varphi_y(s)\, ds \leq \frac{1}{3}g_x(t)$$

(and $n^{-4\varepsilon} \ll g_x(t)$). So we get $X_{uv}(i) = (\varphi_x(t) \pm g_x(t))n$, as desired. □

This concludes the part of the argument corresponding to the $\mathcal{X}_{uv}(i)$ terms.

§4.3.2 Property 4.2: Tracking $\mathcal{Y}_{uv}(i)$

In this section, we'll explain the portion of the argument corresponding to the sets $\mathcal{Y}_{uv}(i)$ (we'll do this more briefly, since it uses the same ideas). This time, we'll define two collections of martingales, one to track configurations entering $\mathcal{Y}_{uv}(i)$ and the other to track configurations leaving it:

- First, we define $\mathcal{Y}_{uv}^-(i+1)$ as the set of configurations that leave $\mathcal{Y}_{uv}(i)$ because we either chose or closed their one open edge on step $i+1$, and $Y_{uv}^-(i+1)$ as its size. (If e_{i+1} is uv itself, then we instead define $\mathcal{Y}_{uv}^-(i+1) = \emptyset$.)
- We define $\mathcal{Y}_{uv}^+(i+1)$ as the set of configurations that come in from $\mathcal{X}_{uv}(i)$ because we chose one of their open edges, and Y_{uv}^+ as its size. Note that this is always either 0 or 1, since e_{i+1} belongs to at most one configuration in $\mathcal{X}_{uv}(i)$.

Then as long as $uv \notin G_{i+1}$, we have

$$Y_{uv}(i+1) = Y_{uv}(i) - Y_{uv}^-(i+1) + Y_{uv}^+(i+1).$$

We define a martingale $M_{uv}(i)$ by $M_{uv}(0) = 0$ and

$$M_{uv}(i+1) - M_{uv}(i) = \begin{cases} Y_{uv}^-(i+1) - \mathbb{E}[Y_{uv}^-(i+1) \mid G_i] & \text{off the event } \mathcal{B}_i \\ 0 & \text{on the event } \mathcal{B}_i. \end{cases}$$

Similarly, we define a martingale $N_{uv}(i)$ by $N_{uv}(0) = 0$ and

$$N_{uv}(i+1) - N_{uv}(i) = \begin{cases} Y_{uv}^+(i+1) - \mathbb{E}[Y_{uv}^+(i+1) \mid G_i] & \text{off the event } \mathcal{B}_i \\ 0 & \text{on the event } \mathcal{B}_i. \end{cases}$$

These martingales accumulate how much the stepwise increases and decreases in $Y_{uv}^+(i)$ differ from their expectations; in particular, as long as $uv \notin G_i$ and $\mathcal{B}_{i-1}$ doesn't occur, we have

$$Y_{uv}(i) = -\sum_{j=0}^{i-1} \mathbb{E}[Y_{uv}^-(j+1) \mid G_j] + \sum_{j=0}^{i-1} \mathbb{E}[Y_{uv}^+(j+1) \mid G_j] - M_{uv}(i) + N_{uv}(i). \tag{4.1}$$

To compute the expectations of these one-step changes, as long as $uv \notin G_i$ and $\mathcal{B}_i$ doesn't occur, we have

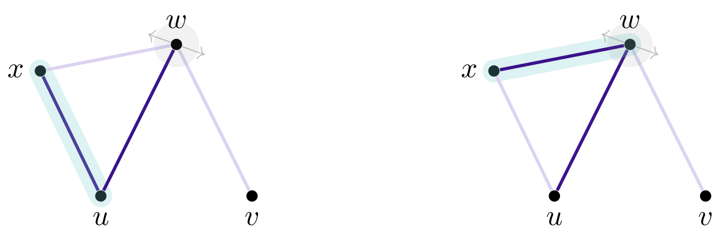
$$\mathbb{E}[Y_{uv}^-(i+1) \mid G_i] = \frac{(\varphi_y(t) \pm 2g_y(t))\sqrt{n}}{\frac{1}{2}n^2} \cdot (\varphi_y(t) \pm 2g_y(t))\sqrt{n} = 2(\varphi_y(t)^2 \pm 5g_y(t)\varphi_y(t))n^{-1}, \tag{4.2}$$

where the second factor corresponds to the number of configurations in $\mathcal{Y}_{uv}(i)$, and the second corresponds to the probability, for each, that we close (or choose) its one open edge. We also have

$$\mathbb{E}[Y_{uv}^+(i+1) \mid G_i] = \frac{2}{\frac{1}{2}n^2} \cdot (\varphi_x(t) \pm 2g_x(t))n = 4(\varphi_x(t) \pm 2g_x(t))n^{-1}, \tag{4.3}$$

where the second factor corresponds to the number of configurations in $\mathcal{X}_{uv}(i)$, and the first factor to the probability that we choose one of its two open edges.

To prove that the martingales $M_{uv}(i)$ and $N_{uv}(i)$ concentrate well, we claim that on the event $\mathcal{B}_i$, we always have $Y_{uv}^-(i+1) \le n^{\varepsilon}$. To see this, if e_{i+1} is incident to u, then letting $e_{i+1} = ux$, for any configuration $\{uw, vw\} \in \mathcal{Y}_{uv}(i)$ that it moves to $\mathcal{Y}_{uv}^-(i+1)$, we must have $\{vw, xw\} \in \mathcal{Z}_{vx}(i)$. And the number of such w is at most n^{ε} by Property 4.3 (for i). Meanwhile, if e_{i+1} isn't incident to u or v, it affects at most one configuration.



Also, we always have $Y_{uv}^+(i+1) \le 1$ (since e_{i+1} belongs to at most one configuration in $\mathcal{X}_{uv}(i)$).

So both $M_{uv}(i)$ and $N_{uv}(i)$ will concentrate well, e.g., within a window of length $n^{1/2-4\varepsilon}$, by Lemma 4.5 — we'll have

$$\frac{\sigma}{b} \ge \frac{n^{1/2-4\varepsilon}}{n^{\varepsilon}} \ge n^{1/2-5\varepsilon} \quad \text{and} \quad \frac{\sigma^2}{iab} \ge \frac{n^{1-8\varepsilon}}{n^{3/2+c} \cdot n^{-1} \cdot n^{\varepsilon}} \ge n^{1/2-10\varepsilon}$$

(where the n^{-1} bound on a comes from (4.2) and (4.3)).

Finally, we need to show that if $\mathcal{B}_{i-1}$ doesn't occur and $|M_{uv}(i)|, |N_{uv}(i)| \leq n^{1/2-4\varepsilon}$ (and $uv \notin G_i$), then $Y_{uv}(i) = (\varphi_y(t) \pm g_y(t))\sqrt{n}$. To do so, plugging (4.2) and (4.3) into (4.1) gives

$$\begin{aligned} Y_{uv}(i) &= -\sum_{j=0}^{i-1} \mathbb{E}[Y_{uv}^-(j+1) \mid G_j] + \sum_{j=0}^{i-1} \mathbb{E}[Y_{uv}^+(j+1) \mid G_j] - M_{uv}(i) + N_{uv}(i) \\ &= -\sum_{j=0}^{i-1} 2(\varphi_y(s)^2 \pm 5g_y(s)\varphi_y(s))n^{-1} + \sum_{j=0}^{i-1} 4(\varphi_x(s) \pm 2g_x(s))n^{-1} \pm 2n^{1/2-4\varepsilon}. \end{aligned}$$

Normalizing by $\sqrt{n}$ and converting the sums into integrals, we get

$$\begin{aligned} \frac{Y_{uv}(i)}{\sqrt{n}} &= -\sum_{j=0}^{i-1} (\varphi_y(s) \pm 5g_y(s)\varphi_y(s)) \cdot 2n^{-3/2} + \sum_{j=0}^{i-1} 2(\varphi_x(s) \pm 2g_x(s)) \cdot 2n^{-3/2} \pm 2n^{-4\varepsilon} \\ &= -\int_0^t (\varphi_y(s) \pm 5g_y(s)\varphi_y(s))\, ds + \int_0^t 2(\varphi_x(s) \pm 2g_x(s))\, ds \pm 3n^{-4\varepsilon}. \end{aligned}$$

The main terms match because (1.4) means that

$$\varphi_y(t) = -\int_0^t \varphi_y(s)^2\, ds + \int_0^t 2\varphi_x(s)\, ds,$$

and for the error terms, Fact 4.6 gives that

$$\int_0^t 5g_y(s)\varphi_y(s)\, ds + \int_0^t 4g_x(s)\, ds \leq \frac{1}{3} g_y(t)$$

(and $n^{-4\varepsilon} \ll g_y(t)$ as well). So this shows

$$Y_{uv}(i) = (\varphi_y(t) \pm g_y(t))\sqrt{n},$$

which concludes the portion of the argument corresponding to $\mathcal{Y}_{uv}(i)$.

§4.3.3 Property 4.3: Tracking $\mathcal{Z}_{uv}(i)$

Finally, we'll run the portion of the argument corresponding to $\mathcal{Z}_{uv}(i)$. For this, we *could* use martingale concentration again if we wanted to, but it's not necessary. As a simpler argument, for each uv, we have that $Z_{uv}(i)$ increases by at most 1 every step, and it increases by 1 only if we choose the one open edge of a configuration in $\mathcal{Y}_{uv}(i)$. By Property 4.2, as long as $\mathcal{B}_i$ doesn't occur, there are at most $\sqrt{n}$ such edges. This shows that off the event $\mathcal{B}_i$, conditioned on G_i we have

$$Z_{uv}(i+1) - Z_{uv}(i) \sim \mathsf{Ber}(p) \quad \text{for some } p \leq 2n^{-3/2}.$$

This means $Z_{uv}(i) \cdot \mathbf{1}_{\mathcal{B}_{i-1}}$ is stochastically dominated by a sum $S_{uv}(i)$ of i independent $\mathsf{Ber}(2n^{-3/2})$ random variables. And we're running up to $i \leq n^{3/2+c}$, so $\mathbb{E}[S_{uv}(i)] \leq 2n^c$, and since c is small relative to ε, we get that $S_{uv}(i) \leq n^{\varepsilon}$ with very high probability (and therefore the same is true of $Z_{uv}(i) \cdot \mathbf{1}_{\mathcal{B}_{i-1}}$).

References

[AKS81] Miklós Ajtai, János Komlós, and Endre Szemerédi. A dense infinite Sidon sequence. *European Journal of Combinatorics*, 2:1–11, 1981.

[BK20] Tom Bohman and Peter Keevash. Dynamic concentration of the triangle-free process. *Random Structures & Algorithms*, 58:221–293, 2020.

[Boh09] Tom Bohman. The triangle-free process. *Advances in Mathematics*, 221(5):1653–1677, 2009. arxiv:0806.4375.

[Kim95] Jeong Han Kim. The Ramsey number $r(3, t)$ has order of magnitude $t^2/\log t$. *Random Structures & Algorithms*, 7(3):173–207, 1995.

[PGM20] Gonzalo Fiz Pontiveros, Simon Griffiths, and Robert Morris. The triangle-free process and the Ramsey number $r(3, k)$. *Memoirs of the American Mathematical Society*, 263(1274), 2020.

[Spe95] Joel Spencer. Maximal triangle-free graphs and Ramsey $r(3, k)$, 1995. Unpublished manuscript, https://cs.nyu.edu/~spencer/papers/potpourri.pdf.

[Wol11] Guy Wolfovitz. Triangle-free subgraphs in the triangle-free process. *Random Structures & Algorithms*, 39(4):539–543, 2011. arxiv:0903.1756.