arXiv:0903.1756v1 [math.CO] 10 Mar 2009

4-cycles at the triangle-free process

Guy Wolfovitz*

Abstract

We consider the triangle-free process: Given an integer n, start by taking a uniformly random permutation of the edges of the complete n-vertex graph K_n. Then, traverse the edges of K_n according to the order imposed by the permutation and add each traversed edge to an (initially empty) evolving graph - unless its addition creates a triangle. We study the evolving graph at around the time where $\Theta(n^{3/2+\varepsilon})$ edges of K_n have been traversed for any fixed $\varepsilon \in (0, 10^{-6})$. At that time, we give a tight concentration result for the number of copies of the 4-cycle in the evolving graph. Our analysis uses in part Spencer's original branching process approach for analysing the triangle-free process, coupled with the semi-random method.

1 Introduction

Consider the next random process for generating a triangle-free graph. Given $n \in \mathbb{N}$, assign every edge f of the complete n-vertex graph K_n a birthtime $\beta(f)$, distributed uniformly at random in the interval $[0, 1]$. Now start with the empty n-vertex graph and iteratively add edges to it as follows. Traverse the edges of K_n in order of their birthtimes (which are all distinct with probability 1), starting with the edge whose birthtime is smallest, and add each traversed edge to the evolving graph, unless its addition creates a triangle. When all edges of K_n have been exhausted, the process ends. Denote by $\mathbb{TF}(n)$ the triangle-free graph which is the result of the above process. Further, denote by $\mathbb{TF}(n, p)$ the intersection of $\mathbb{TF}(n)$ with $\{f : \beta(f) \le p\}$.

Let X be the random variable that counts the number of edges in $\mathbb{TF}(n, p)$. Let X' be the random variable that counts the number of copies of the 4-cycle, C_4, in $\mathbb{TF}(n, p)$. We say that an event holds *asymptotically almost surely* (*a.a.s.*) if the probability of the event goes to 1 as $n \to \infty$. For $x = x(n)$, $y = y(n)$, we write $x \sim y$ if x/y goes to 1 as $n \to \infty$. Let $\ln x$ denote the natural logarithm of x. Our main result follows.

Theorem 1.1. *Let $\varepsilon \in (0, 10^{-6})$. For some $p \sim n^{\varepsilon - 1/2}$, a.a.s.,*

$$X \sim \binom{n}{2} \frac{\sqrt{\ln n^{\varepsilon}}}{\sqrt{n}}, \quad X' \sim \frac{n^4}{aut(C_4)} \left(\frac{\sqrt{\ln n^{\varepsilon}}}{\sqrt{n}} \right)^4.$$

One interesting point worth making with respect to our main result is this. Fix $\varepsilon \in (0, 10^{-6})$ and let $p \sim n^{\varepsilon - 1/2}$ be as guaranteed to exist by Theorem 1.1. Consider the random graph $\mathbb{G}(n, m)$, which is chosen uniformly at random from among those n-vertex graphs with exactly

*Department of Computer Science, Haifa University, Haifa, Israel. Email address: `gwolfovi@cs.haifa.ac.il`.

$m := \lfloor 2^{-1}n^{3/2}\sqrt{\ln n^{\varepsilon}}\rfloor$ edges. Note that by Theorem 1.1, $\mathbb{TF}(n,p)$ and $\mathbb{G}(n,m)$ a.a.s. has asymptotically the same number of edges. This of course follows directly from our choice of the parameter m. The point is that by standard techniques and by Theorem 1.1, we also have that a.a.s., the number of copies of the 4-cycle in $\mathbb{G}(n,m)$ is asympyoyically equal to the number of copies of the 4-cycle in $\mathbb{TF}(n,p)$. Furthermore, $\mathbb{G}(n,m)$ is expected to contain many triangles, and indeed it does contain many triangles a.a.s., whereas $\mathbb{TF}(n,m)$ contains no triangles at all. Therefore, one may argue, at least with respect to the number of 4-cycles, that the graph $\mathbb{TF}(n,p)$ "looks like" a uniformly random graph with m edges–only that it has no triangles.

Related results. Erdős, Suen and Winkler [6] were the first to consider the triangle-free process. They proved that a.a.s., the number of edges in $\mathbb{TF}(n)$ is bounded by $\Omega(n^{3/2})$ and $O(n^{3/2}\ln n)$. Spencer [12] showed that for every two reals a_1, a_2, there exists n_0 such that the expected number of edges in $\mathbb{TF}(n)$ for $n \geq n_0$ is bounded from below by $a_1 n^{3/2}$ and from above by $a_2 n^{3/2}\ln n$. In the same paper, Spencer conjectured that a.a.s., the number of edges in $\mathbb{TF}(n)$ is $\Theta(n^{3/2}\sqrt{\ln n})$. This conjecture was recently proved valid by Bohman [1]. Other results are known for the more general H-free process. In the H-free process, instead of forbidding a triangle, one forbids the appearance of a copy of H. Bollobás and Riordan [2] considered the H-free process for the case where $H \in \{K_4, C_4\}$ and Osthus and Taraz [9] considered the more general case where H is strictly 2-balanced. In both cases, the authors gave upper and lower bounds on the probable number of edges that appear in the resulting H-free graph, bounds that are tight up to poly$(\ln n)$ factors. Bohman [1], in addition to the triangle-free process, considered the H-free process for $H = K_4$ and gave a lower bound on the number of edges in the resulting H-free graph, a bound that improves that given by [2, 9].

2 Preliminaries

2.1 Notation

As usual, we write $[d]$ for the set $\{1, 2, \ldots, d\}$. All asymptotic notation in this paper is with resepct to $n \to \infty$. We write $x = y \pm z$ if $x \in [y-z, y+z]$. We also use $y \pm z$ to denote the interval $[y-z, y+z]$. All inequalities in this paper are valid only for $n > n_0$, for some sufficiently large n_0 which we do not specify.

2.2 Basic definitions

Here we give some definitions and set up some useful parameters which will be used throughout the paper. We start by fixing $\varepsilon \in (0, 10^{-6})$. For brevity, we assume that n^{ε} is an integer. (Our argument can be easily altered so that it holds also for the case where n^{ε} is not an integer.) Define the following functions of n:

$$\delta := n^{-\varepsilon}, \;\; I := n^{2\varepsilon}, \;\; k := n^{50\varepsilon}, \;\; \text{and} \;\;\; K := n^{5000\varepsilon}.$$

Let $\Phi(x)$ be a function, whose derivative is denoted by $\phi(x)$, and which is defined by

$$\Phi(0) = 0, \quad \phi(x) = \exp(-\Phi(x)^2).$$

Furthermore, let erfi(x) be the imaginary error function, given by

$$\operatorname{erfi}(x) = \frac{2}{\sqrt{\pi}} \sum_{j=0}^{\infty} \frac{x^{2j+1}}{j!(2j+1)}.$$

Solving the separable differential equation above, one gets that $\Phi(x)$ satisfies

$$\frac{\sqrt{\pi}}{2} \operatorname{erfi}(\Phi(x)) = x. \tag{1}$$

We have that $\operatorname{erfi}(x) \to \exp(x^2)/(\sqrt{\pi}x)$ as $x \to \infty$. Hence, by the above, we have that as $x \to \infty$,

$$\Phi(x) \to \sqrt{\ln x}, \quad \phi(x) \to \frac{1}{2x\sqrt{\ln x}}, \quad \text{and} \quad \operatorname{erfi}(\Phi(x)) \to \frac{1}{\sqrt{\pi}\Phi(x)\phi(x)}.$$

For every $0 \le i \le I$, define

$$\begin{aligned} \gamma(i) &:= \max\{\delta\Phi(i\delta)\phi(i\delta),\, \delta^2\phi(i\delta)^2\}, \\ \Gamma(i) &:= \begin{cases} n^{-30\varepsilon} & \text{if } i = 0, \\ \Gamma(i-1)\cdot(1+10\gamma(i-1)) & \text{if } i \ge 1. \end{cases} \end{aligned}$$

Fact 2.1. *For all* $0 \le i \le I$,

(i) $\phi(i\delta) \le 1$ *and* $\Phi(i\delta) \le \ln n$.

(ii) $\phi(i\delta) = \Omega(\delta^{1.5})$ *and* $i \ge 1 \implies \Phi(i\delta) = \Omega(\delta)$.

(iii) $\gamma(i) = o(1)$.

(iv) $\gamma(i) = \Omega(\delta^5)$.

(v) $n^{-30\varepsilon} \le \Gamma(i) \le n^{-10\varepsilon}$.

Proof.

(i) We have that $\operatorname{erfi}(x) \ge 0$ if and only if $x \ge 0$. By (1), $\operatorname{erfi}(\Phi(i\delta)) = 2i\delta/\sqrt{\pi} \ge 0$. Hence $\Phi(i\delta) \ge 0$. Therefore $\phi(i\delta) = \exp(-\Phi(i\delta)^2) \le 1$. Next, note that erfi$(x)$ is monotonically increasing with x. We also have by (1) that $\operatorname{erfi}(\Phi(i\delta))$ is monotonically increasing with i. Hence $\Phi(i\delta)$ is monotonically increasing with i and so $\Phi(i\delta) \le \Phi(I\delta)$. The bound on $\Phi(i\delta)$ now follows since $I\delta = n^{\varepsilon}$ and $\Phi(n^{\varepsilon}) \sim \sqrt{\ln n^{\varepsilon}}$.

(ii) As argued in the proof of (i), we have that $\Phi(i\delta)$ is monotonically increasing with i. This in turn implies that $\phi(i\delta)$ is monotonically decreasing with i. Therefore, it is enough to show that $\Phi(\delta) = \Omega(\delta)$ and $\phi(I\delta) = \Omega(\delta^{1.5})$. The fact that $\Phi(\delta) = \Omega(\delta)$ follows directly from (1) and the definition of erfi(x). The fact that $\phi(I\delta) = \Omega(\delta^{1.5})$ follows since $\phi(I\delta) \to 1/(2I\delta\sqrt{\ln I\delta})$.

(iii) By (i) we have $\delta\Phi(i\delta)\phi(i\delta) \le \delta \ln n = o(1)$ and $\delta^2\phi(i\delta)^2 \le \delta^2 = o(1)$.

(iv) Follows directly from the definition of $\gamma(i)$ and (ii).

(v) Since $\Gamma(i)$ is monotonically non-decreasing and $\Gamma(0) = n^{-30\varepsilon}$, it is enough to show that $\Gamma(I) \leq n^{-10\varepsilon}$. We do that by first showing that $\Gamma(\delta^{-1}\lfloor \ln\ln n \rfloor) \leq n^{-30\varepsilon+o(1)}$. For brevity, we shall assume below that $\ln\ln n$ is an integer.

For every $0 \leq i \leq \delta^{-1}\ln\ln n$, $\Phi(i\delta) \leq \ln\ln n$ (crudely) and $\phi(i\delta) \leq 1$. Therefore, we have that for every $0 \leq i \leq \delta^{-1}\ln\ln n$,

$$\begin{aligned} \delta\Phi(i\delta)\phi(i\delta) &\leq \delta\ln\ln n, \text{ and} \\ \delta^2\phi(i\delta)^2 &\leq \delta\ln\ln n. \end{aligned}$$

Hence, for $0 \leq i \leq \delta^{-1}\ln\ln n$, $\gamma(i) \leq \delta\ln\ln n$ and so

$$\Gamma(\delta^{-1}\ln\ln n) \leq n^{-30\varepsilon}(1+10\delta\ln\ln n)^{\delta^{-1}\ln\ln n} = n^{-30\varepsilon+o(1)}.$$

We now note that for every $\delta^{-1}\ln\ln n \leq i \leq I$,

$$\begin{aligned} \delta\Phi(i\delta)\phi(i\delta) &\leq 0.6/i, \text{ and} \\ \delta^2\phi(i\delta)^2 &\leq 0.6/i, \end{aligned}$$

and this follows from the fact that for $\delta^{-1}\ln\ln n \leq i \leq I$, $\Phi(i\delta)\phi(i\delta) \sim 1/(2i\delta)$ and $\phi(i\delta) \leq 1/(2i\delta)$. Hence, for $\delta^{-1}\ln\ln n \leq i \leq I$, $\gamma(i) \leq 0.6/i$ and so we conclude that

$$\Gamma(I) \leq n^{-30\varepsilon+o(1)} \prod_{1\leq i\leq I}(1+6/i) \leq n^{-30\varepsilon+o(1)} \cdot \exp(7\ln I) \leq n^{-10\varepsilon}.$$

■

3 Argument for Theorem 1.1 and statement of Main Lemma

In this section we give our argument for the proof of Theorem 1.1. We begin the argument by giving an alternative definition of the triangle-free process. Under this alternative definition, we formulate an equivalent assertion to the one given in Theorem 1.1. We then give the overall argument for the proof of this new, equivalent assertion, including stating our main technical lemma. The actual proof is deferred to subsequent sections.

For every integer $i \geq 0$ we define a triangle-free graph $\mathbb{TF}(n,i)$. Initially, take $\mathbb{TF}(n,0)$ to be the empty graph over the vertex set of K_n and set $\mathbb{B}_0 := \emptyset$. Given $\mathbb{TF}(n,i)$, $i \geq 0$, define $\mathbb{TF}(n,i+1)$ as follows. Choose uniformly at random a function $\beta_{i+1} : K_n \setminus \mathbb{B}_{\leq i} \to [0,1]$ where $\mathbb{B}_{\leq i} := \bigcup_{j\leq i}\mathbb{B}_j$. Let $\mathbb{B}_{i+1}$ be the set of edges f for which the birthtime $\beta_{i+1}(f)$ satisfies $\beta_{i+1}(f) < \delta n^{-1/2}$. Traverse the edges in $\mathbb{B}_{i+1}$ in order of their birthtimes (starting with the edge whose birthtime is smallest), and add each traversed edge to $\mathbb{TF}(n,i)$, unless its addition creates a triangle. Denote by $\mathbb{TF}(n,i+1)$ the graph thus produced. Observe that $\mathbb{TF}(n,I)$ has the same distribution as $\mathbb{TF}(n,p)$ for some $p \sim n^{\varepsilon-1/2}$. With that observation, the following implies Theorem 1.1.

Theorem 3.1. *Let Y be the random variable that counts the number of edges in $\mathbb{TF}(n,I)$. Let Y' be the random variable that counts the number of copies of C_4 in $\mathbb{TF}(n,I)$. Then a.a.s.,*

$$Y \sim \binom{n}{2}\frac{\Phi(I\delta)}{\sqrt{n}}, \quad Y' \sim \frac{n^4}{aut(C_4)}\left(\frac{\Phi(I\delta)}{\sqrt{n}}\right)^4.$$

The validity of Theorem 3.1, as we show below, follows from the following theorem together with the second moment method.

Theorem 3.2. *For every triangle-free graph $F \subset K_n$ of size $O(1)$,*

$$Pr[F \subseteq \mathbb{TF}(n,I) \,|\, F \subseteq \mathbb{B}_{\le I}] \sim \left(\frac{\Phi(I\delta)}{n^{\varepsilon}}\right)^{e_F}.$$

Proof of Theorem 3.1. The number of copies of C_4 in K_n is $4!\binom{n}{4}/\mathrm{aut}(C_4)$. In addition, for $F \subset K_n$,

$$\Pr[F \subseteq \mathbb{B}_{\le I}] \sim \left(\frac{n^{\varepsilon}}{\sqrt{n}}\right)^{e_F}.$$

Therefore, by Theorem 3.2,

$$\mathbb{E}[Y] \sim \binom{n}{2}\frac{\Phi(I\delta)}{\sqrt{n}}, \qquad \mathbb{E}[Y'] \sim \frac{n^4}{\mathrm{aut}(C_4)}\left(\frac{\Phi(I\delta)}{\sqrt{n}}\right)^4.$$

To complete the proof, it suffices by Chebyshev's inequality (see e.g. [?AS92]) to show that $\mathrm{Var}(Y') = o(\mathbb{E}[Y']^2)$ and $\mathrm{Var}(Y) = o(\mathbb{E}[Y]^2)$.

We bound $\mathrm{Var}(Y')$. For $F \subset K_n$, let I_F be the indicator random variable for the event $\{F \subseteq \mathbb{TF}(n,I)\}$. We have

$$\mathrm{Var}(Y') = \sum_{F,F'} \mathrm{Cov}(I_F, I_{F'}) = \sum_{F,F'} \mathbb{E}[I_F, I_{F'}] - \mathbb{E}[I_F]\,\mathbb{E}[I_{F'}],$$

where the sum ranges over all copies F, F' of C_4 in K_n. We partition the sum above to two sums and show that each is bounded by $o(\mathbb{E}[Y']^2)$. First, let $\sum_{F,F'}$ be the sum over all copies F, F' of C_4 in K_n such that F and F' share no vertex. If F and F' share no vertex then $F \cup F'$ is triangle-free. Hence, by Theorem 3.2,

$$\sum_{F,F'} \mathbb{E}[I_F, I_{F'}] - \mathbb{E}[I_F]\,\mathbb{E}[I_{F'}] = \left(\frac{n^4}{\mathrm{aut}(C_4)}\right)^2 \cdot o\left(\left(\frac{\Phi(I\delta)}{\sqrt{n}}\right)^8\right),$$

which is $o(\mathbb{E}[Y']^2)$. We will now make use of the following observation: If F, F' are two copies of C_4 in K_n with $F \cap F'$ being isomorphic to H, then $\mathbb{E}[I_F, I_{F'}] = O((n^{\varepsilon}n^{-1/2})^{8-e_H})$. This is true since, if the event $\{F, F' \subseteq \mathbb{TF}(n,I)\}$ occurs, we must also have that $\{F \cup F' \subseteq \mathbb{B}_{\le I}\}$ occurs. Let $\sum_H$ be the sum over all $H \subseteq C_4$ with $v_H > 0$. Let $\sum_{F\cap F' \equiv H}$ be the sum over all copies F, F' of C_4 in K_n that share at least 1 vertex such that $F \cap F'$ is isomorphic to H. Then by the observation above,

$$\sum_H \sum_{F\cap F'\equiv H} \mathrm{Cov}(I_F, I_{F'}) \le O(n^{8-v_H}) \cdot (n^{\varepsilon}n^{-1/2})^{8-e_H},$$

which is $o(\mathbb{E}[Y']^2)$, since $n^{-v_H+e_H/2}(n^{\varepsilon})^{8-e_H} = o(1)$ for every $H \subseteq C_4$ with $v_H > 0$. This implies the desired bound on $\mathrm{Var}(Y')$. A similar argument also shows that $\mathrm{Var}(Y) = o(\mathbb{E}[Y]^2)$. ∎

It remains to prove Theorem 3.2. The main lemma used in the proof is stated in the following subsection. The rest of the paper will then be devoted for the proof of this main lemma (Sections 4 , 5 and 6) and for the proof of Theorem 3.2 (Section 7).

3.1 Statement of Main lemma

In this subsection we state our main technical lemma, whose proof is given in the next three sections. This lemma and its proof will be used later to prove Theorem 3.2.

For every edge $g \in K_n$ and for every $0 \le i \le I$ we define $\Lambda_j(g,i)$, $j \in \{0,1,2\}$ as follows. Let $\Lambda_0(g,i)$ be the family of all sets $\{g_1,g_2\} \subseteq \mathbb{TF}(n,i)$ such that $\{g,g_1,g_2\}$ is a triangle. Let $\Lambda_1(g,i)$ be the family of all singletons $\{g_1\} \subseteq K_n$ such that there exists $g_2 \in \mathbb{TF}(n,i)$ for which $\{g,g_1,g_2\}$ is a triangle and it holds that (i) $g_1 \notin \mathbb{B}_{\le i}$ and (ii) $\mathbb{TF}(n,i) \cup \{g_1\}$ is triangle-free. Let $\Lambda_2(g,i)$ be the family of all sets $\{g_1,g_2\} \subseteq K_n$ such that $\{g,g_1,g_2\}$ is a triangle and for which it holds that (i) $g_1,g_2 \notin \mathbb{B}_{\le i}$ and (ii) $\mathbb{TF}(n,i) \cup \{g_j\}$ is triangle-free for both $j \in \{1,2\}$.

Lemma 3.3 (Main Lemma)**.** *Let $f \in K_n$, $0 \le i < I$. Suppose that given $\mathbb{TF}(n,i)$, we have for all $g \in K_n$,*

$$\begin{aligned}
|\Lambda_0(g,i)| &\le in^{1/100},\\
|\Lambda_1(g,i)| &= 2\sqrt{n}\Phi(i\delta)\phi(i\delta)\cdot(1\pm\Gamma(i)),\\
|\Lambda_2(g,i)| &= n\phi(i\delta)^2\cdot(1\pm\Gamma(i)).
\end{aligned}$$

Then:

- *Conditioned on the event $\{f \in \mathbb{B}_{i+1}, \mathbb{TF}(n,i)\cup\{f\}$ is triangle-free$\}$,*

$$Pr[f \in \mathbb{TF}(n,i+1)] = \begin{cases} \frac{\Phi((i+1)\delta)}{\delta}(1\pm 9\Gamma(i)\gamma(i)) & \text{if } i=0,\\ \frac{\Phi((i+1)\delta)-\Phi(i\delta)}{\phi(i\delta)\,\delta}(1\pm 9\Gamma(i)\gamma(i)) & \text{if } i\ge 1.\end{cases}$$

- *Conditioned on the event $\{f \notin \mathbb{B}_{\le i+1}, \mathbb{TF}(n,i)\cup\{f\}$ is triangle-free$\}$,*

$$Pr[\mathbb{TF}(n,i+1)\cup\{f\} \text{ is triangle-free}] = \begin{cases} \phi((i+1)\delta)(1\pm 9\Gamma(i)\gamma(i)) & \text{if } i=0,\\ \frac{\phi((i+1)\delta)}{\phi(i\delta)}(1\pm 9\Gamma(i)\gamma(i)) & \text{if } i\ge 1.\end{cases}$$

- *With probability at least $1-n^{-\omega(1)}$, for all $g \in K_n$,*

$$\begin{aligned}
|\Lambda_0(g,i+1)| &\le (i+1)n^{1/100},\\
|\Lambda_1(g,i+1)| &= 2\sqrt{n}\Phi((i+1)\delta)\phi((i+1)\delta)\cdot(1\pm\Gamma(i+1)),\\
|\Lambda_2(g,i+1)| &= n\phi((i+1)\delta)^2\cdot(1\pm\Gamma(i+1)).
\end{aligned}$$

4 Proof of Main Lemma

Fix f,i as specified in the Main Lemma and assume that the preconditions in the lemma hold.

4.1 Definitions and an observation

Definition 1 (Redefinition of β_{i+1})**.** *Let $\mathbb{B}^{**}_{i+1}$ be a random set of edges, formed by choosing every edge in $K_n \setminus \mathbb{B}_{\le i}$ with probability $Kn^{-1/2}$. Let $\mathbb{B}^{*}_{i+1}$ be a random set of edges, formed by choosing* *every edge in $\mathbb{B}_{i+1}^{**}$ with probability kK^{-1}. For each $g \in \mathbb{B}_{i+1}^{*}$, let $\beta_{i+1}(g)$ be distributed uniformly at random in $[0, kn^{-1/2}]$ and for each $g \in K_n \setminus (\mathbb{B}_{\leq i} \cup \mathbb{B}_{i+1}^{*})$, let $\beta_{i+1}(g)$ be distributed uniformly at random in $(kn^{-1/2}, 1]$.*

Clearly, the above new definition of β_{i+1} is equivalent to the original definition of β_{i+1} that was given in Section 3. Note that $\mathbb{B}_{i+1} \subseteq \mathbb{B}_{i+1}^{*} \subseteq \mathbb{B}_{i+1}^{**}$.

For $g \in K_n$ and $j \in \{1, 2\}$, let $\Lambda_j^{*}(g, i)$ be the family of all $G \in \Lambda_j(g, i)$ such that $G \subseteq \mathbb{B}_{i+1}^{*}$ and let $\Lambda_j^{**}(g, i)$ be the family of all $G \in \Lambda_j(g, i)$ such that $G \subseteq \mathbb{B}_{i+1}^{**}$.

Definition 2. *Let $g \in K_n$, $d \in \mathbb{N}$. We define inductively a labeled, rooted tree $T_{g,d}^{*}$ of height $2d$. The nodes at even distance from the root will be labeled with edges. The nodes at odd distance from the root will be labeled with sets of j edges, $j \in \{1, 2\}$.*

- *$T_{g,1}^{*}$:*
  - *The root v_0 of $T_{g,1}^{*}$ is labeled with the edge g.*
  - *For every set $G \in \Lambda_1^{*}(g, i) \cup \Lambda_2^{*}(g, i)$ do: Set a new node u_1, labeled G, as a child of v_0. Furthermore, for each edge $g_1 \in G$ set a new node v_1, labeled g_1, as a child of u_1.*
- *$T_{g,d}^{*}$, $d \geq 2$: We construct the tree $T_{g,d}^{*}$ by adding new nodes to $T_{g,d-1}^{*}$ as follows. Let $(v_0, u_1, v_1, \ldots, u_{d-1}, v_{d-1})$ be a directed path in $T_{g,d-1}^{*}$ from the root v_0 to a leaf v_{d-1}. Let g_j be the label of v_j.*
  - *For every set $G \in \Lambda_2^{*}(g_{d-1}, i)$ for which $g_{d-2} \notin G$ do: Set a new node u_d, labeled G, as a child of v_{d-1}. Furthermore, for each edge $g_d \in G$ set a new node v_d, labeled g_d, as a child of u_d.*
  - *For every set $G \in \Lambda_1^{*}(g_{d-1}, i)$ for which $g_{d-2} \notin G$ and $G \cup \{g_{d-1}, g_{d-2}\}$ isn't a triangle do: Set a new node u_d, labeled G, as a child of v_{d-1}. Furthermore, for the edge $g_d \in G$ set a new node v_d, labeled g_d, as a child of u_d.*

Define the tree $T_{g,d}^{**}$ exactly as $T_{g,d}^{*}$ above, only that now use $\Lambda_j^{**}(\cdot, i)$ instead of $\Lambda_j^{*}(\cdot, i)$, $j \in \{1, 2\}$, in the definition. Lastly, for $G \subset K_n$, define $T_{G,d}^{*} := \bigcup_{g \in G} T_{g,d}^{*}$.

For the next definition, let $g \in K_n$ and assume that $g \notin \mathbb{B}_{\leq i}$ so that $\beta_{i+1}(g)$ is defined. Let $T \in \{T_{g,d}^{*}, T_{g,d}^{**}\}$ and let v be a node at even distance from the root of T. Let g' be the label of v. Define the event that v *survives* as follows. If v is a leaf then v survives by definition. Otherwise, v survives if and only if for every child u, labeled G, of v, the following holds: If $\beta_{i+1}(g'') < \beta_{i+1}(g')$ for all $g'' \in G$ then u has a child that does not survive. Let $\mathcal{A}_{g,d}^{*}$ ($\mathcal{A}_{g,d}^{**}$) be the event that the root of $T_{g,d}^{*}$ ($T_{g,d}^{**}$) survives. The following is an easy observation.

Proposition 4.1. *Let $g \in K_n$, $d \geq 1$ odd.*

- *Conditioned on $\{g \in \mathbb{B}_{i+1}, \mathbb{TF}(n, i) \cup \{g\}$ is triangle-free$\}$,*

$$\mathcal{A}_{g,d}^{*} \implies \{g \in \mathbb{TF}(n, i+1)\} \implies \mathcal{A}_{g,d+1}^{*}.$$

- *Conditioned on* $\{g \notin \mathbb{B}_{\leq i+1}, \mathbb{TF}(n,i) \cup \{g\}$ *is triangle-free*$\}$,

$$\mathcal{A}^*_{g,d} \implies \{\mathbb{TF}(n,i+1) \cup \{g\} \text{ is triangle-free}\} \implies \mathcal{A}^*_{g,d+1}.$$

In addition, the above also holds if one replaces $\mathcal{A}^*_{g,d}$ *with* $\mathcal{A}^{**}_{g,d}$ *and* $\mathcal{A}^*_{g,d+1}$ *with* $\mathcal{A}^{**}_{g,d+1}$.

4.2 Proof of Main Lemma

Let $\mathcal{E}^{**}$ be the event that the following two items hold:

- For every $g \in K_n$,

$$\begin{aligned} |\Lambda_1^{**}(g,i)| &= 2K\Phi(i\delta)\phi(i\delta) \cdot (1 \pm 1.1\Gamma(i)), \\ |\Lambda_2^{**}(g,i)| &= K^2\phi(i\delta)^2 \cdot (1 \pm 1.1\Gamma(i)). \end{aligned}$$

- For every three vertices $w, x, y \in K_n$:
  - The number of vertices z such that $\{w,z\},\{y,z\} \in \mathbb{TF}(n,i)$ and $\{x,z\} \in \mathbb{B}^{**}_{i+1}$ is at most $(\ln n)^2$.
  - The number of vertices z such that $\{w,z\} \in \mathbb{TF}(n,i)$ and $\{x,z\},\{y,z\} \in \mathbb{B}^{**}_{i+1}$ is at most $(\ln n)^2$.

Let $\mathcal{E}^*$ be the event that for every $g \in K_n$,

$$\begin{aligned} |\Lambda_1^{*}(g,i)| &= 2k\Phi(i\delta)\phi(i\delta) \cdot (1 \pm 1.2\Gamma(i)), \\ |\Lambda_2^{*}(g,i)| &= k^2\phi(i\delta)^2 \cdot (1 \pm 1.2\Gamma(i)). \end{aligned}$$

Fix once and for the rest of the paper $D \in \{40, 41\}$. For $F \subset K_n$, let $\mathcal{E}^*_F$ be the following event: If g is a label of some node at even distance from the root of $T^*_{f,D} \in T^*_{F,D}$ then g is the label of no other node at even distance from the root of a tree in $T^*_{F,D}$.

The next two lemmas are proved in Sections 5 and 6 respectivly.

Lemma 4.2. *Let* $F \subset K_n$ *be a triangle-free graph of size* $O(1)$. *Then*

- $Pr[\mathcal{E}^{**}] \geq 1 - n^{-\omega(1)}$.
- $Pr[\mathcal{E}^* \mid \mathcal{E}^{**}] \geq 1 - n^{-\omega(1)}$.
- $Pr[\mathcal{E}^*_F \mid \mathbb{TF}(n,i) \cup F$ *is triangle-free*, $\mathcal{E}^{**}] \geq 1 - K^{-1/10}$.

Lemma 4.3. *Condition on the event*

$$\{f \notin \mathbb{B}_{\leq i},\ \mathbb{TF}(n,i) \cup \{f\} \text{ is triangle-free},\ \mathcal{E}^*,\ \mathcal{E}^*_{\{f\}}\}.$$

Then:

- *Conditioned on* $\{f \in \mathbb{B}_{i+1}\}$,

$$Pr[\mathcal{A}^*_{f,D}] = \begin{cases} \frac{\Phi((i+1)\delta)}{\delta}(1 \pm 8.5\Gamma(i)\gamma(i)) & \text{if } i = 0, \\ \frac{\Phi((i+1)\delta)-\Phi(i\delta)}{\phi(i\delta)\,\delta}(1 \pm 8.5\Gamma(i)\delta\Phi(i\delta)\phi(i\delta)) & \text{if } i \geq 1. \end{cases}$$

- *Conditioned on* $\{f \notin \mathbb{B}_{\leq i+1}\}$,

$$Pr[\mathcal{A}^*_{f,D}] = \begin{cases} \phi((i+1)\delta)(1 \pm 8.5\Gamma(i)\gamma(i)) & \text{if } i = 0, \\ \frac{\phi((i+1)\delta)}{\phi(i\delta)}(1 \pm 8.5\Gamma(i)\gamma(i)) & \text{if } i \geq 1. \end{cases}$$

Fact 2.1, Proposition 4.1, the choice of D (it being either even or odd) and Lemmas 4.2 and 4.3 imply immediately the validity of the first two items in the Main Lemma. It remains to prove the validity of the last item in the Main Lemma.

In what follows we define three random variables that will be used to estimate the cardinality of $\Lambda_j(g, i+1)$, $j \in \{0,1,2\}$. Let $\lambda_0(g, i+1, d)$ be the number of sets $\{g_1\} \in \Lambda_1(g,i)$ for which it holds that $g_1 \in \mathbb{B}_{i+1}$ and $A^*_{g_1,d}$ holds, *plus* the number of sets $\{g_1, g_2\} \in \Lambda_2(g,i)$ for which it holds that $g_1, g_2 \in \mathbb{B}_{i+1}$ and $A^*_{g_1,d} \cap A^*_{g_2,d}$ occurs. Let $\lambda_1(g, i+1, d)$ be the number of sets $\{g_1\} \in \Lambda_1(g,i)$ for which it holds that $g_1 \notin \mathbb{B}_{\leq i+1}$ and $\mathcal{A}^*_{g_1,d}$ occurs, *plus* the number of sets $\{g_1, g_2\} \in \Lambda_2(g,i)$ for which it holds that $g_1 \in \mathbb{B}_{i+1}$, $g_2 \notin \mathbb{B}_{\leq i+1}$, and $\mathcal{A}^*_{g_1,d} \cap \mathcal{A}^*_{g_2,d}$ occurs. Let $\lambda_2(g, i+1, d)$ be the number of sets $\{g_1, g_2\} \in \Lambda_2(g,i)$ for which it holds that $g_1, g_2 \notin \mathbb{B}_{\leq i+1}$ and $\mathcal{A}^*_{g_1,d} \cap \mathcal{A}^*_{g_2,d}$ occurs.

By the preconditions in the lemma and Proposition 4.1 we have for every $g \in K_n$ and odd $d \geq 1$,

$$\begin{aligned} & & |\Lambda_0(g, i+1)| &\leq \lambda_0(g, i+1, d+1) + in^{1/100}, \\ \lambda_1(g, i+1, d) &\leq & |\Lambda_1(g, i+1)| &\leq \lambda_1(g, i+1, d+1), \\ \lambda_2(g, i+1, d) &\leq & |\Lambda_2(g, i+1)| &\leq \lambda_2(g, i+1, d+1). \end{aligned}$$

By the preconditions in the lemma, the fact that $\Pr[g \notin \mathbb{B}_{\leq i+1} \,|\, g \notin \mathbb{B}_{\leq i}] \geq 1 - \delta n^{-1/2}$ for any $g \in K_n \setminus \mathbb{B}_{\leq i}$ and by Lemmas 4.2 and 4.3 one can verify that

$$\begin{aligned} \mathbb{E}[\lambda_0(g, i+1, D) \,|\, \mathcal{E}^{**}] &\leq n^{1/100}/2, \\ \mathbb{E}[\lambda_1(g, i+1, D) \,|\, \mathcal{E}^{**}] &= 2\sqrt{n} \cdot \Phi((i+1)\delta) \cdot \phi((i+1)\delta) \cdot (1 \pm (\Gamma(i) + 8.6\Gamma(i)\gamma(i))), \\ \mathbb{E}[\lambda_2(g, i+1, D) \,|\, \mathcal{E}^{**}] &= n \cdot \phi((i+1)\delta)^2 \cdot (1 \pm (\Gamma(i) + 8.6\Gamma(i)\gamma(i))). \end{aligned}$$

Now note that conditioned on $\mathcal{E}^{**}$, for every $g \in K_n$, the event $\mathcal{A}^{**}_{g,D}$ depends only on the birthtimes of at most $O(n^{1/10})$. Furthermore, every edge appears as a label in at most $O(n^{1/10})$ trees $T^{**}_{g,D}$. Therfore, $\mathcal{A}^*_{g,D}$ depends only on the birthtimes of at most $O(n^{1/10})$ edges for every $g \in K_n$ and every edge appears as a label in at most $O(n^{1/10})$ trees $T^*_{g,D}$. Hence, one can use Talagrand's inequality to get that conditioned on $\mathcal{E}^{**}$, the following hold with probability $1 - n^{-\omega(1)}$:

$$\begin{aligned} \lambda_0(g, i+1, D) &\leq n^{1/100}, \\ \lambda_1(g, i+1, D) &= 2\sqrt{n} \cdot \Phi((i+1)\delta) \cdot \phi((i+1)\delta) \cdot (1 \pm \Gamma(i+1)). \\ \lambda_2(g, i+1, D) &= n \cdot \phi((i+1)\delta)^2 \cdot (1 \pm \Gamma(i+1)). \end{aligned}$$

Lastly, note that by Lemma 4.2, $\Pr[\mathcal{E}^{**}] = 1 - n^{-\omega(1)}$. This completes the proof.

5 Proof of Lemma 4.2

Given Fact 2.1, the first two items in the lemma follow from Chernoff's bound. We prove the last item. Assume for the rest of this section that $F \subset K_n$ is a triangle-free graph of size $O(1)$ and condition on the event $\{\mathbb{TF}(n,i) \cup F \text{ is triangle-free}, \mathcal{E}^{**}\}$. We prove that $\Pr[\mathcal{E}_F^*] \geq 1 - K^{-1/10}$. For every $g \in K_n$, set $\Lambda(g,i) := \Lambda_1(g,i) \cup \Lambda_2(g,i)$.

Definition 3 (bad-sequence)**.** *Let $S = (G_1, G_2, \ldots, G_d)$ be a sequence of subgraphs of K_n with $1 \leq d \leq 2D$. We say that S is a* bad-sequence *if the following properties hold simultaneously.*

- *For every $j \in [d]$, $G_j \in \Lambda(g,i)$ for some edge $g \in F \cup \bigcup_{l<j} G_l$.*
- *For every $j \in [d-1]$, G_j shares $|G_j|$ vertices and 0 edges with $F \cup \bigcup_{l<j} G_l$.*
- *Either*
  - *G_d shares $|G_d| + 1$ vertices and at most $|G_d| - 1$ edges with $F \cup \bigcup_{l<d} G_l$, or*
  - *G_d shares $|G_d|$ vertices and 0 edges with $F \cup \bigcup_{l<d} G_l$. In addition, let $\{x,y\} \in F \cup \bigcup_{l<d} G_l$ be such that $G_d \in \Lambda(\{x,y\}, i)$. Let z be the vertex of G_d that doesn't belong to $F \cup \bigcup_{l<d} G_l$. Then there is a vertex $w \notin \{x,y,z\}$ in $F \cup \bigcup_{l<d} G_l$ such that $\{w,z\} \in \mathbb{TF}(n,i)$.*

Let $\mathcal{E}$ be the event that for every bad-sequence $S = (G_1, G_2, \ldots, G_d)$ there exists $j \in [d]$ such that $\{G_j \not\subseteq \mathbb{B}_{i+1}^*\}$. The next two propositions imply the desired bound $\Pr[\mathcal{E}_F^*] \geq 1 - K^{-1/10}$, as they state that $\mathcal{E}$ implies $\mathcal{E}_F^*$ and $\Pr[\mathcal{E}] \geq 1 - K^{-1/10}$.

Proposition 5.1. *$\mathcal{E}$ implies $\mathcal{E}_F^*$.*

Proof. Assume $\mathcal{E}$ occurs. We have the following claim.

Claim 5.2. *Let $P = (v_0, u_1, v_1, \ldots, u_D, v_D)$ denote an arbitrary path in $T_{F,D}^*$, starting from some root v_0 and ending with some leaf. Let G_j be the label of node u_j and let g_j be the label of node v_j (so that $g_0 \in F$). For all $j \in [D]$, G_j shares 0 edges with $F \cup \bigcup_{l<j} G_l$.*

Proof. Suppose for the sake of contradiction that the claim is false, and fix the minimal $d \in [D]$ for which G_d shares some edge with $F \cup \bigcup_{l<d} G_l$. Consider the sequence $S = (G_1, G_2, \ldots, G_d)$.

We observe the following: For all $j \in [d-1]$, G_j shares $|G_j|$ vertices and 0 edges with $F \cup \bigcup_{l<j} G_l$. The fact that G_j shares 0 edges with $F \cup \bigcup_{l<j} G_l$ follows from the minimality of d. Now, for all $j \in [d-1]$, G_j shares at least $|G_j|$ vertices with $F \cup \bigcup_{l<j} G_l$. If there exists $j \in [d-1]$ such that G_j shares all $|G_j| + 1$ vertices with $F \cup \bigcup_{l<j} G_l$, then since in that case G_j shares $0 \leq |G_j| - 1$ edges with $\bigcup_{l<j} G_l$, clearly we also have that some prefix of S is a bad-sequence, and this contradicts $\mathcal{E}$.

Suppose that $|G_d| = 2$. By assumption we have that G_d shares some edge with $F \cup \bigcup_{l<d} G_l$, which also implies that G_d shares $|G_d| + 1$ vertices with $F \cup \bigcup_{l<d} G_l$. Hence, by the observation above, in order to show that S is a bad-sequence and reach a contradiction, it remains to show that G_d shares exactly 1 edge with $F \cup \bigcup_{l<d} G_l$. Suppose on the contrary that G_d shares both of its 2 edges with $F \cup \bigcup_{l<d} G_l$. Notice that since F is triangle-free, this implies that $d \geq 2$, so g_{d-2} is well defined. Write $g_{d-2} = \{x,y\}$ and $g_{d-1} = \{x,z\}$ and note that $G_{d-1} \in \Lambda(g_{d-2}, i)$ and

$G_d \in \Lambda(g_{d-1}, i)$. Now, note that the edge in G_d that is adjacent to z must also be an edge in G_{d-1}. This is true since otherwise, the subgraph G_{d-1} will share the vertex z with $F \cup \bigcup_{l<d-1} G_l$, which is clearly not the case as by the observation above G_{d-1} shares only vertices from $\{x, y\}$ with $F \cup \bigcup_{l<d-1} G_l$. The only possible edge to be adjacent in G_d to z and be in G_{d-1} is the edge $\{y, z\}$. Hence we get that y is a vertex of G_d. Therefore, we conclude that $g_{d-2} \in G_d$. But by the definition of $T^*_{F,D}$, $g_{d-2} \notin G_d$. Therefore, G_d shares exactly 1 edge with $F \cup \bigcup_{l<d} G_l$, which implies that S is a bad-sequence–a contradiction.

Next assume that $|G_d| = 1$. By assumption we have that G_d shares its edge with $F \cup \bigcup_{l<d} G_l$. Since $\mathbb{TF}(n, i) \cup F$ is triangle-free, this implies $d \geq 2$ and so g_{d-2} is well defined. Let x, y, z be as defined in the previous paragraph. We have two case.

- Suppose z is a vertex of G_d. Then similarly to the previous paragraph we get that G_d must be the edge $\{y, z\}$. In that case we get that $\{g_{d-2}, g_{d-1}, g_d\}$ is a triangle. This contradicts the definition of $T^*_{F,D}$.
- Suppose x is a vertex of G_d. Note that by the observation above, the vertex z of G_{d-1} doesn't belong to $F \cup \bigcup_{l<d-1} G_l$. Let $w \neq x$ be the other vertex of the edge in G_d. By definition of $T^*_{F,D}$, $w \notin \{x, y, z\}$. Hence, by assumption, w is a vertex of $F \cup \bigcup_{l<d-1} G_l$. Since $|G_d| = 1$, we have that $\{w, z\} \in \mathbb{TF}(n, i)$. With the observation above, this implies that, by definition, that $(G_j)_{j=1}^{d-1}$ is a bad-sequence. This contradicts $\mathcal{E}$.

■

The next claim, when combined with Claim 5.2, implies the proposition.

Claim 5.3. *Let u_d, $1 \leq d \leq D$, be a node at distance $2d-1$ from a root in $T^*_{F,D}$. Let $u'_{d'}$, $1 \leq d' \leq d$, be a different node at distance $2d' - 1$ from a root in $T^*_{F,D}$. Then the labels of u_d and $u'_{d'}$ share no edge.*

Proof. The proof is by induction on d. For the base case $d = 1$, let u_1 and u'_1 be two distinct nodes at distance 1 from the roots of $T^*_{F,D}$. Let G_1 and G'_1 be the labels of u_1 and u'_1. Assume for the sake of contradiction that G_1 and G'_1 share an edge. We claim that either (G_1) or (G_1, G'_1) is a bad-sequence thus reaching the desired contradiction. To see that this indeed holds, note first that by Claim 5.2, G_1 shares $|G_1|$ vertices and 0 edges with F and G'_1 shares $|G'_1|$ vertices and 0 edges with F. Let v_0 and v'_0 be the parents of u_1 and u'_1 respectively. Since G_1 and G'_1 share an edge and $u_1 \neq u'_1$, we have that $v_0 \neq v'_0$. Therefore G_1 and G'_1 share exactly 1 edge. Let g_0 and g'_0 be the labels of v_0 and v'_0, respectively. Now, if $|G'_1| = 2$ we get from the above that G'_1 shares $|G'_1| + 1$ vertices and $|G'_1| - 1$ edges with $F \cup G_1$, which implies that (G_1, G'_1) is a bad-sequence, contradicting $\mathcal{E}$. Hence, assume $|G'_1| = 1$. Let z be the vertex of G_1 and G'_1 that is not in F. Then it is easy to see given the above that there is a vertex w in F (more accurately a vertex in g'_0) such that w is not a vertex of g_0 and $\{w, z\} \in \mathbb{TF}(n, i)$. Hence, (G_1) is a bad-sequence, contradicting $\mathcal{E}$.

Let $2 \leq d \leq D$ and assume the claim is valid for $d - 1$. Let u_d be a node at distance $2d - 1$ from a root in $T^*_{F,D}$ and let $u'_{d'}$, $1 \leq d' \leq d$, be a different node at distance $2d' - 1$ from a root in $T^*_{F,D}$. Let $P = (v_0, u_1, v_1, \ldots, u_{d-1}, v_{d-1}, u_d)$ be the unique path from a root in $T^*_{F,D}$ to u_d and let $P' = (v'_0, u'_1, v'_1, \ldots, u'_{d'-1}, v'_{d'-1}, u'_{d'})$ be the unique path from a root in $T^*_{F,D}$ to $u'_{d'}$. Let G_j be the label of u_j and G'_j the label of u'_j. Assume for the sake of contradiction that G_d shares an edge with $G'_{d'}$. Without loss of generality we further assume that d' is minimal in the sence that there exists no $d'' < d'$ such that G_d shares an edge with $G'_{d''}$.

Let u'_r be the first node in P' which is not a node that appears in P. Consider the sequences $S_1 = (G_1, G_2, \ldots, G_d, G'_r, G'_{r+1}, \ldots, G'_{d'})$ and $S_2 = (G_1, G_2, \ldots, G_{d-1}, G'_r, G'_{r+1}, \ldots, G'_{d'-1}, G_d)$, where $(G'_r, G'_{r+1}, \ldots, G'_{d'-1})$ is the empty sequence in case $r = d'$. (Note that $G'_{d'}$ does not appear in S_2.) We reach a contradiction by showing that either S_1 or S_2 are bad-sequences.

Assume that $|G'_{d'}| = 2$. For brevity rewrite the sequence S_1 as $S_1 = (F_1, F_2, \ldots, F_s)$, where $s = d + d' - r + 1$, $F_d = G_d$ and $F_s = G'_{d'}$. By Claim 5.2, the minimality of d' and the induction hypothesis we have that for every $j \in [s-1]$, F_j shares 0 edges with $F \cup \bigcup_{l<j} F_l$. Therefore, by $\mathcal{E}$ we also have that for every $j \in [s-1]$, F_j shares $|F_j|$ vertices with $F \cup \bigcup_{l<j} F_l$. Now, since $u_d \neq u'_{d'}$ and yet F_d and F_s share an edge, we get that $v_{d-1} \neq v'_{d'-1}$. This implies by the induction hypothesis that the label of v_{d-1} is not the label of $v'_{d'-1}$, and so we get that F_d shares *exactly* 1 edge with F_s. In what follows we show that F_s shares 0 edges with $F \cup \bigcup_{l<d, d<l<s} F_l$. This will imply that F_s shares $|F_s|+1$ vertices and $1 = |F_s| - 1$ edges with $F \cup \bigcup_{l<s}$, which given the above implies that S_1 is a bad-sequence–a contradiction. The fact that F_s shares 0 edges with $F \cup \bigcup_{d<l<s}$ follows from Claim 5.2. We argue that F_s shares 0 edges with $\bigcup_{l<d} F_l$. Indeed, if F_s does share an edge with $\bigcup_{l<d} F_l$, then since F_s also shares an edge with F_d, we get by Claim 5.2 that F_d shares $|F_d|+1$ vertices and $0 \leq |F_d| - 1$ edges with $F \cup \bigcup_{l<d}$. This implies, again by Claim 5.2, that $(F_1, F_2, \ldots, F_d)$ is a bad-sequence–a contradiction.

Assume that $|G'_{d'}| = 1$. For brevity rewrite the sequence S_2 as $S_2 = (F_1, F_2, \ldots, F_s)$, where $s = d + d' - r$ and $F_s = G_d$. By Claim 5.2, the minimality of d' and the induction hypothesis we have that for every $j \in [s]$, F_j shares 0 edges with $F \cup \bigcup_{l<j} F_l$. Therefore, by $\mathcal{E}$ we also have that for every $j \in [s-1]$, F_j shares $|F_j|$ vertices with $F \cup \bigcup_{l<j} F_l$. Let v (resp. v') be the parent of u_d (resp. $u'_{d'}$). Let g (resp. g') be the label of v (resp. v'). Since $u_d \neq u'_{d'}$ and yet G_d and $G'_{d'}$ share an edge, we have that $v \neq v'$. By the induction hypothesis we thus get that $g \neq g'$. Write $g = \{x, y\}$. Let z be the vertex of G_d that is not in $\{x, y\}$ and note that since F_s shares $|F_s|$ vertices with $F \cup \bigcup_{l<s} F_l$ and these vertices are only from $\{x, y\}$, we have that z doesn't belong to $F \cup \bigcup_{l<s} F_l$. Now observe that g' is an edge in $F \cup \bigcup_{l<s} F_l$. Hence, z is not a vertex of g'. Also, since $g \neq g'$ we have that there is a vertex w in g' that is not in $g = \{x, y\}$. Moreover, since $|G_{d'}| = 1$ we have that $\{w, z\} \in \mathbb{TF}(n, i)$. Therefore, there is a vertex $w \notin \{x, y, z\}$ in $F \cup \bigcup_{l<s} F_l$ such that $\{w, z\} \in \mathbb{TF}(n, i)$. This implies that S_2 bad-sequence–a contradiction. ■

With that we complete the proof of the proposition. ■

Proposition 5.4. $Pr[\mathcal{E}] \geq 1 - K^{-1/10}$.

Proof. For a bad-sequence $S = (G_1, G_2, \ldots, G_d)$, write $\{S \subseteq \mathbb{B}^*_{i+1}\}$ for the event that for all $j \in [d]$, $\{G_j \subseteq \mathbb{B}^*_{i+1}\}$. Let Z be the random variable counting the number of bad-sequences S for which $\{S \subseteq \mathbb{B}^*_{i+1}\}$. It suffices to show $\mathbb{E}[Z] \leq K^{-1/10}$.

For $d \in [2D]$, $0 \leq c \leq d$, let $\mathrm{Seq}_{d,c,1}$ denote the set of all bad-sequences $S = (G_1, G_2, \ldots, G_d)$ with $c = |\{j : |G_j| = 1\}|$ such that G_d shares $|G_d|+1$ vertices and at most $|G_d|-1$ edges with $F \cup \bigcup_{l<d} G_l$. For $d \in [2D]$, $0 \leq c \leq d$, let $\mathrm{Seq}_{d,c,2}$ denote the set of all bad-sequences $S = (G_1, G_2, \ldots, G_d)$ with

$c = |\{j : |G_j| = 1\}|$ that are not in $\mathrm{Seq}_{d,c,1}$. Then

$$\mathbb{E}[Z] \quad = \quad \sum_{d \in [2D]} \sum_{0 \le c \le d} \sum_{j \in \{1,2\}} \sum_{S \in \mathrm{Seq}_{d,c,j}} \Pr[S \subseteq \mathbb{B}^*_{i+1}]. \tag{2}$$

Below we show that

$$\forall d \in [2D],\ 0 \le c \le d. \quad \sum_{S \in \mathrm{Seq}_{d,c,1}} \Pr[S \subseteq \mathbb{B}^*_{i+1}] \quad \le \quad K^{-1/9}, \tag{3}$$

$$\forall d \in [2D],\ 0 \le c \le d. \quad \sum_{S \in \mathrm{Seq}_{d,c,2}} \Pr[S \subseteq \mathbb{B}^*_{i+1}] \quad \le \quad K^{-1/9}. \tag{4}$$

From (2), (3) and (4) and since $D = O(1)$, we get that $\mathbb{E}[Z] \le K^{-1/10}$ as required. Hence, it remains to prove (3) and (4).

We prove (3). Fix $d \in [2D]$, $0 \le c \le d$. We first count the number of sequences $S = (G_1, G_2, \ldots, G_d)$ in $\mathrm{Seq}_{d,c,1}$. To do so, we construct such a sequence iteratively. First, we choose the cardinalities of the subgraphs in S. Note that there are $\binom{d}{c} = O(1)$ possible choices for the cardinalities. Now, given that we have chosen the first $j-1$ subgraphs in S for $j \in [d-1]$, we count the number of choices for G_j. There are $O(1)$ possible choices for an edge $g \in F \cup \bigcup_{l<j} G_l$ for which $G_j \in \Lambda(g,i)$. Given g: if $|G_j|$ is to be of size 1 then there are at most $\Lambda_1(g,i)$ choices for G_j and if $|G_j|$ is to be of size 2 then there are at most $\Lambda_2(g,i)$ choices for G_j. Given that we have already chosen the first $d-1$ subgraphs in S, the number of choices for G_d is at most $O(1)$, since the vertices of G_d are all in $F \cup \bigcup_{l<d} G_l$. Therefore, and here we use the occurence of $\mathcal{E}^{**}$, the number of sequences in $\mathrm{Seq}_{d,c,1}$ is at most

$$O(1) \cdot \left(K^2 \phi(i\delta)^2\right)^{d-1-c} \cdot \left(K \Phi(i\delta) \phi(i\delta)\right)^c.$$

Now, given $\mathcal{E}^{**}$, one can verify that the probability of $\{S \subseteq \mathbb{B}^*_{i+1}\}$ for $S \in \mathrm{Seq}_{d,c,1}$ is at most

$$\left(\frac{k^2}{K^2}\right)^{d-1-c} \cdot \left(\frac{k}{K}\right)^c \cdot \frac{k}{K}.$$

Hence,

$$\begin{aligned}
\sum_{S \in \mathrm{Seq}_{d,c,1}} \Pr[S \subseteq \mathbb{B}^*_{i+1}] \quad &\le \quad O(1) \cdot \left(k^2 \phi(i\delta)^2\right)^{d-1-c} \cdot \left(k \Phi(i\delta) \phi(i\delta)\right)^c \cdot \frac{k}{K} \\
&\le \quad O(1) \cdot k^{2d-2-2c} \cdot (k \ln n)^c \cdot \frac{k}{K} \\
&\le \quad K^{-1/9},
\end{aligned}$$

where the second inequality follows by Fact 2.1 (i) and the last inequality follows by the definition of D, k and K. This gives us the validity of (3).

It remains to prove (4). Fix $d \in [2D]$, $0 \le c \le d$. As before, we first count the number of sequences $S = (G_1, G_2, \ldots, G_d)$ in $\mathrm{Seq}_{d,c,2}$ and we do it by constructing such a sequence iteratively. First, we choose the cardinalities of the subgraphs in S and we note that there are $\binom{d}{c} = O(1)$ possible choices for these cardinalities. Now, given that we have chosen the first $j-1$ subgraphs in

S for $j \in [d-1]$, we count the number of choices for G_j. There are $O(1)$ possible choices for an edge $g \in F \cup \bigcup_{l<j} G_l$ for which $G_j \in \Lambda(g,i)$. Given g: if $|G_j|$ is to be of size 1 then there are at most $\Lambda_1(g,i)$ choices for G_j and if $|G_j|$ is to be of size 2 then there are at most $\Lambda_2(g,i)$ choices for G_j. Suppose we have already chosen the first $d-1$ subgraphs in S. We claim that the number of choices for G_d is atmost $O((\ln n)^2)$. Indeed, there are $O(1)$ choices for an edge $g = \{x,y\} \in F \cup \bigcup_{l<d} G_l$ such that $G_d \in \Lambda(g,i)$. Given g, there are at most $O(1)$ choices for a vertex w in $F \cup \bigcup_{l<d} G_l$ such that $w \notin \{x,y\}$ and given $\mathcal{E}^{**}$, there are at most $(\ln n)^2$ choices for the vertex z of G_d which is not a vertex of $F \cup \bigcup_{l<d} G_l$ such that $\{w,z\} \in \mathbb{TF}(n,i)$. Therefore, and here we use again the occurence of $\mathcal{E}^{**}$, the number of sequences in $\mathrm{Seq}_{d,c,2}$ is at most

$$O(1) \cdot \left(K^2\phi(i\delta)^2\right)^{d-1-c} \cdot \left(K\Phi(i\delta)\phi(i\delta)\right)^c \cdot (\ln n)^2.$$

Given $\mathcal{E}^{**}$, one can verify that the probability of $\{S \subseteq \mathbb{B}^*_{i+1}\}$ for $S \in \mathrm{Seq}_{d,c,1}$ is at most

$$\left(\frac{k^2}{K^2}\right)^{d-1-c} \cdot \left(\frac{k}{K}\right)^c \cdot \frac{k}{K}.$$

Therefore,

$$\begin{aligned}
\sum_{S \in \mathrm{Seq}_{d,c,2}} \Pr[S \subseteq \mathbb{B}^*_{i+1}] &\leq O(1) \cdot \left(k^2\phi(i\delta)^2\right)^{d-1-c} \cdot \left(k\Phi(i\delta)\phi(i\delta)\right)^c \cdot \frac{k}{K} \cdot (\ln n)^2 \\
&\leq O(1) \cdot k^{2d-2-2c} \cdot (k \ln n)^c \cdot \frac{k}{K} \cdot (\ln n)^2 \\
&\leq K^{-1/9},
\end{aligned}$$

where as before, the second inequality follows by Fact 2.1 (i) and the last inequality follows by the definition of d, k and K. This gives us the validity of (4). With that we complete the proof. ■

6 Proof of Lemma 4.3

Let f, i be as specified in the Main Lemma. Condition throughout this section on:

$$\{f \notin \mathbb{B}_{\leq i},\ \mathbb{TF}(n,i) \cup \{f\} \text{ is triangle-free},\ \mathcal{E}^*,\ \mathcal{E}^*_{\{f\}}\}.$$

Definition 4 (T_∞, T_d)**.**

- *Let T_∞ be an infinite rooted tree, defined as follows. Every node g at even distance from the root has two sets of children. One set consists of children which are singletons and the other set consists of children which are sets of size 2. Every node G at odd distance from the root of T_∞, which is a set of size $|G| \in \{1,2\}$ has exactly $|G|$ children. Lastly, for every node g at even distance from the root:*

$$\begin{aligned}
\textit{Number of children of } g \textit{ that are of size } 1 &= \lceil 2k\Phi(i\delta)\phi(i\delta)\rceil, \\
\textit{Number of children of } g \textit{ that are of size } 2 &= k^2\phi(i\delta)^2.
\end{aligned}$$

- *Let $1 \leq d \leq D$. Define T_d to be the tree that is obtained by removing from T_∞ every node whose distance from the root is larger than $2d$.*

Remark 6.1: We assume from now on that $2k\Phi(i\delta)\phi(i\delta)$ is an integer. Hence, for example, the number of children of the root of T_∞ that are of size 1 is exactly $2k\Phi(i\delta)\phi(i\delta)$. As we discuss in Section 8, our proof can be modified for the case where $2k\Phi(i\delta)\phi(i\delta)$ is not an integer.

Some remarks regarding $T^*_{f,D}$ follow. The event $\mathcal{E}^*_{\{f\}}$ says that every label that appears at some node in $T^*_{f,D}$ appears exactly once. Therefore, we shall refer from now on to the nodes of $T^*_{f,D}$ *by their labels*. Given $\mathcal{E}^*$ and by definition of $T^*_{f,D}$ and Fact 2.1, it is easily seen that for every non-leaf node g at even distance from the root f of $T^*_{f,D}$,

$$\begin{aligned}\text{Number of children of } g \text{ that are of size 1} &= 2k\Phi(i\delta)\phi(i\delta)(1\pm 1.3\Gamma(i)),\\ \text{Number of children of } g \text{ that are of size 2} &= k^2\phi(i\delta)^2(1\pm 1.3\Gamma(i)).\end{aligned}$$

Note that for every node $g \neq f$ at even distance from the root of $T^*_{f,D}$, $\beta_{i+1}(g)$ is mapped uniformly at random to the interval $[0, kn^{-1/2}]$. Also note that by the condition $\{f \notin \mathbb{B}_{\leq i}\}$, $\beta_{i+1}(f)$ is distributed uniformly at random in $[0,1]$. We extend the definition of β_{i+1} so that in addition, for every node g at even distance from the root of T_∞ (and hence from the root of T_D), $\beta_{i+1}(g)$ is mapped uniformly at random to the interval $[0, kn^{-1/2}]$. We recall the definition of survival in $T^*_{f,D}$ and extend it to the trees T_D and T_∞. Let $T \in \{T^*_{f,D}, T_D, T_\infty\}$. Let g be a node at even distance from the root of T. We define the event that g *survives* as follows. If g is a leaf (so that $T \neq T_\infty$) then g survives by definition. Otherwise, g survives if and only if for every child G of g, the following holds: If $\beta_{i+1}(g') < \beta_{i+1}(g)$ for all $g' \in G$ then G has a child that does not survive.

For a node g at height $2d$ in $T^*_{f,D}$, let $p_{g,d}(x)$ be the probability that g survives under the assumption that $\beta_{i+1}(g) = xn^{-1/2}$. Let $q_d(x)$ be the probability that the root of T_d survives under the assumption that $\beta_{i+1}(g) = xn^{-1/2}$, where g here denotes the root of T_d. Let $r(x)$ be the probability that the root of T_∞ survives under the assumption that $\beta_{i+1}(g) = xn^{-1/2}$, where g here denotes the root of T_∞. One can show that $p_{g,d}(x), q_d(x)$ and $r(x)$ are all continuous and bounded in the interval $[0,\delta]$. Hence, we can define the following functions on the interval $[0,\delta]$:

$$P_{g,d}(x) := \int_0^x p_{g,d}(y)dy, \qquad Q_d(x) := \int_0^x q_d(y)dy \quad \text{and} \quad R(x) := \int_0^x r(y)dy.$$

Observe that for all $x \in (0,\delta]$:

$$\begin{aligned}\Pr[\mathcal{A}^*_{f,D} \,|\, \beta_{i+1}(f) < xn^{-1/2}] &= \frac{P_{g,d}(x)}{x},\\ \Pr[\text{The root } g \text{ of } T_d \text{ survives} \,|\, \beta_{i+1}(g) < xn^{-1/2}] &= \frac{Q_d(x)}{x},\\ \Pr[\text{The root } g \text{ of } T_\infty \text{ survives} \,|\, \beta_{i+1}(g) < xn^{-1/2}] &= \frac{R(x)}{x}.\end{aligned}$$

The next lemma, when combined with the discussion above, implies Lemma 4.3.

Lemma 6.2.

(i)

$$R(\delta) = \begin{cases} \Phi((i+1)\delta) & \text{if } i=0,\\ \frac{\Phi((i+1)\delta)-\Phi(i\delta)}{\phi(i\delta)} & \text{if } i \geq 1,\end{cases} \tag{5}$$

$$r(\delta) = \begin{cases} \phi((i+1)\delta) & \text{if } i=0,\\ \frac{\phi((i+1)\delta)}{\phi(i\delta)} & \text{if } i \geq 1.\end{cases} \tag{6}$$

(ii) For all $x \in [0, \delta]$, $q_D(x) = r(x)(1 \pm o(\Gamma(i)\gamma(i)))$.

(iii) For all $x \in [0, \delta]$, $p_{f,D}(x) = q_D(x)(1 \pm 8\Gamma(i)\gamma(i))$.

The proof of Lemma 6.2 is given in the next three subsections.

6.1 Proof of Lemma 6.2 (i)

Clearly $R(0) = 0$ and $r(0) = 1$. Hence, from the definition of survival and the definition of $r(x)$ and $R(x)$, we get that for every $x \in [0, \delta]$, at the limit as $n \to \infty$,

$$\begin{aligned} r(x) &= \left(1 - \frac{R(x)^2}{k^2}\right)^{k^2\phi(i\delta)^2} \cdot \left(1 - \frac{R(x)}{k}\right)^{2k\Phi(i\delta)\phi(i\delta)} \\ &= \exp\Big(-R(x)^2\phi(i\delta)^2 - 2R(x)\Phi(i\delta)\phi(i\delta)\Big). \end{aligned} \tag{7}$$

By the fundamental theorem of calculus, $r(x)$ is the derivarive of $R(x)$. Hence, we view (7) as the separable differential equation that it is. This equation has the following as an implicit solution:

$$\int \exp\left(R^2\phi(i\delta)^2 + 2R\phi(i\delta)\Phi(i\delta)\right) dR = x.$$

Solving the above integral, we get

$$\frac{\sqrt{\pi}}{2} \cdot \operatorname{erfi}\big(\Phi(i\delta) + \phi(i\delta)R\big) = x + C. \tag{8}$$

The initial condition is $x = 0$, $R = 0$. Hence, from (8) we get

$$C = \frac{\sqrt{\pi}}{2} \cdot \operatorname{erfi}(\Phi(i\delta)).$$

Let $z > 0$ satisfy

$$\exp(-z^2\phi(i\delta)^2 - 2z\phi(i\delta)\Phi(i\delta)) = \frac{\phi((i+1)\delta)}{\phi(i\delta)}.$$

A simple analysis shows that

$$z = \frac{\Phi((i+1)\delta) - \Phi(i\delta)}{\phi(i\delta)}.$$

Taking $R = z$ and C as above, we solve (8) for x to get

$$x = \frac{\sqrt{\pi}}{2} \cdot \operatorname{erfi}\big(\Phi(i\delta) + \phi(i\delta)R\big) - C = \frac{\sqrt{\pi}}{2} \cdot \big(\operatorname{erfi}(\Phi((i+1)\delta)) - \operatorname{erfi}(\Phi(i\delta))\big) = \delta,$$

where the last equality is by (1). Hence, $R(\delta) = \frac{\Phi((i+1)\delta) - \Phi(i\delta)}{\phi(i\delta)}$ and $r(\delta) = \frac{\phi((i+1)\delta)}{\phi(i\delta)}$.

6.2 Proof of Lemma 6.2 (ii)

Assume first that D is odd. Let g_0 be the root of T_D and T_∞. Assume $\beta_{i+1}(g_0) = xn^{-1/2}$ for some $x \in [0, \delta]$. Clearly if g_0 survives in T_D then g_0 survives in T_∞. Hence $q_D(x) \le r(x)$. Below we show that $q_D(x) \ge r(x) - n^{-36\varepsilon}$. We claim that this last inequality implies $q_D(x) \ge r(x)(1 - o(\Gamma(i)\gamma(i)))$, which gives the lemma. Indeed, using the fact that $x \le \delta = o(1)$, it is easy to verify that $r(x) \sim 1$. In addition, by Fact 2.1 we have that $\Gamma(i)\gamma(i) = \Omega(n^{-35\varepsilon})$. Therefore we get, as needed,

$$q_D(x) \ge r(x)(1 - n^{-36\varepsilon}/r(x)) \ge r(x)(1 - o(\Gamma(i)\gamma(i))).$$

Say that a node g at even distance from the root of T_D is *relevant* if g and its sibling (if exists) are born before their grandparent, and in addition, their grandparent is either relevant or the root. Observe that if the root of T_∞ survives then either the root of T_D survives, or else, there is a relevant node in T_D at distance $2D$ from the root. It remains to show that under the assumption $\beta_{i+1}(g_0) = xn^{-1/2}$, the expected number of relevant nodes at distance $2D$ from the root of T_D is at most $n^{-36\varepsilon}$.

Say that a node g_D at distance $2D$ from the root of T_D is a c-node if the path leading from the root to g_D contains exactly c nodes G at odd distance from the root, which are sets of size 1. Consider a path $(g_0, G_1, g_1, \ldots, G_D, g_D)$ from the root to a node g_D at distance $2D$ from the root of T_D, where g_D is a c-type. Let $\mathcal{G}$ be the union of $\{g_j : j \in [D]\}$ together with the set $\{g : g \text{ is a sibling of some } g_j,\ j \in [D]\}$. Since g_D is a c-type, we have $|\mathcal{G}| = 2D - c$. Now if g_D is relevant, then for every node $g \in \mathcal{G}$, $\{\beta_{i+1}(g) < \beta_{i+1}(g_0)\}$ holds. This event occurs with probability $(x/k)^{2D-c}$. Hence, the probability that g_D is relevant is at most

$$\left(\frac{x}{k}\right)^{2D-c} \quad = \quad \left(\frac{x}{k}\right)^{c} \cdot \left(\frac{x^2}{k^2}\right)^{D-c}.$$

The number of c-nodes at distance $2D$ from the root of T_D is at most

$$2^D \cdot (4k\Phi(i\delta)\phi(i\delta))^c \cdot (2k^2\phi(i\delta)^2)^{D-c} \quad \le \quad (8k \ln n)^c \cdot (8k^2)^{D-c},$$

where the inequality is by Fact 2.1. Hence, the expected number of relevant c-nodes at distance $2D$ from the root of T_D is at most

$$\left(\frac{x}{k}\right)^{c} \cdot \left(\frac{x^2}{k^2}\right)^{D-c} \cdot (8k \ln n)^c \cdot (8k^2)^{D-c} \quad \le \quad (8x \ln n)^{2D-c}.$$

Now, $(8x \ln n)^{2D-c} \le \delta^{D-1} = n^{-39\varepsilon}$, where the inequality is by $x \le \delta$ and $c \le D$, while the equality is by the choice of D. We complete the proof by noting that there are at most $D+1 = O(1)$ choices for c, which by the union bound implies that the expected number of relevant nodes at distance $2D$ from the root of T_D is at most $n^{-36\varepsilon}$.

Assume D is even, let g_0 be as above and assume $\beta_{i+1}(g_0) = xn^{-1/2}$. The proof for this case is similar to the previous case and we only outline it. It is easy to verify that if g_0 doesn't survive in T_D then g_0 doesn't survive in T_∞. Hence $q_D(x) \ge r(x)$. Now, if g_0 doesn't survive in T_∞ then either the root of T_D doesn't survive, or else, there is a relevant node in T_D at distance $2D$ from the root. One can now show using the same argument as above that the expected number of relevant nodes at distance $2D$ from the root of T_D is at most $n^{-36\varepsilon}$. This completes the proof.

6.3 Proof of Lemma 6.2 (iii)

The following implies Lemma 6.2 (iii).

Proposition 6.3. *Let g be a node at height $2d$ in $T^*_{f,D}$. Let $x \in [0, \delta]$. Then assuming $i \geq 1$,*

$$q_d(x)(1 - 8\Gamma(i)\gamma(i)) \leq p_{g,d}(x) \leq q_d(x)(1 + 8\Gamma(i)\gamma(i)).$$

To prove Proposition 6.3, we need the following inequalities.

Claim 6.4. *For all $d \in [D], x \in (0, \delta]$,*

(i) $\left(1 - \frac{x}{k} \cdot \frac{(1-8\Gamma(i)\gamma(i))Q_{d-1}(x)}{x}\right)^{|\Lambda_1(g,i)|} \leq \left(1 - \frac{Q_{d-1}(x)}{k}\right)^{2k\Phi(i\delta)\phi(i\delta)}(1 + 3\Gamma(i)\gamma(i)).$

(ii) $\left(1 - \frac{x^2}{k^2} \cdot \frac{(1-8\Gamma(i)\gamma(i))^2 Q_{d-1}(x)^2}{x^2}\right)^{|\Lambda_2(g,i)|} \leq \left(1 - \frac{Q_{d-1}(x)^2}{k^2}\right)^{k^2\phi(i\delta)^2}(1 + 2\Gamma(i)\gamma(i)).$

(iii) $\left(1 - \frac{x}{k} \cdot \frac{(1+8\Gamma(i)\gamma(i))Q_{d-1}(x)}{x}\right)^{|\Lambda_1(g,i)|} \geq \left(1 - \frac{Q_{d-1}(x)}{k}\right)^{2k\Phi(i\delta)\phi(i\delta)}(1 - 3\Gamma(i)\gamma(i)).$

(iv) $\left(1 - \frac{x^2}{k^2} \cdot \frac{(1+8\Gamma(i)\gamma(i))^2 Q_{d-1}(x)^2}{x^2}\right)^{|\Lambda_2(g,i)|} \geq \left(1 - \frac{Q_{d-1}(x)^2}{k^2}\right)^{k^2\phi(i\delta)^2}(1 - 2\Gamma(i)\gamma(i)).$

Proof. (i) We have

$$\begin{aligned}
&\left(1 - \frac{x}{k} \cdot \frac{(1 - 8\Gamma(i)\gamma(i))Q_{d-1}(x)}{x}\right)^{|\Lambda_1(g,i)|} \\
\leq\ & \left(1 - \frac{x}{k} \cdot \frac{(1 - 8\Gamma(i)\gamma(i))Q_{d-1}(x)}{x}\right)^{2k\Phi(i\delta)\phi(i\delta)} \cdot \left(1 - \frac{x}{k} \cdot \frac{(1 - 8\Gamma(i)\gamma(i))Q_{d-1}(x)}{x}\right)^{-2k\Phi(i\delta)\phi(i\delta)\Gamma(i)} \\
\leq\ & \left(1 - \frac{x}{k} \cdot \frac{(1 - 8\Gamma(i)\gamma(i))Q_{d-1}(x)}{x}\right)^{2k\Phi(i\delta)\phi(i\delta)} \cdot \left(1 - \frac{\delta}{k}\right)^{-2k\Phi(i\delta)\phi(i\delta)\Gamma(i)},
\end{aligned}$$

where the first inequality is by $|\Lambda_1(g,i)| \geq 2k\Phi(i\delta)\phi(i\delta)(1 - \Gamma(i))$ and the second inequality is by $(1 - 8\Gamma(i)\gamma(i))Q_{d-1}(x)/x \leq 1$ and $x \leq \delta$. We further have

$$\begin{aligned}
&\left(1 - \frac{x}{k} \cdot \frac{(1 - 8\Gamma(i)\gamma(i))Q_{d-1}(x)}{x}\right)^{2k\Phi(i\delta)\phi(i\delta)} \\
\leq\ & \left(1 - \frac{x}{k} \cdot \frac{Q_{d-1}(x)}{x}\right)^{2k\Phi(i\delta)\phi(i\delta)(1-36\Gamma(i)\gamma(i))(1-1/k)} \\
\leq\ & \left(1 - \frac{x}{k} \cdot \frac{Q_{d-1}(x)}{x}\right)^{2k\Phi(i\delta)\phi(i\delta)(1-150\Gamma(i)\gamma(i))} \\
\leq\ & \left(1 - \frac{x}{k} \cdot \frac{Q_{d-1}(x)}{x}\right)^{2k\Phi(i\delta)\phi(i\delta)} \cdot \left(1 - \frac{\delta}{k}\right)^{-300k\Phi(i\delta)\phi(i\delta)\Gamma(i)\gamma(i)},
\end{aligned}$$

where the first inequality follows essentially from the fact that for every $z > 1$, $\exp(-1/(z-1)) < 1 - 1/z < \exp(-1/z)$, the second inequality follows from the fact that $1/k \leq 36\Gamma(i)\gamma(i)$ and the last ineqaulity follows by $Q_{d-1}(x)/x \leq 1$ and $x \leq \delta$. Lastly, by Fact 2.1 we have

$$\begin{aligned}
\left(1 - \frac{\delta}{k}\right)^{-300k\Phi(i\delta)\phi(i\delta)\Gamma(i)\gamma(i)} &= 1 + o(\Gamma(i)\gamma(i)), \text{ and} \\
\left(1 - \frac{\delta}{k}\right)^{-2k\Phi(i\delta)\phi(i\delta)\Gamma(i)} &\leq 1 + 2.5\Gamma(i)\gamma(i).
\end{aligned}$$

This implies the claim.

(ii) We have

$$
\begin{aligned}
& \Big(1 - \frac{x^2}{k^2} \cdot \frac{(1-8\Gamma(i)\gamma(i))^2 Q_{d-1}(x)^2}{x^2}\Big)^{|\Lambda_2(g,i)|} \\
\leq\ & \Big(1 - \frac{x^2}{k^2} \cdot \frac{(1-8\Gamma(i)\gamma(i))^2 Q_{d-1}(x)^2}{x^2}\Big)^{k^2\phi(i\delta)^2} \cdot \Big(1 - \frac{x^2}{k^2} \cdot \frac{(1-8\Gamma(i)\gamma(i))^2 Q_{d-1}(x)^2}{x^2}\Big)^{-k^2\phi(i\delta)^2\Gamma(i)} \\
\leq\ & \Big(1 - \frac{x^2}{k^2} \cdot \frac{(1-8\Gamma(i)\gamma(i))^2 Q_{d-1}(x)^2}{x^2}\Big)^{k^2\phi(i\delta)^2} \cdot \Big(1 - \frac{\delta^2}{k^2}\Big)^{-k^2\phi(i\delta)^2\Gamma(i)},
\end{aligned}
$$

where the first inequality is by $|\Lambda_2(g,i)| \geq k^2\phi(i\delta)^2(1-\Gamma(i))$ and the second inequality is by $(1-8\Gamma(i)\gamma(i))^2 Q_{d-1}(x)^2/x^2 \leq 1$ and $x \leq \delta$. We further have

$$
\begin{aligned}
& \Big(1 - \frac{x^2}{k^2} \cdot \frac{(1-8\Gamma(i)\gamma(i))^2 Q_{d-1}(x)^2}{x^2}\Big)^{k^2\phi(i\delta)^2} \\
\leq\ & \Big(1 - \frac{x^2}{k^2} \cdot \frac{Q_{d-1}(x)^2}{x^2}\Big)^{k^2\phi(i\delta)^2(1-36\Gamma(i)\gamma(i))(1-1/k)} \\
\leq\ & \Big(1 - \frac{x^2}{k^2} \cdot \frac{Q_{d-1}(x)^2}{x^2}\Big)^{k^2\phi(i\delta)^2(1-150\Gamma(i)\gamma(i))} \\
\leq\ & \Big(1 - \frac{x^2}{k^2} \cdot \frac{Q_{d-1}(x)^2}{x^2}\Big)^{k^2\phi(i\delta)^2} \cdot \Big(1 - \frac{\delta^2}{k^2}\Big)^{-150k^2\phi(i\delta)^2\Gamma(i)\gamma(i)},
\end{aligned}
$$

where the first inequality follows essentially from the fact that for every $z > 1$, $\exp(-1/(z-1)) < 1 - 1/z < \exp(-1/z)$, the second inequality follows from the fact that $1/k \leq 36\Gamma(i)\gamma(i)$ and the last ineqaulity follows by $Q_{d-1}(x)/x \leq 1$ and $x \leq \delta$. Lastly, by Fact 2.1 we note that

$$
\begin{aligned}
\Big(1 - \frac{\delta^2}{k^2}\Big)^{-150k^2\phi(i\delta)^2\Gamma(i)\gamma(i)} &= 1 + o(\Gamma(i)\gamma(i)), \text{ and} \\
\Big(1 - \frac{\delta^2}{k^2}\Big)^{-k^2\phi(i\delta)^2\Gamma(i)} &\leq 1 + 1.5\Gamma(i)\gamma(i).
\end{aligned}
$$

This implies the claim.

(iii) We have

$$
\begin{aligned}
& \Big(1 - \frac{x}{k} \cdot \frac{(1+8\Gamma(i)\gamma(i)) Q_{d-1}(x)}{x}\Big)^{|\Lambda_1(g,i)|} \\
\geq\ & \Big(1 - \frac{x}{k} \cdot \frac{(1+8\Gamma(i)\gamma(i)) Q_{d-1}(x)}{x}\Big)^{2k\Phi(i\delta)\phi(i\delta)} \cdot \Big(1 - \frac{x}{k} \cdot \frac{(1+8\Gamma(i)\gamma(i)) Q_{d-1}(x)}{x}\Big)^{2k\Phi(i\delta)\phi(i\delta)\Gamma(i)} \\
\geq\ & \Big(1 - \frac{x}{k} \cdot \frac{(1+8\Gamma(i)\gamma(i)) Q_{d-1}(x)}{x}\Big)^{2k\Phi(i\delta)\phi(i\delta)} \cdot \Big(1 - \frac{1.1\delta}{k}\Big)^{2k\Phi(i\delta)\phi(i\delta)\Gamma(i)},
\end{aligned}
$$

where the first inequality is by $|\Lambda_1(g,i)| \leq 2k\Phi(i\delta)\phi(i\delta)(1+\Gamma(i))$ and the second inequality is by $(1+8\Gamma(i)\gamma(i))Q_{d-1}(x)/x \leq 1.1$ and $x \leq \delta$. We further have

$$
\begin{aligned}
& \Big(1 - \frac{x}{k} \cdot \frac{(1+8\Gamma(i)\gamma(i)) Q_{d-1}(x)}{x}\Big)^{2k\Phi(i\delta)\phi(i\delta)} \\
\geq\ & \Big(1 - \frac{x}{k} \cdot \frac{Q_{d-1}(x)}{x}\Big)^{2k\Phi(i\delta)\phi(i\delta)(1+36\Gamma(i)\gamma(i))(1+1/k)} \\
\geq\ & \Big(1 - \frac{x}{k} \cdot \frac{Q_{d-1}(x)}{x}\Big)^{2k\Phi(i\delta)\phi(i\delta)(1+150\Gamma(i)\gamma(i))} \\
\geq\ & \Big(1 - \frac{x}{k} \cdot \frac{Q_{d-1}(x)}{x}\Big)^{2k\Phi(i\delta)\phi(i\delta)} \cdot \Big(1 - \frac{\delta}{k}\Big)^{300k\Phi(i\delta)\phi(i\delta)\Gamma(i)\gamma(i)},
\end{aligned}
$$

where the first inequality follows essentially from the fact that for every $z > 1$, $\exp(-1/(z-1)) < 1 - 1/z < \exp(-1/z)$, the second inequality follows from the fact that $1/k \leq 36\Gamma(i)\gamma(i)$ and the last ineqaulity follows by $Q_{d-1}(x)/x \leq 1$ and $x \leq \delta$. Lastly, by Fact 2.1 we have

$$\begin{aligned}\left(1-\frac{\delta}{k}\right)^{300k\Phi(i\delta)\phi(i\delta)\Gamma(i)\gamma(i)} &= 1 - o(\Gamma(i)\gamma(i)), \text{ and} \\ \left(1-\frac{1.1\delta}{k}\right)^{2k\Phi(i\delta)\phi(i\delta)\Gamma(i)} &\geq 1 - 2.5\Gamma(i)\gamma(i).\end{aligned}$$

This implies the claim.

(iv) We have

$$\begin{aligned}&\left(1-\frac{x^2}{k^2}\cdot\frac{(1+8\Gamma(i)\gamma(i))^2Q_{d-1}(x)^2}{x^2}\right)^{|\Lambda_2(g,i)|} \\ \geq\ &\left(1-\frac{x^2}{k^2}\cdot\frac{(1+8\Gamma(i)\gamma(i))^2Q_{d-1}(x)^2}{x^2}\right)^{k^2\phi(i\delta)^2}\cdot\left(1-\frac{x^2}{k^2}\cdot\frac{(1+8\Gamma(i)\gamma(i))^2Q_{d-1}(x)^2}{x^2}\right)^{k^2\phi(i\delta)^2\Gamma(i)} \\ \geq\ &\left(1-\frac{x^2}{k^2}\cdot\frac{(1+8\Gamma(i)\gamma(i))^2Q_{d-1}(x)^2}{x^2}\right)^{k^2\phi(i\delta)^2}\cdot\left(1-\frac{1.1\delta^2}{k^2}\right)^{k^2\phi(i\delta)^2\Gamma(i)},\end{aligned}$$

where the first inequality is by $|\Lambda_2(g,i)| \leq k^2\phi(i\delta)^2(1+\Gamma(i))$ and the second inequality is by $(1+8\Gamma(i)\gamma(i))^2Q_{d-1}(x)^2/x^2 \leq 1.1$ and $x \leq \delta$. We further have

$$\begin{aligned}&\left(1-\frac{x^2}{k^2}\cdot\frac{(1+8\Gamma(i)\gamma(i))^2Q_{d-1}(x)^2}{x^2}\right)^{k^2\phi(i\delta)^2} \\ \geq\ &\left(1-\frac{x^2}{k^2}\cdot\frac{Q_{d-1}(x)^2}{x^2}\right)^{k^2\phi(i\delta)^2(1+36\Gamma(i)\gamma(i))(1+1/k)} \\ \geq\ &\left(1-\frac{x^2}{k^2}\cdot\frac{Q_{d-1}(x)^2}{x^2}\right)^{k^2\phi(i\delta)^2(1+150\Gamma(i)\gamma(i))} \\ \geq\ &\left(1-\frac{x^2}{k^2}\cdot\frac{Q_{d-1}(x)^2}{x^2}\right)^{k^2\phi(i\delta)^2}\cdot\left(1-\frac{\delta^2}{k^2}\right)^{150k^2\phi(i\delta)^2\Gamma(i)\gamma(i)},\end{aligned}$$

where the first inequality follows essentialy from the fact that for every $z > 1$, $\exp(-1/(z-1)) < 1 - 1/z < \exp(-1/z)$, the second inequality follows from the fact that $1/k \leq 36\Gamma(i)\gamma(i)$ and the last ineqaulity follows by $Q_{d-1}(x)/x \leq 1$ and $x \leq \delta$. Lastly, by Fact 2.1 we note that

$$\begin{aligned}\left(1-\frac{\delta^2}{k^2}\right)^{150k^2\phi(i\delta)^2\Gamma(i)\gamma(i)} &= 1 - o(\Gamma(i)\gamma(i)), \text{ and} \\ \left(1-\frac{1.1\delta^2}{k^2}\right)^{k^2\phi(i\delta)^2\Gamma(i)} &\leq 1 - 1.5\Gamma(i)\gamma(i).\end{aligned}$$

This implies the claim.

■

Proof of Proposition 6.3. The proof is by induction on d. For the base case, $d = 0$, the assertion holds, since by definition $p_{g,0}(x) = q_0(x) = 1$ for all $x \in [0,\delta]$. Let $d > 0$ and assume that the proposition holds for d. Note that the assertion now holds for $x = 0$, since $p_{g,d}(0) = q_d(0) = 1$. Hence, it is enough to prove the proposition for $x \in (0,\delta]$, which we fix. The induction hypothesis implies

$$Q_{d-1}(x)(1-8\Gamma(i)\gamma(i)) \leq P_{g,d-1}(x) \leq Q_{d-1}(x)(1+8\Gamma(i)\gamma(i)).$$

By definition of survival and the structural properties of $T^*_{f,D}$ as implied by $\mathcal{E}_{\{f\}} \cap \mathcal{E}^*$, we have

$$p_{g,d}(x) = \prod_{G \in \Lambda(g,i)} \Big(1 - \prod_{g' \in G} \frac{x}{k} \cdot \frac{P_{g',d-1}(x)}{x}\Big). \tag{9}$$

By the induction hypothesis, (9) and Claim 6.4 (i) and (ii), we get

$$\begin{aligned} p_{g,d}(x) &\leq \Big(1 - \frac{Q_{d-1}(x)}{k}\Big)^{2k\Phi(i\delta)\phi(i\delta)} \Big(1 - \frac{Q_{d-1}(x)^2}{k^2}\Big)^{k^2\phi(i\delta)^2} (1 + 8\Gamma(i)\gamma(i)) \\ &= q_d(x)(1 + 8\Gamma(i)\gamma(i)). \end{aligned}$$

In addition, by the induction hypothesis, (9) and Claim 6.4 (iii) and (iv), we get

$$\begin{aligned} p_{g,d}(x) &\geq \Big(1 - \frac{Q_{d-1}(x)}{k}\Big)^{2k\Phi(i\delta)\phi(i\delta)} \Big(1 - \frac{Q_{d-1}(x)^2}{k^2}\Big)^{k^2\phi(i\delta)^2} (1 - 8\Gamma(i)\gamma(i)) \\ &= q_d(x)(1 - 8\Gamma(i)\gamma(i)). \end{aligned}$$

This completes the proof. ■

7 Proof of Theorem 3.2

Fix a triangle-free graph $F \subset K_n$ of size $O(1)$. Say that the triangle-free process *well-behaves* if for all $0 \leq i < I$, the preconditions in Lemma 3.3 hold and the events $\mathcal{E}^*$ and $\mathcal{E}^*_F$ as defined in Section 4.2 with respect to each i hold as well. Observe that the preconditions in the lemma hold for $i = 0$. Hence, by Lemma 3.3, the preconditions in the lemma hold for all $0 \leq i < I$ with probability $1 - o(1)$. Given that the preconditions hold for all $0 \leq i < I$, it follows from Lemma 4.2 that the events $\mathcal{E}^*$ and $\mathcal{E}^*_F$ both hold with probability $1 - o(1)$ for all $0 \leq i < I$. Hence it remains to prove Theorem 3.2 under the condition that the triangle-free process well-behaves, as we indeed do.

We begin the proof of Theorem 3.2 by proving it for the special case of $F = \{f\}$.

Proposition 7.1. *$Pr[f \in \mathbb{TF}(n, I) \mid f \in \mathbb{B}_{\leq I}] \sim \frac{\Phi(I\delta)}{n^{\varepsilon}}$.*

Proof. First we estimate the probability of $\{f \in \mathbb{TF}(n, I) \mid f \in \mathbb{B}_{i+1}\}$ for any $0 \leq i < I$. Since the triangle-free process well-behaves, we can apply the first two items in Lemma 3.3 for any $0 \leq i < I$. By the first item in Lemma 3.3 and by Fact 2.1, for $0 \leq i < I$,

$$\Pr[f \in \mathbb{TF}(n, i+1) \mid f \in \mathbb{B}_{i+1}, \mathbb{TF}(n, i) \cup \{f\} \text{ is triangle-free}] \sim \frac{\Phi((i+1)\delta) - \Phi(i\delta)}{\phi(i\delta)\delta}.$$

In addition, by the second item in Lemma 3.3 and by Fact 2.1, one can verify that for $0 \leq i < I$,

$$\Pr[\mathbb{TF}(n, i) \cup \{f\} \text{ is triangle-free} \mid f \notin \mathbb{B}_{\leq i}] \sim \phi(i\delta).$$

Hence,

$$\Pr[f \in \mathbb{TF}(n, i+1) \mid f \in \mathbb{B}_{i+1}] \sim \frac{\Phi((i+1)\delta) - \Phi(i\delta)}{\delta}.$$

Therefore,

$$\begin{aligned}
\Pr[f \in \mathbb{TF}(n, I) \,|\, f \in \mathbb{B}_{\leq I}] &= \sum_{0 \leq i < I} \Pr[f \in \mathbb{B}_{i+1} \,|\, f \in \mathbb{B}_{\leq I}] \Pr[f \in \mathbb{TF}(n, i+1) \,|\, f \in \mathbb{B}_{i+1}] \\
&\sim \frac{\delta}{n^{\varepsilon}} \sum_{0 \leq i < I} \Pr[f \in \mathbb{TF}(n, i+1) \,|\, f \in \mathbb{B}_{i+1}] \\
&\sim \frac{1}{n^{\varepsilon}} \sum_{0 \leq i < I} \Phi((i+1)\delta) - \Phi(i\delta) \\
&= \frac{\Phi(I\delta)}{n^{\varepsilon}}.
\end{aligned}$$

■

Note that since the triangle-free process well-behaves, we could have proved Proposition 7.1 using the two items in Lemma 4.3 instead of the first two items in Lemma 3.3. Indeed, under the condition that the triangle-free process well-behaves, the proof of the first two items in Lemma 3.3 follow directly from Lemma 4.3. In other words, the probability of the event $\{f \in \mathbb{TF}(n, I) \,|\, f \in \mathbb{B}_{\leq I}\}$ was essentially estimated in the proof of Proposition 7.1 by the probabilities of the events $\{\mathcal{A}^*_{f,D} \,|\, f \in \mathbb{B}_{i+1}\}$ and $\{\mathcal{A}^*_{f,D} \,|\, f \notin \mathbb{B}_{\leq i+1}\}$ as defined for each $0 \leq i < I$ in Section 4.2. Now observe that for all $0 \leq i < I$, since $\mathcal{E}^*_F$ holds, we have that $\{\mathcal{A}^*_{f,D} \,|\, f \notin \mathbb{B}_{\leq i}\}$ is independent from $\{\mathcal{A}^*_{f',D} \,|\, f' \notin \mathbb{B}_{\leq i}\}$. Therefore, it follows from Proposition 7.1 that

$$\Pr[F \subseteq \mathbb{TF}(n, I) \,|\, F \subseteq \mathbb{B}_{\leq I}] \sim \left(\frac{\Phi(I\delta)}{n^{\varepsilon}}\right)^{e_F}.$$

8 Concluding remarks

- In Section 6 we have defined the tree T_∞ so that the number of children of the root (say) of T_∞ that are of size 1 is exactly $\lceil 2k\Phi(i\delta)\phi(i\delta) \rceil$. We further made the simplifying assumption that $2k\Phi(i\delta)\phi(i\delta)$ is an integer. Here we briefly outline a modified argument to the one given in Section 6 for the case in which $2k\Phi(i\delta)\phi(i\delta)$ is not an integer. The basic idea, as described below, is to take a random subtree of $T^*_{f,D}$ at the beginning of Section 6, to adjust appropriately the distribution of the birthtimes of the edges that appear in that random subtree and to redefine the tree T_∞ according to the properties of the random subtree.

  To describe the random subtree, let $\zeta \in [0.1, 0.9]$ be such that $\zeta \cdot 2k\Phi(i\delta)\phi(i\delta)$ is an integer. Keep every subtree of $T^*_{f,D}$ that is rooted by a set of size 1 with probability ζ and keep every subtree of $T^*_{f,D}$ that is rooted by a set of size 2 with probability 1. This gives us a random subtree of $T^*_{f,D}$. One can show, given $\mathcal{E}^*$, which we assume to hold at the beginning of Section 6, the following. With probability $1-n^{-\omega(1)}$, for every non-leaf node v at even distance from the root of $T^*_{f,D}$, the number of children of v that are sets of size 1 is $\zeta \cdot 2k\Phi(i\delta)\phi(i\delta)(1 \pm 1.25\Gamma(i))$ and the number of children of v that are sets of size 2 is as implied by $\mathcal{E}^*$. Since we condition on $\mathcal{E}^*_{\{f\}}$, we refer to the nodes of $T^*_{f,D}$ below by their labels. By $\mathcal{E}^*_{\{f\}}$ we can also redefine the birthtimes of the edges that are nodes in $T^*_{f,D}$ as follows. The birthtime of an edge that appears in a set of size 1 in $T^*_{f,D}$ is redefined so that it is uniformly distributed in

$[0, \zeta \cdot kn^{-1/2}]$, whereas the birthtime of an edge that appears in a set of size 2 in $T^*_{f,D}$ remains uniformly distributed in $[0, kn^{-1/2}]$. Given that, we define an infinite tree T_∞ as before, only that now we take $\zeta \cdot 2k\Phi(i\delta)\phi(i\delta)$ instead of $2k\Phi(i\delta)\phi(i\delta)$. Furthermore, every node in T_∞ which is a child of a set of size 1 has a birthtime that is distributed uniformly at random in $[0, \zeta \cdot kn^{-1/2}]$, whereas every node in T_∞ which is a child of a set of size 2 has, as before, a birthtime that is distributed uniformly at random in $[0, kn^{-1/2}]$. The rest of the argument is essentially the same as the one presented in Section 6.

- Consider the following random greedy algorithm for constructing a matching in a given hypergraph H. Start by ordering the edges of H uniformly at random. Then, traverse the ordered edges and add each traversed edge to an evolving set (which is initially empty), unless the addition of the edge creates a set which is not a matching. Let $\mathbb{M}(H)$ denote the matching created by the algorithm. One can show using a similar argument to the one presented above that if H is a k-uniform, d-regular, simple hypergraph on n vertices then the number of vertices not covered by $\mathbb{M}(H)$ is $O\big(\frac{n(\ln n)^{O(1)}}{d^{1/(k-1)}}\big)$. This confirms a conjecture of Alon, Kim and Spencer [Israel J. Math. 100 (1997)]. This is the subject of a forthcoming paper.

- Coupled with an idea from [12], we can use the same argument presented in this paper in order to provide an upper bound on the length of the K_4-free process, a bound which matches, up to a constant factor, the recent lower bound provided by Bohman [1]. Namely, we can show that the expected number of edges that are accepted by the K_4-free process is $O(n^{8/5}(\ln n)^{1/5})$. This again is the subject of a forthcoming paper.

References

[1] T. Bohman, *The triangle-free process* (2008), available at `arXiv:0806.4375v1`.

[2] B. Bollobás and O. Riordan, *Constrained graph processes*, Electr. J. Comb. **7** (2000).

[3] W. G. Brown, *On graphs that do not contain a Thomsen graph*, Canad. Math. Bull. **9** (1966), 281–285.

[4] P. Erdős, A. Rényi, and V. T. Sós, *On a problem of graph theory*, Studia Sci. Acad. Math. Hungar. **1** (1966), 215–235.

[5] P. Erdős and J. H. Spencer, *Probabilistic methods in combinatorics*, 1974.

[6] P. Erdős, S. Suen, and P. Winkler, *On the size of a random maximal graph*, Random Struct. Algorithms **6** (1995), no. 2/3, 309–318.

[7] S. Janson, T. Luczak, and A. Ruciński, *Random graphs*, 2000.

[8] S. Janson, *Poisson approximation for large deviations*, Random Struct. Algorithms **1** (1990), no. 2, 221–230.

[9] D. Osthus and A. Taraz, *Random maximal h-free graphs*, Random Struct. Algorithms **18** (2001), no. 1, 61–82.

[10] A. Rucinski and N. C. Wormald, *Random graph processes with degree restrictions*, Combinatorics, Probability & Computing **1** (1992), 169–180.

[11] J. H. Spencer, *Threshold functions for extension statements*, J. Comb. Theory, Ser. A **53** (1990), no. 2, 286–305.

[12] J. H. Spencer, *Maximal triangle-free graphs and Ramsey $r(3,t)$*, unpublished manuscript (1995).