Random $K_k$-removal algorithm *

Fang Tian$^{1\dagger}$ Zi-Long Liu$^2$ Xiang-Feng Pan$^3$

$^1$ Department of Applied Mathematics
Shanghai University of Finance and Economics, Shanghai, 200433, China
tianf@mail.shufe.edu.cn

$^2$School of Optical-Electrical and Computer Engineering
University of Shanghai for Science and Technology, Shanghai, 200093, China
liuzl@usst.edu.cn

$^3$School of Mathematical Sciences
Anhui University, Hefei, Anhui, 230601, China
xfpan@ahu.edu.cn

Abstract

One interesting question is how a graph develops from some constrained random graph process, which is a fundamental mechanism in the formation and evolution of dynamic networks. The problem here is referred to the random $K_k$-removal algorithm. For a fixed integer $k \geq 3$, it starts with a complete graph on $n \to \infty$ vertices and iteratively removes the edges of an uniformly chosen $K_k$. This algorithm terminates once no $K_k$s remain and at the same time it generates one linear $k$-uniform hypergraph. For $k = 3$, it was shown that the size in the final graph is $n^{3/2+o(1)}$. Less results are on the cases when $k \geq 4$. In this paper, we prove that the exact expected trajectories of various key parameters in the algorithm to some iteration such that the final size in the algorithm is at most $n^{2-1/(k(k-1)-2)+o(1)}$ for $k \geq 4$. We also show the bound is a natural barrier.

Keywords: random greedy algorithm, $K_k$-free, the critical interval method, dynamic concentration.

Mathematics Subject Classifications: 05D40, 68R10

1 Introduction

Extremal problems are central research issues in random graph algorithms, which are also fundamental mechanisms in the formation and evolution of dynamic networks. A better understanding of the underlying graph offers us opportunities to study how a graph develops from some constrained random greedy process. Recently, the power of random greedy algorithm is illustrated in [9] by showing the existence of mathematical objects with better properties. Each time random greedy algorithms go beyond classical applications of the probabilistic method used in previous work.

*The work was partially supported by NSFC.
$\dagger$Corresponding Author: tianf@mail.shufe.edu.cn(Email Address).
The problem here is referred to the random $K_k$-removal algorithm. Given a fixed integer $k \geq 3$, the random $K_k$-removal algorithm for generating one $K_k$-free graph, and at the same time creating a linear $k$-uniform hypergraph, is defined as follows. Start from a complete graph on vertex set $[n]$, denoted by $G(0)$, and $G(i + 1)$ is the remaining graph from $G(i)$ by selecting one $K_k$ uniformly at random out of all $K_k$'s in $G(i)$ and deleting all its edges. Let the hitting time $M$ be $M = \min \{i : G(i)$ is $K_k$-free$\}$ and $E(i)$ denote the edge set of $G(i)$, thus $|E(M)|$ is the number of edges in the final $K_k$-free graph.

Work on finding the exact values of $|E(M)|$ has evolved over the past 20 years and is a nontrivial task even for $k = 3$. Bollobás and Erdős [6] conjectured that with high probability $|E(M)| = n^{3/2+o(1)}$ when $k = 3$. It was shown $|E(M)| = o(n^2)$ by Spencer [13] and independently by Rödl and Thoma [12]. Grable [8] improved this bound to $|E(M)| \leq n^{7/4+o(1)}$. Bohman et al. [3] introduced the critical interval method for proving dynamic concentrations. They [4] confirmed the exponent in a breakthrough by generalizing the approach in [3]. Less results directly studied the random $K_k$-removal algorithms when $k \geq 4$. Bennett and Bohman [1] conjectured that $|E(M)| \leq n^{2k/(k+1)+o(1)}$ as a folklore for $k \geq 3$ when they investigated the random greedy hypergraph matching algorithm. It is exactly the one proposed by Bollobás and Erdős when $k = 3$.

A different recipe for obtaining a random $K_k$-free graph is the so-called “$K_k$-free process”. In that algorithm, starting with an empty graph, the $\binom{n}{2}$ edges are randomly inserted so long as no $K_k$'s are formed in the current graph. Despite the high similarity between the two protocols, it was shown [4] that the random $K_k$-removal algorithm has proved quite challenging at the level of acquiring the correct exponent of the final number of edges. A pseudo-random heuristic for divining the evolution of various key parameters plays a central role in the understanding of these algorithms that produce interesting combinatorial objects [1–5, 11, 14].

In this paper, we directly discuss the structure of random $K_k$-removal algorithm for $k \geq 4$. We design an ensemble of appropriate random variables including the number of $K_k$'s, using a heuristic assumption to find the trajectories of these variables when the process evolves. Compared with the random $K_3$-removal algorithm, it is challenging to make use of these auxiliary variables to analyze the one-step change of the number of $K_k$'s when $k \geq 4$ and show a rigorous proof of their expressions. At last, we verified that

**Theorem 1.1.** Given a fixed integer $k \geq 4$, consider the random $K_k$-removal algorithm on $n$ vertices. Let $M$ be the number of steps it takes the process to terminate and $E(M)$ be the size of the resulting $K_k$-free graph. With high probability, $|E(M)| \leq n^{2-1/(k(k+1)-2)+o(1)}$.

Though our bound exists a gap with $|E(M)| \leq n^{2k/(k+1)+o(1)}$ conjectured in [1], we will show our result corresponds to the inherent barrier of the algorithm.

The remainder of this paper is organized as follows. In the next section, notations and some lemmas for analyzing the random $K_k$-removal algorithm are presented. In Section 3, we discuss the evolution of the algorithm in detail and estimate the trajectories of these random variables. We formally prove the concentrations in Section 4.
2 Notations and Some Lemmas

Let \((\Omega, \mathcal{F}, \mathbb{P})\) be an arbitrary probability space. Note that our probability space is the set of all maximal sequences of edge-disjoint \(K_k\)s on vertex set \([n]\) with probability measure given by the uniform random choice at each step. Let \((\mathcal{F}_i)_{i \geq 0}\) be the filtration given by the evolutionary algorithm.

Given a sequence of random variables \(X_i\), let \(\Delta X = X_{i+1} - X_i\) denote the one-step change for the random variables \(X_i\) and the pair \(\{X_i, \mathcal{F}_i\}_{i \geq 0}\) is then called a submartingale (resp. supermartingale) if \(X_i\) is \(\mathcal{F}_i\)-measurable and \(\mathbb{E}[\Delta X | \mathcal{F}_i] \geq 0\) (resp. \(\mathbb{E}[\Delta X | \mathcal{F}_i] \leq 0\)) for all \(i \geq 0\). An event is said to occur with high probability (w.h.p. for short), if the probability that it holds tends to 1 when \(n \to \infty\). Furthermore, for two positive-valued functions \(f, g\) on the variable \(n\), we write \(f \ll g\) to denote \(\lim_{n \to \infty} f(n)/g(n) = 0\) and \(f \sim g\) to denote \(\lim_{n \to \infty} f(n)/g(n) = 1\). Let \(a = b \pm c\) be short for \(a \in [b - c, b + c]\). \(\binom{S}{b} = 0\) and \(\binom{a}{b} = 0\) if \(b > |S|\) and \(b > a\). We also use the standard asymptotic notation \(o, O, \Omega\) and \(\Theta\). All logarithms are natural, and the floor and ceiling signs are omitted whenever they are not crucial. Throughout the following sections we assume that \(n \to \infty\).

For \(2 \leq m \leq k, u \in [n]\) and \(U_m = \{u_1, \cdots, u_m\} \in \binom{[n]}{m}\), let \(N_u = N_u(i) = \{x \in [n] : xu \in E(i)\}\), \(N_{U_m} = N_{U_m}(i) = \cap_{u \in U_m} N_u\) and \(K_m(i)\) be the set of complete graph \(K_m\) in \(G(i)\). Our goal is to estimate the number of \(K_m\)s in \(G(i)\), that is \(|K_m(i)|\), which is particularly denoted by \(Q_k(i)\). Define the random variable \(R_{k, U_m}(i)\) to be

\[
R_{k, U_m}(i) = \begin{cases} 
|K_{k-m} \cap \binom{N_{U_m}}{k-m}|, & 2 \leq m \leq k-1; \\
1_{U_k}, & m = k. 
\end{cases}
\]  (2.1)

For \(2 \leq m \leq k-1\), \(R_{k, U_m}(i)\) counts the number of \(K_{k-m}\)s in \(G(i)\) such that every vertex in \(K_{k-m}\) is in \(N_{U_m}\); particularly \(R_{k, U_{k-1}}(i) = |N_{U_{k-1}}|\) is the codegree of the vertex subset \(U_{k-1}\). \(1_{U_k}\) is the indicator random variable with \(1_{U_k} = 1\) if the subgraph induced by \(U_k\) in \(G(i)\) is complete, instead \(1_{U_k} = 0\) otherwise. Bennett et al. [1] ever added more assumptions on codegrees of larger vertex subsets to obtain stronger results on random greedy hypergraph matching algorithm. Sometimes for shorthand we will suppress \(i\). These random variables in (2.1) yield important information about the underlying process.

Suppose that the vertex set of the \((i+1)\)-th taken \(K_k\) is denoted by \(U_k\). Let \(U_m \in \binom{U_k}{m}\) with \(2 \leq m \leq k\) and

\[
Q^U_k(i) = |\{K_k \in G(i) | K_k \cap U_k = U_m\}|,
\]

namely, \(Q^U_k(i)\) denotes the number of \(K_k\)s in \(G(i)\) that exactly contains the vertices \(U_m\) in \(U_k\). In particular, \(Q^U_k(i) = 1\). Thus, we have

\[
Q_k(i) - Q_k(i+1) = \sum_{m=2}^{k} \left( \sum_{U_m \in \binom{U_k}{m}} Q^U_k(i) \right).
\]  (2.2)

It is observed that \(R_{k, U_m}\) in (2.1) denotes the number of extensions to one copy of \(K_k\) from \(U_m\)
Thus, the expectation $\mathbb{E} \[ \sum \]_{T_1 \in \left( U_k \setminus U_m \right)} (-1)^1 \mathbf{R}_{k, U_m \cup T_1} + \cdots + \sum \mathbf{R}_{k, U_m \cup T_{k-m-1}} + (-1)^{k-m} \mathbf{R}_{k, U_k}$.  \hspace{1cm} \text{(2.3)}$

Note that
\[ \sum_{U_m \in\left( \begin{smallmatrix} U_k \\ m \end{smallmatrix} \right)} \left( \sum_{T_i \in \left( U_k \setminus U_m \right)} \mathbf{R}_{k, U_m \cup T_i} \right) = \left( \frac{m+i}{m} \right) \sum_{U_m+i \in \left( \begin{smallmatrix} U_k \\ m+i \end{smallmatrix} \right)} \mathbf{R}_{k, U_m+i} \]
for $0 \leq i \leq k - m$ because each element $\mathbf{R}_{k, U_m+i}$ on the right side is counted $\left( \frac{m+i}{m} \right)$ times on the left side. Sum the above corresponding displays (2.3) for all $U_m \in\left( \begin{smallmatrix} U_k \\ m \end{smallmatrix} \right)$ with $2 \leq m \leq k$ altogether into the equation (2.2), then it follows that
\[ \mathbf{Q}_k(i) - \mathbf{Q}_k(i+1) = \sum_{U_2 \in \left( \begin{smallmatrix} U_k \\ 2 \end{smallmatrix} \right)} \mathbf{R}_{k, U_2} + \sum_{U_3 \in \left( \begin{smallmatrix} U_k \\ 3 \end{smallmatrix} \right)} (-1)^1 \left( \begin{smallmatrix} 3 \\ 2 \end{smallmatrix} \right) + (-1)^0 \left( \begin{smallmatrix} 3 \\ 3 \end{smallmatrix} \right) \mathbf{R}_{k, U_3} + \cdots \]
\[ + \sum_{U_{k-1} \in \left( \begin{smallmatrix} U_k \setminus U_m \end{smallmatrix} \right)} (-1)^{k-3} \left( \begin{smallmatrix} k-1 \\ 2 \end{smallmatrix} \right) + \cdots + (-1)^0 \left( \begin{smallmatrix} k-1 \\ k-1 \end{smallmatrix} \right) \mathbf{R}_{k, U_{k-1}} \]
\[ + (-1)^{k-2} \left( \begin{smallmatrix} k \\ 2 \end{smallmatrix} \right) + \cdots + (-1)^0 \left( \begin{smallmatrix} k \\ k \end{smallmatrix} \right) \mathbf{R}_{k, U_k}. \]

Since $\sum_{j=2}^r (-1)^{r-j} \binom{r}{j} = (-1)^r (r - 1)$ for any given integer $r \geq 2$ and $\mathbf{R}_{k, U_k} = 1$ in (2.1),
\[ \mathbf{Q}_k(i) - \mathbf{Q}_k(i+1) = \sum_{U_2 \in \left( \begin{smallmatrix} U_k \\ 2 \end{smallmatrix} \right)} \mathbf{R}_{k, U_2} - \sum_{U_3 \in \left( \begin{smallmatrix} U_k \\ 3 \end{smallmatrix} \right)} \mathbf{R}_{k, U_3} + \cdots \]
\[ + (-1)^{k-1} (k - 2) \sum_{U_{k-1} \in \left( \begin{smallmatrix} U_k \setminus U_{k-1} \end{smallmatrix} \right)} \mathbf{R}_{k, U_{k-1}} + (-1)^k (k - 1). \hspace{1cm} \text{(2.4)} \]

Thus, the expectation $\mathbb{E} [\Delta \mathbf{Q}_k | \mathcal{F}_i]$ of $\Delta \mathbf{Q}_k$ is
\[ \mathbb{E} [\Delta \mathbf{Q}_k | \mathcal{F}_i] = \sum_{U_k \in \mathcal{K}_k(i)} \frac{\sum_{U_2 \in \left( \begin{smallmatrix} U_k \\ 2 \end{smallmatrix} \right)} \mathbf{R}_{k, U_2} + \cdots + (-1)^{k-1} (k - 2) \sum_{U_{k-1} \in \left( \begin{smallmatrix} U_k \setminus U_{k-1} \end{smallmatrix} \right)} \mathbf{R}_{k, U_{k-1}} + (-1)^k (k - 1)}{\mathbf{Q}_k(i)} \]
\[ = (-1)^{k+1} (k - 1) - \frac{1}{\mathbf{Q}_k(i)} \sum_{U_2 \in \mathcal{K}_2(i)} \mathbf{R}_{k, U_2}^2 + \cdots + \frac{(-1)^k (k - 2)}{\mathbf{Q}_k(i)} \sum_{U_{k-1} \in \mathcal{K}_{k-1}(i)} \mathbf{R}_{k, U_{k-1}}^2, \hspace{1cm} \text{(2.5)} \]
where the last equality is true because
\[ \sum_{U_k \in \mathcal{K}_k(i)} \sum_{U_m \in \left( \begin{smallmatrix} U_k \\ m \end{smallmatrix} \right)} \mathbf{R}_{k, U_m} = \sum_{U_m \in \mathcal{K}_m(i)} \mathbf{R}_{k, U_m}^2 \]
for $2 \leq m \leq k - 1$ by double counting.

We also need the following lemmas to establish dynamic concentrations on variables $Q_k(i)$ and $R_{k,U_m}$ for any $U_m \in \binom{[n]}{m}$ with $2 \leq m \leq k - 1$, which were also used in [1-5, 11, 14].

**Lemma 2.1** (Bohman et al. [4]). Let $a_1, \cdots, a_\ell \in \mathbb{R}$ and some $a \in \mathbb{R}$. Suppose that $|a_i - a| \leq \varepsilon$ for all $1 \leq i \leq \ell$, then $\frac{(\sum_{i=1}^\ell a_i)^2}{\ell} \leq \sum_{i=1}^\ell a_i^2 \leq \frac{(\sum_{i=1}^\ell a_i)^2}{\ell} + 4\ell\varepsilon^2$.

**Lemma 2.2** (Hoeffding and Azuma [10]). Suppose a sequence of random variables $\{X_i\}_{i \geq 0}$ is a supermartingale (resp. submartingale) and $|X_i - X_{i-1}| < c_i$, then for any positive integer $\ell$ and any positive real number $a$, $\mathbb{P}[X_\ell - X_0 \geq a] \leq \exp\left[-\frac{a^2}{2\sum_{i=1}^\ell c_i}\right]$. (resp. $\mathbb{P}[X_\ell - X_0 \leq -a] \leq \exp\left[-\frac{a^2}{2\sum_{i=1}^\ell c_i}\right]$)

Let $\eta, N > 0$ be constants. A sequence of random variables $\{X_i\}_{i \geq 0}$ is $(\eta, N)$-bounded if $X_i - \eta \leq X_{i+1} \leq X_i + N$ for all $i \geq 0$. For $(\eta, N)$-bounded supermartingales and submartingales, Bohman [2] showed that

**Lemma 2.3** (Bohman [2]). Suppose $\{X_i\}_{i \geq 0}$ is an $(\eta, N)$-bounded supermartingale (resp. submartingale) with initial value 0 and $\eta \leq \frac{N}{\ell}$. Then for any positive integer $\ell$ and any positive real number $a$ with $a < \eta\ell$, $\mathbb{P}[X_\ell \geq a] \leq \exp\left[-\frac{a^2}{3\eta\ell N}\right]$. (resp. $\mathbb{P}[X_\ell \leq -a] \leq \exp\left[-\frac{a^2}{3\eta\ell N}\right]$)

Finally, in order to explain it is definitely possible to further improve our results. the lemma below in [7] is also required.

**Lemma 2.4** ([7]). For $X \sim \text{Bin}(n,p)$ and any $0 < \xi \leq np$, $\mathbb{P}[|X - np| > \xi] < 2 \exp\left[-\xi^2 / (3np)\right]$.

## 3 Estimates on the variables in $G(i)$

In the following, we use some heuristics to anticipate the likely values of the auxiliary random variables throughout the process. We assume the random $K_k$-removal algorithm produces a graph whose variables are roughly the same as they would be in a random graph $G(n,p)$ with the same edge density. The classical Erdős-Rényi random graph $G(n,p)$ is on vertex set $[n] = \{1, \cdots, n\}$ and any two vertices appear as an edge independently with probability $p$.

In order to describe the expected trajectories of $Q_k(i)$ and $R_{k,U_m}$ as smooth functions for any $U_m = \{u_1, u_2, \cdots, u_m\} \in \binom{[n]}{m}$ with $2 \leq m \leq k - 1$, we appropriately rescale the number of steps $i$ to be $t = t(i) = \frac{i}{n^2}$ and introduce a notion of edge density as

$$p = p(i, n) = 1 - \frac{k(k-1)i}{n^2} = 1 - k(k-1)t. \quad (3.1)$$

Note that $p$ can be viewed as either a continuous function of $t$ or as a function of the discrete variable $i$. We pass between these interpretations without comment. With this notation, we have

$$|E(i)| = \binom{n}{2} - \binom{k}{2} i = \binom{n}{2} - \frac{1}{2}(1-p)n^2 = \frac{1}{2}(n^2p - n) \quad (3.2)$$

such that the number of edges in $G(i)$ with edge density $p$ is approximately equal to the one in the Erdős-Rényi graph $G(n,p)$ up to the negligible linear term when $p$ lies in some range.
For a fixed integer $k \geq 4$, $2 \leq m \leq k - 1$ and $U_m \in \binom{[m]}{m}$, under the assumption that $G(i)$ resembles $G(n,p)$, we anticipate that the expressions of $Q_k(i)$ and $R_{k,U_m}$ are

$$Q_k(i) \sim \frac{n^k}{k!} p_{G(n,p)}^{(k)}$$

and

$$R_{k,U_m} \sim \frac{n^{k-m}}{(k-m)!} p_{G(n,p)}^{(k)-(m)}$$

where $\frac{n^k}{k!} p_{G(n,p)}^{(k)}$ counts the expected number of $K_k$s in $G(n,p)$; $(n-m)(k-m)! p_{G(n,p)}^{(k)-(m)}$ counts the expected number of $K_{k-m}$s in which every vertex is in $N_{U_m}$. Our main theorem is as follows:

**Theorem 3.1.** Given a fixed integer $k \geq 4$, let $U_m \in \binom{[m]}{m}$ with $2 \leq m \leq k - 1$, then there exist absolute constants $\mu$, $\gamma_m$ and $\lambda$ such that, with high probability,

$$Q_k(i) \leq \frac{n^k}{k!} p_{G(n,p)}^{(k)} + \frac{n^{k-1}}{2} p_{G(n,p)}^{(k)-4}, \quad (3.3)$$

$$Q_k(i) \geq \frac{n^k}{k!} p_{G(n,p)}^{(k)} - \sigma^2 n^\alpha p^{-1} \log^\mu n, \quad (3.4)$$

$$R_{k,U_m} = \frac{n^{k-m}}{(k-m)!} p_{G(n,p)}^{(k)-(m)} \pm \sigma n^{\beta_m} \log^{\gamma_m} n \quad (3.5)$$

holding for every $i \leq i_0$ with $i_0 = \frac{n^2}{k(k-1)} - \frac{\sqrt{2}}{k(k-1)} n^2 \frac{1}{k(k-1)^2} \log^2 n$, where

$$\alpha = k - \frac{k+1}{2(k-2)}, \quad (3.6)$$

$$\beta_m = k - m - \frac{k+1}{2(k-2)} \quad (3.7)$$

and the error function $\sigma = \sigma(t)$ is taken with initial value $\sigma(0) = 1$ that slowly grows to be

$$\sigma = \sigma(t) = 1 - \frac{k(k-1)}{4} \log p(t). \quad (3.8)$$

Theorem 3.1 is proved in Section 4. It implies that for these specific choices of constants satisfying the equations in (3.6) and (3.7), and the error function $\sigma$ in (3.8), these random variables are around the heuristical trajectories to the stopping time $\tau = i_0$ with high probability. These dynamic concentrations in turn show that the algorithm produces a graph of size at most $|E(i_0)|$ with high probability. We make no attempt to optimize the constants $\mu$, $\lambda$ and $\gamma_m$ in all error terms with $2 \leq m \leq k - 1$. There are many choices of them that can be balanced to satisfy certain inequalities, such as $[\frac{k+1}{2}] \lambda > \mu + 2, [\frac{k-1}{2}] \lambda > \gamma_m + 1$ with $2 \leq m \leq k - 1$, and $\gamma_2 > \frac{1}{2}$, can support our analysis of Theorem 3.1. We do not replace them with their actual values. This is for the interest of understanding the role of these constants played in the calculations.

**Proof of Theorem 1.1.** We recover the number of edges when $p = p_0$ to be

$$|E(i_0)| = \frac{n^2}{2} - \binom{k}{2} i_0 \sim \frac{\sqrt{2}}{2} n^2 \frac{1}{k(k-1)^2} \log^2 n.$$

Theorem 1.1 follows directly from Theorem 3.1 by $|E(M)| \leq |E(i_0)|$ with room to sparse in the power of the logarithmic factor.
Remark 3.2. The variation equations in (3.3)-(3.5) are verified in a straightforward manner below. According to (3.1), define

\[ p_0 = p(i_0, n) = 1 - \frac{k(k-1)i_0}{n^2} = \sqrt{2}n - \frac{1}{k(k-1) - 2} \log \lambda n. \] (3.9)

Since \( i \leq i_0 \) in Theorem 3.1, we have \( p \geq p_0 \) in (3.9). Note that \( \frac{k^i}{k!}p^{(k)}_0 \gg \sigma^2 n^{\alpha-1}p^{-1} \log^\mu n \) when \( \alpha \) is in (3.6), and appropriate choices of \( \lambda \) and \( \mu \). It follows that \( Q_k(i) = (1 + o(1))n^{k}p^{(k)}_0/k! \) in (3.3) and (3.4). Similarly all the error terms in (3.5) are negligible compared to their respective corresponding main terms.

Remark 3.3. Our bound in Theorem 1.1 exists a gap with \( |E(M)| \leq n^{2k/(k+1) + o(1)} \) conjectured in [1]. In fact, the term \( n^{2-1/(k(k-1))} \) corresponds to a natural barrier in the random \( K_k \)-removal algorithm. To illustrate this, as stated in Theorem 3.1, we know \( G(i) \) is roughly the same with \( G(n,p) \), while we notice that the standard variations of \( R_k,U_m \) for any \( U_m \in \binom{n}{m} \) with \( 2 \leq m \leq k-1 \) would be as large as their main trajectories when \( p \) is around \( n^{-1/(k(k-1))} \) (up to logarithmic factors), which means that the control over \( R_k,U_m \) for any \( U_m \in \binom{n}{m} \) is lost.

Remark 3.4. As stated in Theorem 3.1, we know \( G(i) \) is roughly the same with \( G(n,p) \) for \( i \leq i_0 \). Thus, when \( p \) is around \( p_0 \) in (3.9), by a union bound, it follows that the probability that there exists one \( U_m \in \binom{[m]}{m} \) with \( 2 \leq m \leq k-1 \) such that \( |R_k,U_m - \frac{n}{(k-m)!} \binom{k}{m} \binom{m}{2} p_0| > \sigma n^{\beta_m} \log^\gamma m \) is at most

\[
\sum_{U_m \in \binom{[m]}{m}, 2 \leq m \leq k-1} \Pr \left[ \left| R_k,U_m - \frac{n^{k-m}}{(k-m)!} \binom{k}{m} \binom{m}{2} p_0 \right| > \sigma n^{\beta_m} \log^\gamma m \right] < 2 \sum_{m=2}^{k-1} \binom{n}{m} \exp \left[ - \frac{(k-m)! \sigma n^{\beta_m} \log^\gamma m n^2}{3 n^{k-m} \binom{k}{2} \binom{m}{2}} \right]
\]

\[
= \sum_{m=2}^{k-1} \binom{n}{m} \exp \left[ - \Theta \left( \frac{k-m}{2} \binom{m}{2} \right) \right]. \quad (3.10)
\]

by applying Lemma 2.4 with \( \xi = \sigma n^{\beta_m} \log^\gamma m n \), where the last equality is true because \( \beta_m \) is in (3.7). Since the summand in (3.10) is increasing in \( m \) for fixed \( k \geq 4 \), it suffices to take the number of terms times the last term when \( m = k-1 \). Thus, we have

\[
\sum_{m=2}^{k-1} \binom{n}{m} \exp \left[ - \Theta \left( \frac{k-m-\frac{1}{2}}{2} \binom{m}{2} \right) \right] = O \left( n^{k-1} \exp \left[ - \Theta \left( \frac{1-\frac{1}{2}}{2} \binom{k}{2} \right) \right] \right) = o(1).
\]

In fact, we could show the similar phenomenon even when we take \( \xi = \Theta(n^\theta) \) with \( \frac{1}{2} \beta_m < \theta < \beta_m \), instead our main results in Theorem 3.1 cannot support us. Like [4], in order to prove better bounds on \( |E(M)| \), it is possible to design new random variables such that their variations decrease as the process evolves.
4 Proof of Theorem 3.1

Recall the outline of the critical interval method \cite{1, 3, 4} to control some graph parameters when the process evolves. Let the stopping time \( \tau \) be the minimum of \( i_0 \) and the smallest index \( i \) such that any one of the random variables violates its corresponding trajectory. Let the event \( \mathcal{E}_X \) be of the form \( X(i) = x(i) + e(i) \) for all \( i \leq i_0 \), where \( X(i) \) is some random variable, \( x(i) \) is the expected trajectory and \( e(i) \) is the error term. We show that the event \( \{ \tau = i_0 \} \) holds by means of \( \{ \tau = i_0 \} = \cap_{X \in \mathcal{I}} \mathcal{E}_X \), where \( |\mathcal{I}| \) is polynomial in \( n \).

For each such random variable \( X(i) \), we define a critical interval \( I_X \) for its bound (upper and lower) that has one endpoint at the bound we are trying to maintain and the other slightly closer to the expected trajectory of the random variable. Consider a fixed step \( j \leq i_0 \) such that \( X(j) \in I_X \). Define the stopping time \( \tau_{X,j} \) to be \( \tau_{X,j} = \min\{i_0, \max\{j, \tau\} \} \), the smallest \( i \geq j \) such that \( X(i) \notin I_X \), which made us possible to establish the martingale condition and apply the martingale inequality in Lemma 2.2 or Lemma 2.3. Establish bounds on the events that the designed variable crosses its critical interval in the process, such that a simple application of the union bound over all starting point \( j \) shows that the probability of the occurrence of any event in the collection is low to complete the proof.

As a supplement, we list some necessary inequalities that we need in the following proof of Theorem 3.1. By Lemma 2.1, we have

\[
\sum_{U_m \in \mathcal{K}_m(i)} R_{k,U_m}^2 \geq \frac{(\sum_{U_m \in \mathcal{K}_m(i)} R_{k,U_m})^2}{|\mathcal{K}_m(i)|}
\]

for any \( U_m \in \mathcal{K}_m(i) \) with \( 2 \leq m \leq k - 1 \). Firstly, note that \( \sum_{U_m \in \mathcal{K}_m(i)} R_{k,U_m} = \binom{k}{m} Q_k(i) \) because each element on the right side is counted \( \binom{k}{m} \) times on the left side. Next, note that \( |\mathcal{K}_2(i)| = |E(i)| \sim \frac{n^2}{p} \) in (3.2) when \( p \geq p_0 \) in (3.9), and we recursively apply the equation \( |\mathcal{K}_m(i)| \leq \frac{n}{m} |\mathcal{K}_{m-1}(i)| \) to achieve \( |\mathcal{K}_m(i)| \leq \frac{n}{m!} p \) with \( 2 \leq m \leq k - 1 \). Thus, we have

\[
\sum_{U_m \in \mathcal{K}_m(i)} R_{k,U_m}^2 \geq \frac{m! \binom{k}{m}^2 Q_k^2(i)}{n^m p}. \tag{4.1}
\]

Conditioned on the estimates in (3.5) hold on \( R_{k,U_m} \) for any \( U_m \in \binom{m}{n} \) with \( 2 \leq m \leq k - 1 \), we also have the upper bounds of \( \sum_{U_m \in \mathcal{K}_m(i)} R_{k,U_m}^2 \). For \( m = 2 \), we have \( \beta_2 = k - \frac{1}{2} \) in (3.7) and \( |\mathcal{K}_2(i)| \sim \frac{n^2}{2p} \), then by Lemma 2.1,

\[
\sum_{U_2 \in \mathcal{K}_2(i)} R_{k,U_2}^2 \leq \frac{(\sum_{U_2 \in \mathcal{K}_2(i)} R_{k,U_2})^2}{|\mathcal{K}_2(i)|} + 4|\mathcal{K}_2(i)|((\sigma n\beta_2 \log^\gamma n)^2)
\approx \frac{2! \binom{k}{2} Q^2_k(i)}{n^2 p} + 2\sigma^2 n^{2k-3} p \log^\gamma n. \tag{4.2}
\]

For \( 3 \leq m \leq k - 1 \), by the estimates in (3.5) and \( |\mathcal{K}_m(i)| \leq \frac{n^m}{m!} p \), the trivial upper bound is

\[
\sum_{U_m \in \mathcal{K}_m(i)} R_{k,U_m}^2 \leq \frac{n^m p}{m!} \left( \frac{n^{k-m} p^{(k-m)/(2)}}{(k-m)!} + \sigma n^{\beta m} \log^\gamma n \right)^2. \tag{4.3}
\]
4.1 Tracking $Q_k(i)$

For the upper bound of $Q_k(i)$, we introduce a critical interval as

$$I^u_k = \left(\frac{n^k}{k!}p(k) + Bn^{-1}p(k)^{-4}, \frac{n^k}{k!}p(k) + \frac{n^{k-1}}{2}p(k)^{-4}\right),$$

where

$$B = \frac{1}{2} - \frac{1}{2(\frac{k}{2})^2} + \frac{1}{3(\frac{k}{2})(k-4)!} \leq \frac{1}{2}.$$  \hfill (4.5)

Consider a fixed step $j \leq i_0$. Suppose $Q_k(i) \in I^u_k$. Define

$$\tau^u_{k,j} = \min\{i_0, \max\{j, \tau\}\},$$

the smallest $i \geq j$ such that $Q_k(i) \notin I^u_k$. \hfill (4.6)

Let $j \leq i \leq \tau^u_{k,j}$, thus all calculations in this subsection are conditioned on the estimates in (3.5) hold on $R_{k,U_m}$ for any $U_m \in \left(\frac{\lfloor n \rfloor}{m}\right)$ with $2 \leq m \leq k - 1$.

By the equation shown in (2.5), it follows that

$$\mathbb{E}[\Delta Q_k | F_i] = (-1)^{k+1}(k-1) - \frac{1}{Q_k(i)} \sum_{U_2 \in K_2(i)} R_{k,U_2}^2 + \cdots + \frac{(-1)^{k}(k-2)}{Q_k(i)} \sum_{U_{k-1} \in K_{k-1}(i)} R_{k,U_{k-1}}^2$$

$$< (-1)^{k+1}(k-1) - \frac{2(k)^2}{n^2p} Q_k(i) + 2 \frac{n^3p}{3!} \frac{n^{k-3}}{(k-3)!} p(k)^{-3} + \sigma n^{\beta_3} \log^{\gamma_3} n \right)^2$$

$$+ O(n^{k-4}p(k)^{-1}),$$

where $\sum_{U_2 \in K_2(i)} R_{k,U_2}^2$ and $\sum_{U_3 \in K_3(i)} R_{k,U_3}^2$ are replaced by the equations in (4.1) and (4.3), the last term $O(n^{k-4}p(k)^{-1})$ comes from $\sum_{U_4 \in K_4(i)} R_{k,U_4}^2$ in (4.1) that dominates all the remaining terms.

Since $Q_k(i) \in I^u_k$ is in (4.4), we further have

$$\mathbb{E}[\Delta Q_k | F_i] < (-1)^{k+1}(k-1) - \frac{2(k)^2}{k!} \frac{n^{k-2}}{p(k)^{-1}} - 2 \left(\frac{k}{2}\right)^2 Bn^{-3} p(k)^{-5}$$

$$+ \frac{k!n^{k-3}}{3(k-3)!2} p(k)^{-5} + O(\sigma n^{\beta_3} p^{-2} \log^{\gamma_3} n),$$

where $O(n^{k-4}p(k)^{-1})$ is absorbed into $O(\sigma n^{\beta_3} p^{-2} \log^{\gamma_3} n)$ when $\beta_3$ is in (3.7).

For all $i$ with $j \leq i \leq \tau^u_{k,j}$, define the sequence of random variables to be

$$U(i) = Q_k(i) - \frac{n^k}{k!} p(k) - \frac{n^{k-1}}{2} p(k)^{-4}. \hfill (4.8)$$

Claim 4.1: The sequence $U(i), U(i+1), \cdots, U(\tau^u_{k,j})$ is a supermartingale and the maximum one step $\Delta U$ is $O(\sigma n^{k-5/2} \log^{\gamma_2} n)$.

Proof of Claim 4.1. To see this, for $j \leq i \leq \tau^u_{k,j}$, as the equation in (4.8), we have

$$\mathbb{E}[\Delta U | F_i] = \mathbb{E}[\Delta Q_k | F_i] - \frac{n^k}{k!} \left[p(k)(i+1) - p(k)(i)\right] - \frac{n^{k-1}}{2} \left[p(k)^{-4}(i+1) - p(k)^{-4}(i)\right].$$
Note that \( p = p(i) = 1 - k(k-1)t, p(i+1) = 1 - k(k-1)(t + \frac{1}{n^r}) \) in (3.1), then by Taylor’s expansion, we have

\[
\mathbb{E}[\Delta U | F_i] = \mathbb{E}[\Delta Q_k | F_i] - \frac{n^k}{k!} \left[ -\binom{k}{2} \frac{k(k-1)}{n^2} p^{(k)} - 1 + O\left(\frac{1}{n^4} p^{(k)-2}\right) \right] - \frac{n^{k-1}}{2} \left[ -\left( \binom{k}{2} - 4 \right) \frac{k(k-1)}{n^2} p^{(k)-5} + O\left(\frac{1}{n^4} p^{(k)-6}\right) \right]
\]

\[
= \mathbb{E}[\Delta Q_k | F_i] + \frac{2(\binom{k}{2})^2 n^{k-2}}{k!} p^{(k)-1} + \left[ \binom{k}{2} - 4 \right] \binom{k}{2} n^{k-3} p^{(k)-5} + O(n^{k-4} p^{(k)-2})
\]

(4.9)

where \( O(n^{k-5} p^{(k)-6}) \) is absorbed into \( O(n^{k-4} p^{(k)-2}) \) when \( p \geq p_0 \) in (3.9). With the help of the equation in (4.7), we further have

\[
\mathbb{E}[\Delta U | F_i] < (-1)^{k+1}(k - 1) - \left[ 2\binom{k}{2} B - \binom{k}{2} + 4\binom{k}{2} - \frac{k!}{3(k-3)!^2} \right] n^{k-3} p^{(k)-5} + O(\sigma n^{\beta_3} p^{-2} \log^{\gamma_3} n)
\]

\[
< (-1)^{k+1}(k - 1) - 2\binom{k}{2} n^{k-3} p^{(k)-5} + O(\sigma n^{\beta_3} p^{-2} \log^{\gamma_3} n),
\]

where

\[
2\binom{k}{2} B - \binom{k}{2} + 4\binom{k}{2} - \frac{k!}{3(k-3)!^2} = 3\binom{k}{2} - \binom{k}{2} \frac{2}{3(k-3)!} > 2\binom{k}{2}
\]

by \( B \) shown in (4.5), and \( O(n^{k-4} p^{(k)-2}) \) is absorbed into \( O(\sigma n^{\beta_3} p^{-2} \log^{\gamma_3} n) \) by \( \beta_3 \) shown in (3.7). Note that \( \binom{k}{2} n^{k-3} p^{(k)-5} > O(\sigma n^{\beta_3} p^{-2} \log^{\gamma_3} n) + (-1)^{k+1}(k - 1) \) when \( p \geq p_0 \) in (3.9), and appropriate choices of \( \lambda \) and \( \gamma_3 \), then we have \( \mathbb{E}[\Delta U | F_i] < 0 \) and the sequence \( U(j), U(j + 1), \cdots, U(\tau^{\beta_{Q_k}}) \) is a supermartingale.

Next, we show the maximum one step \( \Delta U \) is \( O(\sigma n^{k-5/2} \log^{\gamma_2} n) \). As the equations shown in (4.8) and (4.9), we have

\[
\Delta U = \Delta Q_k + \frac{2(\binom{k}{2})^2}{k!} n^{k-2} p^{(k)-1} + \left[ \binom{k}{2} - 4 \right] \binom{k}{2} n^{k-3} p^{(k)-5} + O(n^{k-4} p^{(k)-2}).
\]

Apply the equation of \( \Delta Q_k \) shown in (2.4) to the above display, by the equation of \( R_{k,U_m} \) shown in (3.5) for any \( U_m \in \left(\frac{\lambda}{m}\right) \), and \( \beta_m \) shown in (3.7) with \( 2 \leq m \leq k - 1 \), then we finally have

\[
\Delta U \leq -\binom{k}{2} \left( \frac{n^{k-2}}{(k-2)!} p^{(k)-1} - \sigma n^{\beta_2} \log^{\gamma_2} n \right) + \binom{k}{3} \left( \frac{n^{k-3}}{(k-3)!} p^{(k)-5} + \sigma n^{\beta_3} \log^{\gamma_3} n \right) + \cdots
\]

\[
+ \frac{2(\binom{k}{2})^2}{k!} n^{k-2} p^{(k)-1} + \left[ \binom{k}{2} - 4 \right] \binom{k}{2} n^{k-3} p^{(k)-5} + O(n^{k-4} p^{(k)-2})
\]

\[
= O(\sigma n^{k-5/2} \log^{\gamma_2} n).
\]

The claim follows. \( \square \)
Now, apply Lemma 2.2 to the sequence $U(j), U(j+1), \ldots, U(\tau_{Q_k,j})$. The number of steps in this sequence is $O(n^2p)$ because $|E(i)| \sim \frac{n^2}{2}p$ in (3.2) when $p \geq p_0$ in (3.9). Since $Q_k(j) \in I^u_{Q_k}$ in (4.4), we have the initial value $U(j) \geq -(\frac{1}{2(\frac{3}{2})} - \frac{1}{3(\frac{3}{2})(k-4)})n^{k-1}p^{(\frac{3}{2})-4}$. Then, for all $i$ with $j \leq i \leq \tau_{Q_k,j}^u$, the probability of a large deviation for $Q_k(i)$ beginning at the step $j$ is at most

$$
\Pr[Q_k(i) \geq \frac{n^k}{k!}b_{(\frac{3}{2})}^{(k)} + \frac{n^{k-1}}{2}p^{(\frac{3}{2})-4}] = \Pr[U(i) \geq 0] \leq \exp[-\Omega\left(\frac{(n^{k-1}p^{(\frac{3}{2})-4})^2}{(n^2p)(\sigma n^{k-5/2} \log^{\gamma} n)^2}\right)] = \exp[-\Omega\left(\frac{n^p^{2(k/2)-9}}{\sigma^2 \log^{2\gamma} n}\right)].
$$

By the union bound, note that there are at most $n^2$ possible values of $j$ in (3.1) and $p \geq p_0$ in (3.9), then we have

$$n^2 \exp[-\Omega\left(\frac{n^p^{2(k/2)-9}}{\sigma^2 \log^{2\gamma} n}\right)] = o(1).
$$

W.h.p., $Q_k(i)$ never crosses its critical interval $I^u_{Q_k}$ in (4.1), and so the upper bound of $Q_k(i)$ in (3.3) is true.

**Remark 4.1.** Proving the lower bound of $Q_k(i)$ is similar. We show the proof in the appendix for reference.

### 4.2 Tracking $R_{k,U_m}$ for any $U_m \in \binom{[n]}{m}$ with $2 \leq m \leq k - 1$

We prove the dynamic concentration of $R_{k,U_m}$ for any $U_m \in \binom{[n]}{m}$ with $2 \leq m \leq k - 1$ in this subsection. Fix one subset $U_{m^*} \in \binom{[n]}{m^*}$ for some $m^*$ with $2 \leq m^* \leq k - 1$. We start with the upper bound of $R_{k,U_{m^*}}$. Our critical interval for the upper bound of $R_{k,U_{m^*}}$ is

$$I^u_{R_{k,U_{m^*}}} = \left(\frac{n^{k-m^*}}{(k-m^*)1}\right)p^{(\frac{3}{2})-(\frac{m^*}{2})} + (\sigma - 1) n^{\beta_{m^*} \log^{\gamma_{m^*}} n},$$

$$\frac{n^{k-m^*}}{(k-m^*)1}p^{(\frac{3}{2})-(\frac{m^*}{2})} + \sigma n^{\beta_{m^*} \log^{\gamma_{m^*}} n},$$

where $\beta_{m^*} = k - m^* - \frac{(\frac{3}{2})-(\frac{m^*}{2})}{2(\frac{3}{2})-2}$ in (3.7). Consider a fixed step $j \leq i_0$. Suppose $R_{k,U_{m^*},(j)} \in I^u_{R_{k,U_{m^*}}}$. Define

$$\tau_{R_{k,U_{m^*}},j}^u = \min\{i_0, \max\{j, \tau\}, \text{the smallest } i \geq j \text{ such that } R_{k,U_{m^*}} \notin I^u_{R_{k,U_{m^*}}}\}. \quad (4.11)$$

Let $j \leq i \leq \tau_{R_{k,U_{m^*}},j}^u$, thus all calculations are conditioned on the events that the estimates in (3.3) and (3.4) hold on $Q_k(i)$, and the estimates in (3.5) hold on $R_{k,U_m}$ for all $U_m \in \binom{[n]}{m}$ with $2 \leq m \leq k - 1$ and $U_m \neq U_{m^*}$.
Take one \( U^c_{m*} \in K_{k-\rho} \cap N_{U^c_{m*}} \) in \( G(i) \) and let \( Q_{k,U^c_{m*},U^c_{m*}} \) be the number of \( K_k \)s in \( G(i) \) such that the removal of the edges in any one of these \( K_k \)s results in \( U^c_{m*} \notin K_{k-\rho} \cap N_{U^c_{m*}} \) in \( G(i+1) \). Then, we have

\[
\mathbb{E}[\Delta R_{k,U^c_{m*}} | F_i] = -\sum_{U^c_{m*} \in K_{k-\rho} \cap N_{U^c_{m*}}} \frac{Q_{k,U^c_{m*},U^c_{m*}}}{Q_k(i)}.
\] (4.12)

In order to count \( Q_{k,U^c_{m*},U^c_{m*}} \), let \( H \subseteq U_{m*} \cup U^c_{m*} \) and \( Q^H_{k,U^c_{m*},U^c_{m*}} \) be the number of \( K_k \)s in \( Q_{k,U^c_{m*},U^c_{m*}} \) such that these \( K_k \)s satisfy \( K_k \subseteq (U_{m*} \cup U^c_{m*}) = H \). Define \( |H| = h \). To ensure that the removal of the edges in any one of these \( K_k \)s results in \( U^c_{m*} \notin K_{k-\rho} \cap N_{U^c_{m*}} \) in \( G(i+1) \), it is observed that \( H \cap U^c_{m*} \neq \emptyset \) and \( 2 \leq h \leq k \).

Choose \( H \in \cup_{\rho=0}^{h-1} (U^c_{m*}) \oplus (U^c_{\rho}) \), where \((U^c_{m*}) \oplus (U^c_{\rho})\) denotes the collection of union sets consisting of \( \rho \) vertices in \( U_{m*} \) and \( h-\rho \) vertices in \( U^c_{m*} \). Hence, \( Q_{k,U^c_{m*},U^c_{m*}} \) is decomposed into

\[
Q_{k,U^c_{m*},U^c_{m*}} = \sum_{h=2}^{k} \sum_{\rho=0}^{h-1} \sum_{H \in (U^c_{m*}) \oplus (U^c_{\rho})} Q^H_{k,U^c_{m*},U^c_{m*}}.
\] (4.13)

Following the inclusion-exclusion counting technique shown in (2.4), we have

\[
Q^H_{k,U^c_{m*},U^c_{m*}} = 1_H \cdot R_{k,H} - \sum_{T_1 \in (U^c_{m*}) \cup (U^c_{m*}) \setminus H} 1_{H \cup T_1} \cdot R_{k,H \cup T_1} + \cdots
\]

\[
+ \sum_{T_{k-h} \in (U^c_{m*}) \cup (U^c_{m*}) \setminus H} (-1)^{k-h} 1_{H \cup T_{k-h}} \cdot R_{k,H \cup T_{k-h}}
\]

\[
= \sum_{z=0}^{k-h} \sum_{T_z \in (U^c_{m*}) \cup (U^c_{m*}) \setminus H} (-1)^z 1_{H \cup T_z} \cdot R_{k,H \cup T_z},
\]

where \( 1_{H \cup T_z} \) with \( 0 \leq z \leq k-h \) is the indicator random variable depending on whether the subgraph induced by \( H \cup T_z \) in \( G(i) \) is complete or not. Combining with the equation in (4.13), we further have

\[
Q_{k,U^c_{m*},U^c_{m*}} = \sum_{h=2}^{k-1} \sum_{\rho=0}^{h-1} \sum_{z=0}^{h-z} (-1)^z 1_{H \cup T_z} \cdot R_{k,H \cup T_z}.
\]

In the above display, for fixed integers \( h \) and \( z \), we recount the union \( H \cup T_z \) as a subset \( H_{h+z} \in (U^c_{m*}) \oplus (U^c_{h+z-\zeta}) \) with \( 0 \leq \zeta \leq h+z \), then each \( H_{h+z} \) is counted \([h+z] - \binom{h}{\zeta}] \) times in \( H \cup T_z \) because \( H \cap U^c_{m*} \neq \emptyset \), which means that

\[
= \sum_{z=0}^{h+z} \sum_{H_{h+z} \in (U^c_{m*}) \oplus (U^c_{h+z-\zeta})} [h+z] - \binom{h}{\zeta} (-1)^z 1_{H_{h+z}} \cdot R_{k,H_{h+z}}.
\]
It follows that

$$Q_{k,U_m^*, U_m^{c*}} = \sum_{h=2}^{k-k-h+z} \sum_{z=0}^{h} \sum_{\zeta \in \{U_m^*, U_m^{c*}\}} \sum_{Hh+z \in \{U_m^*, U_m^{c*}\}} \left[ \binom{h+z}{h} - \frac{1}{h} \right] (-1)^z 1_{Hh+z} \cdot R_{k,Hh+z}. \quad (4.14)$$

In fact, $Q_{k,U_m^*, U_m^{c*}}$ is the sum of all elements in the upper triangular matrix below

$$
\begin{pmatrix}
\sum_{\zeta=0}^{2} \sum_{Hk \in \{U_m^*, U_m^{c*}\}} (1-0)(\zeta) 1_{Hk} \cdot R_{k,Hk} & \cdots & \sum_{\zeta=0}^{2} \sum_{Hk \in \{U_m^*, U_m^{c*}\}} (1-2)(k) 1_{Hk} \cdot R_{k,Hk} \\
\sum_{\zeta=0}^{2} \sum_{Hk \in \{U_m^*, U_m^{c*}\}} (1-0)(\zeta) 1_{Hk} \cdot R_{k,Hk} & \cdots & \sum_{\zeta=0}^{2} \sum_{Hk \in \{U_m^*, U_m^{c*}\}} (1-2)(k) 1_{Hk} \cdot R_{k,Hk} \\
\sum_{\zeta=0}^{2} \sum_{Hk \in \{U_m^*, U_m^{c*}\}} (1-0)(\zeta) 1_{Hk} \cdot R_{k,Hk} & \cdots & \sum_{\zeta=0}^{2} \sum_{Hk \in \{U_m^*, U_m^{c*}\}} (1-2)(k) 1_{Hk} \cdot R_{k,Hk} \\
\sum_{\zeta=0}^{2} \sum_{Hk \in \{U_m^*, U_m^{c*}\}} (1-0)(\zeta) 1_{Hk} \cdot R_{k,Hk} & \cdots & \sum_{\zeta=0}^{2} \sum_{Hk \in \{U_m^*, U_m^{c*}\}} (1-2)(k) 1_{Hk} \cdot R_{k,Hk} \\
\end{pmatrix}
$$

with the line corresponds to the index $h$ and the column corresponds to the index $z$ in (4.14), respectively. Recalculate $Q_{k,U_m^*, U_m^{c*}}$ according to every back diagonal lines to be

$$Q_{k,U_m^*, U_m^{c*}} = \sum_{h=2}^{k-k-h+z} \sum_{\zeta=0}^{h} \sum_{Hk \in \{U_m^*, U_m^{c*}\}} (-1)^{h-i} \binom{h}{s} - \binom{\zeta}{s} 1_{Hk} \cdot R_{k,Hk}. \quad (4.15)$$

Note that there is no $R_{k,U_m}$ on the right side of (4.14) because $R_{k,U_m}$ corresponds to the case when $\zeta = h$. Thus, the estimates on $Q_k(i)$ in (3.3) and (3.4), the estimates on $R_{k,U_m}$ in (3.5) for all $U_m \in \{\binom{m}{n}\}$, $U_m \neq U_m^*$ with $2 \leq m \leq k-1$, already support the calculation of $Q_{k,U_m^*, U_m^{c*}}$ in (4.15).

Furthermore, according to the expressions of $R_{k,Hk}$ for $2 \leq h \leq k-1$ in (3.5), the term $R_{k,Hk}$ dominates the sum on the right side of (4.15). Thus, we have $\zeta = 0, 1$ and $s = 2$. It follows that,

$$Q_{k,U_m^*, U_m^{c*}} = \left[ \binom{k-m^*}{2} + m^*(k-m^*) \right] \left[ \binom{n-2}{k-2} \right] p^*(n) - \sigma n^{\beta_2} \log n + O(n^{k-3}p^*(n)^{-3}), \quad (4.16)$$

where $\binom{k-m^*}{2}$ counts the number of $R_{k,H_2}$ when $\zeta = 0$ and $s = 2$, $m^*(k-m^*)$ counts the number of $R_{k,H_2}$ when $\zeta = 1$ and $s = 2$. Note that $\binom{k-m^*}{2} + m^*(k-m^*) = \binom{k}{2} - \binom{m^*}{2}$ and $\beta_2 = k - \frac{5}{2}$ in (3.7), combining the equations in (4.12) and (4.16), and applying the estimates of $Q_k(i)$ in (3.3), we have

$$\mathbb{E} \left[ \Delta R_{k,U_m^*, U_m^{c*}} \right] < - \sum_{U_m^{c*} \in K_{k-m^*} \cap N_{U_m^*}} \frac{\left[ \binom{k}{2} - \binom{m^*}{2} \right] \left( n^{k-2} \right) \left( \frac{k-2!}{k-m^*} \right) \left( \frac{k-m^*}{k-m^*} \right) \left( \frac{k-m^*}{k-m^*} \right) }{n^k p^*(n)}. \quad (4.17)$$

The ways to choose $U_m^{c*} \in K_{k-m^*} \cap N_{U_m^*}$ is $R_{k,U_m^*}$ and $R_{k,U_m^*} \in I_{R_{k,U_m^*}}$ in (4.10), then it further follows that

$$\mathbb{E} \left[ \Delta R_{k,U_m^*, U_m^{c*}} \right] < - \frac{\left[ \binom{k}{2} - \binom{m^*}{2} \right] \left( n^{k-2} \right) \left( \frac{k-2!}{k-m^*} \right) \left( \frac{k-m^*}{k-m^*} \right) \left( \frac{k-m^*}{k-m^*} \right) }{n^k p^*(n)} + O(n^{k-3}p^*(n)^{-3}).$$
Rearrange the above equation to be

\[
\mathbb{E}[\Delta R_{k,U_m^*} | \mathcal{F}_i] < -\left[\frac{(k)}{2} - \binom{m^*}{2}\right]k(k-1)n^{k-m^*-2}\frac{(k)}{2}p^{(k)}(k-1)^{(m^*)-1}
+ \left[\frac{(k)}{2} - \binom{m^*}{2}\right]k\sigma n^{k-m^*-\frac{2}{2}}\log \gamma^2 n
+ \left[\frac{(k)}{2} - \binom{m^*}{2}\right]k(k-1)(\sigma - 1)n^{\beta m^*-2}\log \gamma^m n
+ \left[\frac{(k)}{2} - \binom{m^*}{2}\right]k\sigma(\sigma - 1)n^{\beta m^*-\frac{2}{2}}\log \gamma^m n
+ O(n^{k-m^*-3}p^{(k)}(k-1)^{(m^*)-3}).
\] (4.17)

For all \( i \) with \( j \leq i \leq \tau_{R_k,U_m^*,j} \), define the sequence of random variables to be

\[
Z_{U_m^*}(i) = R_{k,U_m^*} - \frac{n^{k-m^*}}{(k-m^*)!}p^{(k)}(k-1)^{(m^*)-1} - (\sigma - 1)n^{\beta m^*}\log \gamma^m n.
\] (4.18)

In order to prove the upper bound of \( R_{k,U_m^*} \) is the equation in (3.5), we prove the following two claims.

**Claim 4.2:** Removing the edges of one \( K_k \) in \( G(i) \), we have

\[
R_{k,U_m^*}(i) - R_{k,U_m^*}(i + 1) = O\left(n^{k-m^*-1}p^{(k)}(k-1)^{(m^*)-1}\right).
\]

**Proof of Claim 4.2.** When we remove the edges of one \( K_k \) from \( G(i) \), note that \( R_{k,U_m^*}(i) \) is the number of \( K_{k-m^*} \)s in which every vertex is in \( N_{U_m^*} \), then it is clearly true that \( R_{k,U_m^*}(i) - R_{k,U_m^*}(i + 1) \geq 0 \). Suppose the removed \( K_k \) contains one vertex in \( U_m^* \), denoted by \( u \), that is in \( N_{U_m^*} \). Then the number of \( K_{k-m^*} \)s in which every vertex is in \( N_{U_m^*} \cup \{w\} \) is at most \( R_{k,U_m^*} \cup \{w\}(i) \). By the equation in (3.5), we complete the proof. \( \square \)

**Claim 4.3:** The sequence \(-Z_{U_m^*}(j), -Z_{U_m^*}(j + 1), \ldots, -Z_{U_m^*}(\tau_{R_k,U_m^*,j}^m)\) is an \((\eta, N)\)-bounded submartingale, where \( \eta = \Theta(n^{k-m^*-2}p^{(k)}(k-1)^{(m^*)-1}) \) and \( N = \Theta(n^{k-m^*-1}p^{(k)}(k-1)^{(m^*)-1}) \) for \( 2 \leq m^* \leq k - 1 \).

**Proof of Claim 4.3.** For all \( i \) with \( j \leq i \leq \tau_{R_k,U_m^*,j}^m \), as the equation in (4.18), we have

\[
\mathbb{E}[\Delta Z_{U_m^*} | \mathcal{F}_i] = \mathbb{E}[\Delta R_{k,U_m^*} | \mathcal{F}_i] - \frac{n^{k-m^*}}{(k-m^*)!}\left[p^{(k)}(k-1)^{(m^*)-1}(i + 1) - p^{(k)}(k-1)^{(m^*)-1}(i)\right]
- n^{\beta m^*}\log \gamma^m n[\sigma(i + 1) - \sigma(i)].
\]

Note that \( p = p(i) = 1 - k(k-1)t, p(i + 1) = 1 - k(k-1)(t + \frac{1}{n^2}) \) in (3.1), \( \sigma(i) = 1 - \frac{k(k-1)}{4}\log p(i) \),
and \( \sigma(i + 1) = 1 - \frac{k(k - 1)}{4} \log p(i + 1) \) in (3.8), then

\[
\mathbb{E}[\Delta Z_{U_{m^*}} | \mathcal{F}_i] = \mathbb{E}[\Delta R_{k,U_{m^*}} | \mathcal{F}_i] - \frac{n^{k-m^*}}{(k-m^*)!} \left[ -\binom{k}{2} - \binom{m^*}{2} \right] \frac{k(k-1)}{n^2} p^{(k)}_{(2)} - \binom{m^*}{2} - 1 \\
+ O\left( \frac{1}{n}\frac{k}{p^{(k)}_{(2)}} - \binom{m^*}{2} - 2 \right) \right] - n^{\beta_{m^*}} \log^{\gamma_{m^*}} n \left[ \frac{\sigma'}{n^2} + O\left( \frac{\sigma''}{n^2} \right) \right] \\
= \mathbb{E}[\Delta R_{k,U_{m^*}} | \mathcal{F}_i] + \left[ \binom{k}{2} - \binom{m^*}{2} \right] \frac{k(k-1)n^{k-m^*-2}}{(k-m^*)!} p^{(k)}_{(2)} - \binom{m^*}{2} - 1 \\
- \sigma' n^{\beta_{m^*} - 2} \log^{\gamma_{m^*}} n + O\left( n^{k-m^*-4} p^{(k)}_{(2)} - \binom{m^*}{2} - 2 \right),
\]

(4.19)

where \( O(\sigma'' n^{\beta_{m^*} - 4} \log^{\gamma_{m^*}} n) \) is absorbed into \( O(n^{k-m^*-4} p^{(k)}_{(2)} - \binom{m^*}{2} - 2) \) because \( \sigma'' = O(p^{-2}) \) in (3.6), \( \beta_{m^*} \) shown in (3.7), \( p \geq p_0 \) in (3.9), and appropriate choices of the constants \( \lambda \) and \( \gamma_{m^*} \).

Combining the equations in (4.17) and (4.19), we further have

\[
\mathbb{E}[\Delta Z_{U_{m^*}} | \mathcal{F}_i] < \frac{\binom{k}{2} - \binom{m^*}{2}}{(k-m^*)!} k! \sigma \frac{n^{k-m^*-\frac{5}{2}}}{p^{(m^*)}_{(2)}} \log^{\gamma_{m^*}} n \\
- \frac{k(k-1)(\sigma - 1)}{2p} n^{\beta_{m^*} - 2} \log^{\gamma_{m^*}} n \\
+ \frac{k! \sigma}{p_{(2)}^{\frac{1}{2}}} \frac{n^{\beta_{m^*} - \frac{5}{2}}}{\log^{\gamma_{m^*}} n} \\
- \sigma' n^{\beta_{m^*} - 2} \log^{\gamma_{m^*}} n + O\left( n^{k-m^*-3} p^{(k)}_{(2)} - \binom{m^*}{2} - 3 \right),
\]

(4.20)

where \( O(n^{k-m^*-4} p^{(k)}_{(2)} - \binom{m^*}{2} - 2) \) in (4.19) is absorbed into \( O(n^{k-m^*-3} p^{(k)}_{(2)} - \binom{m^*}{2} - 3) \) in (4.17). At last, we have \( \mathbb{E}[\Delta Z_{U_{m^*}} | \mathcal{F}_i] < 0 \) in (4.20) because the following inequalities

\[
\frac{k! \sigma}{(k-m^*)! p^{(m^*)}_{(2)}} n^{k-m^*-\frac{5}{2}} \log^{\gamma_{m^*}} n < \frac{k(k-1)(\sigma - 1)}{2p} n^{\beta_{m^*} - 2} \log^{\gamma_{m^*}} n,
\]

\[
\frac{k! \sigma}{p_{(2)}^{\frac{1}{2}}} \frac{n^{\beta_{m^*} - \frac{5}{2}}}{\log^{\gamma_{m^*}} n} < \frac{k(k-1)}{2p} n^{\beta_{m^*} - 2},
\]

\[
O\left( n^{k-m^*-3} p^{(k)}_{(2)} - \binom{m^*}{2} - 3 \right) < \sigma' n^{\beta_{m^*} - 2} \log^{\gamma_{m^*}} n
\]

are obviously true when \( \beta_{m^*} \) is in (3.7), \( \sigma' = k^2(k-1)^2/4p \) in (3.8), \( p \geq p_0 \) in (3.9), and appropriate choices of \( \lambda \) and \( \gamma_{m^*} \). We have proved that the sequence \( -Z_{U_{m^*}(j)} - Z_{U_{m^*}(j+1)} - \cdots - Z_{U_{m^*}(\tau_{R_{k,U_{m^*}},j})} \) is a submartingale for any \( 2 \leq m^* \leq k - 1 \).

In the following, we show the sequence is \( (\eta,N) \)-bounded. By the equation in (4.18) and the calculation in (4.19), we have

\[
-Z_{U_{m^*}(i+1)} + Z_{U_{m^*}(i)} = R_{k,U_{m^*}}(i) - R_{k,U_{m^*}}(i+1) + \frac{n^{k-m^*}}{(k-m^*)!} \left[ p^{(k)}_{(2)} - \binom{m^*}{2} \right] (i+1) - p^{(k)}_{(2)} - \binom{m^*}{2} (i) \\
+ n^{\beta_{m^*}} \log^{\gamma_{m^*}} n \left[ \sigma(i+1) - \sigma(i) \right] \\
= R_{k,U_{m^*}}(i) - R_{k,U_{m^*}}(i+1) - \binom{k}{2} - \binom{m^*}{2} \frac{k(k-1)n^{k-m^*-2}}{(k-m^*)!} p^{(k)}_{(2)} - \binom{m^*}{2} - 1 \\
+ \sigma' n^{\beta_{m^*} - 2} \log^{\gamma_{m^*}} n + O\left( n^{k-m^*-4} p^{(k)}_{(2)} - \binom{m^*}{2} - 2 \right).
\]
Note that \( p \geq p_0 \) in (3.9), and appropriate choices of \( \lambda \) and \( \gamma_{m^*} \), we have
\[
\left[ \frac{k}{2} - \frac{m^*}{2} \right] \frac{k(k - 1)n^{k-m^*-2} - p(k-2) - (m^*)^{-1}}{(k - m^*)!} > \sigma n^{\beta_{m^*} - 2} \log \gamma_{m^*} n + O(n^{k-m^*-4} p(k-2) - (m^*)^{-2}).
\]
Thus, we take
\[
\eta = \left[ \frac{k}{2} - \frac{m^*}{2} \right] \frac{k(k - 1)n^{k-m^*-2} - p(k-2) - (m^*)^{-1}}{(k - m^*)!} = \Theta(n^{k-m^*-2} p(k-2) - (m^*)^{-1}).
\]
Since \(-Z_{(n^*)}(i+1) + Z_{(n^*)}(i) \leq R_{(n^*)}(i) - R_{(n^*)}(i+1)\), applying Claim 4.2, we take
\[
N = \Theta(n^{k-m^*-1} p(k-2) - (m^*)^{-1} + 1).
\]
We complete the proof of Claim 4.3. \( \square \)

The number of the sequence \(-Z_{(n^*)}(j), -Z_{(n^*)}(j+1), \ldots, -Z_{(n^*)}(\gamma_{(n^*)}(\gamma_{(n^*)}))\) is also \( O(n^2 p) \) in Lemma 2.3. Choose \( a = n^{\beta_{m^*} \log \gamma_{m^*} n} \), then \( a = o(\eta \ell) \). Lemma 2.3 yields that,
\[
\mathbb{P}[R_{(n^*)} \geq n^{k-m^*} (k - m^*)! p(k-2) - (m^*)^{-1} + \sigma n^{\beta_{m^*} \log \gamma_{m^*} n}]
\]
\[
= \mathbb{P}[-Z_{(n^*)}(i) \leq -n^{\beta_{m^*} \log \gamma_{m^*} n}]
\]
\[
\leq \exp \left[-\Omega \left( \frac{n^{2\beta_{m^*} \log 2 \gamma_{m^*} n}}{n^{k-m^*} \cdot n^{k-m^*+1}} \right) \right]
\]
\[
= \exp \left[-\Omega \left( n^{(\frac{m^*}{2})^{-1}} \log 2 \gamma_{m^*} n \right) \right].
\]
By the union bound, note that the choice to choose \( j, m^* (2 \leq m^* \leq k - 1) \) and \( U_m \in [\binom{n}{m}] \) is at most \( (k - 2)n^{m^*+2} \), then we also have
\[
(k - 2)n^{m^*+2} \exp \left[-\Omega \left( n^{(\frac{m^*}{2})^{-1}} \log 2 \gamma_{m^*} n \right) \right] = o(1)
\]
because it is clearly true when \( 3 \leq m^* \leq m - 1 \), and taking \( \gamma_2 > \frac{1}{2} \) for \( m^* = 2 \). In a conclusion, w.h.p., none of \( R_{k,U_m} \) for any \( U_m \in [\binom{n}{m}] \) with \( 2 \leq m \leq k - 1 \) have such a large upward deviations.

**Remark 4.2.** The argument for the lower bound of \( R_{k,U_m} \) in (3.5) for any \( U_m \in [\binom{n}{m}] \) with \( 2 \leq m \leq k - 1 \) is the symmetric analogue of the above analysis.

## 5 Conclusions

For the random \( K_k \)-removal algorithm, there are less direct results when \( k \geq 4 \) because their evolutionary structures are more complicated than the case \( k = 3 \) to investigate. We establish dynamic concentrations of complete higher codegree around the expected trajectories that are derived by
their pseudorandom properties. The final size of the random $K_k$-removal algorithm is at most $n^{2-1/(k(k-1)-2)+o(1)}$ for $k \geq 4$. In order to improve the result, it is observed that the main obstacle is the parameter $R_{k,U_m}$ for $2 \leq m \leq k-1$. The control over $R_{k,U_m}$ loses when $p$ around $p_0$ shown in Remark 3.3, while the probabilities of these extreme events are very low shown in Remark 3.4. The behaviors of these chosen random variables $R_{k,U_m}$ for $2 \leq m \leq k-1$ are not in a position to analyze the structures of the process further, and it is definitely possible to find some new ideas to track the random $K_k$-removal algorithm. This will be investigated in future work.

Acknowledgement

Fang Tian thanks X.-F. Pan for helping us to point out the faults in the equations (3.6) and (3.7), and some useful discussions in Remark 3.3 and 3.4. Fang Tian was supported by the National Natural Science Foundation of China (Grant No. 12071274). X.-F. Pan was supported by University Natural Science Research Project of Anhui Province under Grant No. KJ2020A0001.

References

[1] P. Bennett and T. Bohman, A natural barrier in random greedy hypergraph matching. *Combinator. Probab. Comp.*, 28 (2019), 816-825.

[2] T. Bohman, The triangle-free process. *Adv. Math.*, 221 (2009), 1653-1677.

[3] T. Bohman, A. Frieze and E. Lubetzky, A note on the random greedy triangle-packing algorithm. *J. Comb.*, 1 (2010), 477-488.

[4] T. Bohman, A. Frieze and E. Lubetzky, Random triangle removal. *Adv. Math.*, 280 (2015), 379-438.

[5] T. Bohman and L. Warnke, Large girth approximate Steiner triple systems. *J. Lond. Math. Soc.*, 100 (2019), 895-913.

[6] B. Bollobás, To prove and conjecture: Paul Erdős and his mathematics. *Amer. Math.*, 105(3) (1998), 209-237.

[7] H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis bases on the sum of observations. *Annals of Mathematical Statistics.*, 23 (1952), 493-507.

[8] D. Grable, On random greedy triangle packing. *Electron. J. Comb.*, 4 (1997), R11.

[9] H. Guo and L. Warnke, On the power of random greedy algorithms. *arXiv.2104.07854*.

[10] W. Hoeffding, Probability inequalities for sums of bounded variables. *J. Amer. Statist. Assoc.*, 58 (1963), 13-30.

[11] M. E. Picollelli, The final size of the $C_r$-free process. *SIAM J. Discrete Math.*, 28(3) (2014), 1276-1305.
Appendix

Appendix: Lower bound of $Q_k(i)$ (for Remark 4.1)

For the lower bound of $Q_k(i)$, we work with the critical interval

$$I^k_{Q_k} = \left( \frac{n^k}{k!} p(i) - \sigma^2 n^\alpha p^{-1} \log^\mu n, \frac{n^k}{k!} p(i) - \sigma (\sigma - 1) n^\alpha p^{-1} \log^\mu n \right),$$

(1)

where $\alpha$ is shown in (3.6). Consider a fixed step $j \leq i_0$. Similarly, suppose $Q_k(j) \in I^k_{Q_k}$ and define

$$\tau^k_{Q_k,j} = \min\{i_0, \max\{j, \tau\}, \text{the smallest } i \geq j \text{ such that } Q_k(i) \notin I^k_{Q_k} \}. \quad (2)$$

Let $j \leq i \leq \tau^k_{Q_k,j}$. All calculations in this subsection are conditioned on the estimates in (3.5) hold on $R_{k,U_m}$ for any $U_m \in \binom{[m]}{m}$ with $2 \leq m \leq k - 1$.

By the equations shown in (2.5), we get the estimate on $\mathbb{E}[\Delta Q_k|F_i]$ in reverse direction,

$$\mathbb{E}[\Delta Q_k|F_i] = (-1)^{k+1}(k-1) - \frac{1}{Q_k(i)} \sum_{U_2 \in K_2(i)} R^2_{k,U_2} + \cdots + \frac{(-1)^k(k-2)}{Q_k(i)} \sum_{U_{k-1} \in K_{k-1}(i)} R^2_{k,U_{k-1}}$$

$$> (-1)^{k+1}(k-1) - \frac{1}{Q_k(i)} \left( 2 \left( \frac{k}{2} \right)^2 Q_k^2(i) + 2 \sigma^2 n^{k-3} p \log^{2\gamma_2} n \right)$$

$$+ \frac{12 \left( \frac{k}{3} \right)^2 Q_k^2(i)}{n^3 p} + O(n^{k-4} p^{k-11})$$

$$= (-1)^{k+1}(k-1) - \frac{2 \left( \frac{k}{2} \right)^2 Q_k(i)}{n^2 p} + \frac{12 \left( \frac{k}{3} \right)^2 Q_k(i)}{n^3 p} + O(n^{k-3} \sigma^2 p^{-(k-1)+1} \log^{2\gamma_2} n),$$

where $\sum_{U_2 \in K_2(i)} R^2_{k,U_2}$ and $\sum_{U_3 \in K_3(i)} R^2_{k,U_3}$ are replaced by the equations in (4.1) and (4.2), the term $O(n^{k-4} p^{k-11})$ comes from $\sum_{U_4 \in K_4(i)} R^2_{k,U_4}$ in (4.3) that dominates all the remaining terms.

Since $Q_k(j) \in I^k_{Q_k}$ shown in (1), we further have

$$\mathbb{E}[\Delta Q_k|F_i] > (-1)^{k+1}(k-1) - \frac{2 \left( \frac{k}{2} \right)^2 \left( \frac{n^k}{k!} p(i) - \sigma (\sigma - 1) n^\alpha p^{-1} \log^\mu n \right)}{n^2 p}$$

$$+ \frac{12 \left( \frac{k}{3} \right)^2 \left( \frac{n^k}{k!} p(i) - \sigma^2 n^\alpha p^{-1} \log^\mu n \right)}{n^3 p} + O(n^{k-3} \sigma^2 p^{-(k-1)+1} \log^{2\gamma_2} n)$$

$$= (-1)^{k+1}(k-1) - \frac{2 \left( \frac{k}{2} \right)^2 n^{k-2} p^{(k-2)-1}}{k!} + \frac{2 \left( \frac{k}{2} \right)^2 \sigma (\sigma - 1) n^{\alpha-2} \log^\mu n}{p^2}$$

$$+ \frac{12 \left( \frac{k}{3} \right)^2 n^{k-3}}{k!} p^{(k-2)-1} + O(n^{k-3} \sigma^2 p^{-(k-1)+1} \log^{2\gamma_2} n),$$

(3)
where $\alpha$ is in (3.6), and $O(\sigma^2 n^{a-3} p^{-2} \log^\mu n)$ is absorbed into $O(n^{k-3} \sigma^2 p^{-\frac{k}{2}+1} \log^{2\gamma_2} n)$.

For all $i$ with $j \leq i \leq \tau_{Q_{k,j}}^\ell$, define the sequence of random variables to be

$$L(i) = Q_k(i) - \frac{n^k}{k!} p^{(k)}(i) + \sigma^2 n^a p^{-1} \log^\mu n. \quad (4)$$

**Claim A:** The sequence $L(j), L(j+1), \ldots, L(\tau_{Q_{k,j}}^k)$ is a submartingale and the maximum one step $\Delta L$ is $O(\sigma n^{k-5/2} \log^{\gamma_2} n)$.

**Proof of Claim A.** Similarly, for all $i$ with $j \leq i < \tau_{Q_{k,j}}^\ell$, as the equation shown in (4), we have

$$E[\Delta L|\mathcal{F}_i] = E[\Delta Q_k|\mathcal{F}_i] - \frac{n^k}{k!} \left[ p^{(k)}(i+1) - p^{(k)}(i) \right] + n^a \log^\mu n \left[ \frac{\sigma^2(i+1)}{p(i+1)} - \frac{\sigma^2(i)}{p(i)} \right].$$

Note that $p(i) = 1 - k(k-1)t$, $p(i+1) = 1 - k(k-1)(t + \frac{1}{n^2})$ in (3.1), $\sigma(i) = 1 - \frac{k(k-1)}{4} \log p(i)$, $\sigma(i+1) = 1 - \frac{k(k-1)}{4} \log p(i+1)$ in (3.8), then by Taylor’s expansion, we have

$$E[\Delta L|\mathcal{F}_i] = E[\Delta Q_k|\mathcal{F}_i] - \frac{n^k}{k!} \left[ \left( -\left( k \right)^2 \frac{k(k-1)}{n^2} \right) p^{(k)-1} + O(\frac{1}{n^4} p^{(k)-2}) \right]$$

$$+ n^a \log^\mu n \left[ \frac{2\sigma' p - \sigma^2 p'}{n^2 p^2} + O\left( \frac{\sigma^2}{n^4 p^3} \right) \right]$$

$$= E[\Delta Q_k|\mathcal{F}_i] + \frac{2\left( k^2 \right)^2 n^{k-2} p^{(k)-1}}{k!} + \frac{2\sigma' n^{a-2} \log^\mu n}{p}$$

$$+ O\left( n^{-4} p^{(k)-2} \right),$$

where $O(n^{a-4} \sigma^2 p^{-3} \log^\mu n)$ is absorbed into $O(n^{k-4} p^{(k)-2})$ because $\alpha$ is in (3.6), $p \geq p_0$ in (3.9), and appropriate choices of $\lambda$ and $\mu$. Combining the equations in (3) and (5), we have

$$E[\Delta L|\mathcal{F}_i] > (-1)^{k+1}(k-1) + \frac{\left[ \left( k^2 \right)^2 + k(k-1) \right] \sigma^2 n^{a-2} \log^\mu n}{p^2} - \frac{2\left( k^2 \right)^2 \sigma n^{a-2} \log^\mu n}{p^2}$$

$$+ 12\left( \frac{k^2}{3} \right) n^{k-3} p^{(k)-1} + \frac{2\sigma' n^{a-2} \log^\mu n}{p} + O\left( n^{k-3} \sigma^2 p^{\frac{k}{2}+1} \log^{2\gamma_2} n \right),$$

where $O(n^{k-4} p^{(k)-2})$ in (5) is absorbed into $O(n^{k-3} \sigma^2 p^{-\frac{k}{2}+1} \log^{2\gamma_2} n)$ in (3). We have

$$2\sigma' n^{a-2} \log^\mu n p^{-1} = 2\left( \frac{k}{2} \right)^2 \sigma n^{a-2} p^{-2} \log^\mu n$$

by $\sigma' = \frac{1}{4} k^2 (k-1)^2 p^{-1}$ in (3.8). It follows that

$$E[\Delta L|\mathcal{F}_i] > (-1)^{k+1}(k-1) + \frac{\left[ \left( k^2 \right)^2 + k(k-1) \right] \sigma^2 n^{a-2} \log^\mu n}{p^2}$$

$$+ 12\left( \frac{k^2}{3} \right) n^{k-3} p^{(k)-1} + O\left( n^{k-3} \sigma^2 p^{\frac{k}{2}+1} \log^{2\gamma_2} n \right). \quad (6)$$
Note that
\[
\left[2\binom{k}{2}^2 + k(k-1)\sigma^2 n^{α-2}p^{-2}\log n\right] > O(n^{k-3} \sigma^2 p^{-\binom{k}{2}+1} \log n) + (-1)^{k+1}(k-1)
\]
when \(α\) is in (3.6), \(p \geq p_0\) in (3.9), and appropriate choices of \(α\), \(μ\) and \(γ_2\). We have \(\mathbb{E}[\Delta L_1|\mathcal{F}_1] > 0\). The sequence \(L(j), L(j+1), \ldots, L(\tau_{Q,j})\) is a submartingale.

Next, we show the maximum one step \(\Delta L\) is \(O(\sigma n^{k-5/2} \log n)\). As the equation in (4) and the calculation in (5), we have
\[
\Delta L = \Delta Q_k + \frac{2(k^2)}{k!} n^{k-2} \frac{p^{\binom{k}{2}-1}}{p} + \frac{2\sigma n^{α-2} \log n}{p} + \frac{k(k-1)\sigma^2 n^{α-2} \log n}{p^2} + O(n^{k-4} p^{\binom{k}{2}-2}).
\]
Apply the equation of \(\Delta Q_k\) in (2.4) and the estimates on \(R_{k,U_m}\) for any \(U_m \in \binom{[n]}{m}\) when \(2 \leq m \leq k-1\) in (3.5) to the above display, then
\[
\Delta L \leq -\binom{k}{2} (\frac{n^{k-2}}{(k-2)!} p^{\binom{k}{2}-1} - \sigma n^{α} \log n) + \binom{k}{3} (\frac{n^{k-3}}{(k-3)!} p^{\binom{k}{2}-\binom{3}{2}} + \sigma n^{α} \log n) + \cdots
\]
\[
+ \frac{2(k^2)}{k!} n^{k-2} \frac{p^{\binom{k}{2}-1}}{p} + \frac{2\sigma n^{α-2} \log n}{p} + \frac{k(k-1)\sigma^2 n^{α-2} \log n}{p^2} + O(n^{k-4} p^{\binom{k}{2}-2})
\]
where \(α\) and \(β_2\) are in (3.6) and (3.7), the term \(O(\sigma^2 n^{α-2} p^{-2} \log n)\) is absorbed into \(O(\sigma n^{k-5/2} \log n)\) when \(p \geq p_0\) shown in (3.9), and appropriate choices of \(λ\), \(μ\) and \(γ_2\).

The number of steps in this sequence is also \(O(n^2 p)\). Since \(Q_k(j) \in I_{Q_k}^k\) in (1), we have \(L(j) < \sigma n^{α} p^{-1} \log n\) from (4). For all \(i\) with \(j \leq i \leq τ_{Q_k,j}\), Lemma 2.3 yields that the probability of such a large deviation beginning at the step \(j\) is at most
\[
\mathbb{P}\left[Q_k(i) \leq \frac{n^k}{k!} p^{\binom{k}{2}} - \sigma^2 n^{α} p^{-1} \log n\right] = \mathbb{P}\left[L(i) \leq 0\right]
\]
\[
\leq \exp\left[-Ω\left(\frac{(\sigma n^{α} p^{-1} \log n)^2}{(n^2 p)(\sigma n^{k-5/2} \log n)^2}\right)\right]
\]
\[
= \exp\left[-Ω\left(\frac{n^{2α-2k+3} \log^2 n}{p^3 \log n}\right)\right].
\]
By the union bound, note that there are at most \(n^2\) possible values of \(j\) shown in (3.1), then we have
\[
n^2 \exp\left[-Ω\left(\frac{n^{2α-2k+3} \log^2 n}{p^3 \log n}\right)\right] = o(1)
\]
with \(α\) is shown in (3.6). W.h.p. \(Q_k(i)\) never crosses its critical interval \(I_{Q_k}^k\) in (4), and so the lower bound on \(Q_k(i)\) in (3.4) is true. \(\square\)