The Triangle-Free Process

Tom Bohman*

March 24, 2009

Abstract

Consider the following stochastic graph process. We begin with G_0, the empty graph on n vertices, and form G_i by adding a randomly chosen edge e_i to G_{i-1} where e_i is chosen uniformly at random from the collection of pairs of vertices that neither appear as edges in G_{i-1} nor form triangles when added as edges to G_{i-1}. Let the random variable M be the number of edges in the maximal triangle free graph generated by this process. We prove that asymptotically almost surely $M = \Theta(n^{3/2}\sqrt{\log n})$. This resolves a conjecture of Spencer. Furthermore, the independence number of G_M is asymptotically almost surely $\Theta(\sqrt{n \log n})$, which implies that the Ramsey number $R(3,t)$ is bounded below by a constant times $t^2/\log t$ (a fact that was previously established by Jeong Han Kim). The methods introduced here extend to the K_4-free process, thereby establishing the bound $R(4,t) = \Omega(t^{5/2}/\log^2 t)$.

1 Introduction

Consider the following constrained random graph process. We begin with the empty graph on n vertices, which we denote G_0. At step i we form the graph G_i by adding an edge to G_{i-1} chosen uniformly at random from the collection of pairs of vertices that neither appear as edges in G_{i-1} nor form triangles (i.e. copies of K_3) when added as edges to G_{i-1}. The process terminates with a maximal triangle-free graph on n vertices, which we denote G_M (thus the random variable M is the number of steps in the process). We are interested in the likely structural properties of G_M as n tends to infinity; for example, we would like to know the value of M and the independence number of G_M.

The first result on a process which iteratively adds edges chosen uniformly at random from the collection of potential edges that maintain some graph property is due to Ruciński and Wormald, who answered a question of Erdős regarding the process in which we maintain a bound on the maximum degree [21]. Erdős, Suen and Winkler considered both the triangle-free process (but this process was studied earlier, if not in print, see [7]) and the odd-cycle-free process [12]. The H-free process where H is a fixed graph was treated by Bollobás and Riordan [8] as well as Osthus and Taraz [19]. While these papers establish interesting bounds on the likely number edges in the graph produced by the H-free process for some graphs H, for no graph H that contains a cycle has the exact order of magnitude been determined.

*Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA, USA. Supported in part by NSF grant DMS 0701183. E-mail: tbohman@math.cmu.edu

Another motivation for the triangle-free process comes from Ramsey theory. The Ramsey number $R(k, \ell)$ is the minimum integer n such that any graph on n vertices contains a clique on k vertices or an independent set on ℓ vertices. The Ramsey numbers play a central role in combinatorics and are the subject of many notoriously difficult problems, most of which remain widely open (see, for example, [13] [9]). One problem regarding the Ramsey numbers that has been resolved is the order of magnitude of $R(3, t)$ as t tends to infinity: Ajtai, Komlós and Szemerédi proved the upper bound $R(3, t) = O(t^2/\log t)$ and Kim established the lower bound $R(3, t) = \Omega(t^2/\log t)$. (There were a number of significant steps over the course of about 30 years that led up to these final results; see [10], [11], [14], [23], [22], [12], [18], [24].) The problem of determining the asymptotic behavior of $R(3, t)$ was one of the motivations for the introduction of the triangle-free process. Indeed, we establish a lower bound of the form $R(3, t) > n$ by proving the existence of a graph on n vertices with neither a triangle nor an independent set on t vertices, and the triangle-free process should produce such a graph as it should include enough 'random' edges to eliminate all large independent sets. In a certain sense, Kim's celebrated result verified this intuition as he used a semi-random variation on the triangle-free process. (This was an application of the powerful Rödl nibble that was inspired by an approach to the triangle-free process proposed by Spencer [24].) However, the problem of whether or not the triangle-free process itself is likely to produce a Ramsey $R(3, t)$ graph remained open.

Our main results (Theorems 4 and 5 below) have the following Corollaries.

Theorem 1. *Let the random variable M be the number of edges in the graph on n vertices formed by the triangle-free process. There are constants c_1, c_2 such that asymptotically almost surely we have*

$$c_1\sqrt{\log n} \cdot n^{3/2} \leq M \leq c_2\sqrt{\log n} \cdot n^{3/2}.$$

Theorem 2. *There is a constant c_3 such that the following holds: If $n = n(t) < c_3 \cdot t^2/\log t$ then a.a.s. the triangle-free process on n vertices produces a graph with no independent set of cardinality t. Thus $R(3, t) \geq c_3 \cdot \frac{t^2}{\log t}$ for t sufficiently large.*

Theorem 1 proves a conjecture of Spencer [24]. Theorem 2, which establishes that the triangle-free process is an effective randomized algorithm for producing a Ramsey $R(3, t)$ graph, is a direct consequence of Theorem 5 below. These results reveal a visionary aspect of the 1961 paper of Erdős [10] which established the bound $R(3, t) = \Omega(t^2/\log^2 t)$. When the probabilistic method was in its infancy, Erdős established his bound by analyzing a greedy algorithm applied to a random graph, and it turns out that a random greedy algorithm produces a Ramsey $R(3, t)$ graph. For an explicit construction of a triangle-free graph on $\Theta(t^{3/2})$ vertices with independence number less than t see Alon [2].

The methods introduced here can be applied to other processes. In fact, our methods immediately suggest an approach to the H-free process for general H. This applies to hypergraph processes as well. As an example, we analyze the K_4-free process to prove the following result:

Theorem 3. *There is a constant c_4 such that for t sufficiently large we have*

$$R(4, t) > c_4 \cdot \frac{t^{5/2}}{\log^2 t}.$$

This is a minor improvement on the previously best known lower bound, $\Omega\left((t/\log t)^{5/2}\right)$, which was established by Spencer via an application of the Lovász Local Lemma [23]. Alon, Ben-Shimon and Krivelevich recently gave a heuristic for a construction that, given a K_s-free graph on n vertices that is nearly regular, produces a K_s-free graph on $2n$ vertices that is is both regular and has independence number at most twice the independence number of the original graph [3]. They showed that this construction can be applied to the graph given by the K_3-free process to produce a *regular* Ramsey $R(3,t)$ graph. (Their construction requires a bound on the degrees in the graph produced by the K_3-free process that is not formally proved here. However, this bound follows directly from the methods introduced; in particular, it is a simple refinement of Claim 11.)

We analyze the triangle-free process by an application of the so-called differential equations method for random graph processes (see Wormald [26] for an introduction to the method). The main idea is to identify a collection of random variables whose one-step expected changes can be written in terms of the random variables in the collection. These expressions yield an autonomous system of ordinary differential equations, and we prove that the random variables in our collection (appropriately scaled) are tightly concentrated around the trajectory given by the solution of the o.d.e. Recent applications of this method include results that link the emergence of a giant component in a random graph process to a blow-up point in an associated o.d.e. [6], [25], [4] and an analysis of a randomized matching algorithm that hinges on the existence of an invariant set in an associated o.d.e. [5].

We track the following random variables through the evolution of the triangle-free process. Recall that G_i is the graph given by the first i edges selected by the process. The graph G_i partitions $\binom{[n]}{2}$ into three parts: E_i, O_i and C_i. The set E_i is simply the edge set of G_i. A pair $\{u,v\} \in \binom{[n]}{2}$ is **open**, and in the set O_i, if it can still be added as an edge without violating the triangle-free condition. A pair $\{u,v\} \in \binom{[n]}{2}$ is **closed**, and in the set C_i, if it is neither an edge in the graph nor open; that is, the pair $\{u,v\}$ is in C_i if there some vertex w such that $\{u,w\},\{v,w\} \in E_i$. Note that e_{i+1} is chosen uniformly at random from O_i. Set $Q(i) = |O_i|$; this is one of the random variables we track. For each pair $\{u,v\} \in \binom{[n]}{2}$ we track three random variables. Let $X_{u,v}(i)$ be the set of vertices w such that $\{u,w\},\{v,w\} \in O_i$. Let $Y_{u,v}(i)$ be the set of vertices w such that

$$|\{\{u,w\},\{v,w\}\} \cap O_i| = |\{\{u,w\},\{v,w\}\} \cap E_i| = 1.$$

Finally, let $Z_{u,v}(i)$ be the set of vertices w such that $\{u,w\},\{v,w\} \in E_i$. Note that if $Z_{u,v}(i) \neq \emptyset$ then we have $\{u,v\} \in C_i$. We dub vertices in $X_{u,v}$ **open** with respect to $\{u,v\}$, vertices in $Y_{u,v}$ **partial** with respect to $\{u,v\}$ and vertices in $Z_{u,v}$ **complete** with respect to $\{u,v\}$. We track the variables $|X_{u,v}(i)|$, $|Y_{u,v}(i)|$ and $|Z_{u,v}(i)|$ for all pairs $\{u,v\}$ such that $\{u,v\} \notin E_i$. (In fact, we only show that $|Z_{u,v}(i)|$ does not get too large; so we track this random variable in the sense that we bound it). We emphasize that we make no claims regarding the number of open, partial and complete vertices with respect to pairs $\{u,v\}$ that are edges in the graph. Formally, we set $X_{u,v}(i) = X_{u,v}(i-1)$, $Y_{u,v}(i) = Y_{u,v}(i-1)$ and $Z_{u,v}(i) = Z_{u,v}(i-1)$ if $\{u,v\} \in E_i$.

In order to motivate our main results (Theorems 4 and 5 below) we present a heuristic derivation of the trajectory that the random variables $Q(i), |X_{u,v}(i)|$ and $|Y_{u,v}(i)|$ should follow. We stress that this discussion does not constitute a proof that the random variables follow this trajectory; the proof itself comes in Section 3 below. We begin by choosing appropriate scaling. We introduce a continuous variable t and relate this to the steps G_i in the process by setting $t = t(i) = i/n^{3/2}$. Our trajectories are given by three functions: $q(t), x(t)$ and $y(t)$. We suppose $Q(i)$ is approximately $q(t)n^2$, $|X_{u,v}(i)|$ is approximately $x(t)n$ for all $\{u,v\} \in \binom{[n]}{2} \setminus E_i$ and $|Y_{u,v}(i)|$ is approximately $y(t)\sqrt{n}$ for all $\{u,v\} \in \binom{[n]}{2} \setminus E_i$. Consider a fixed step i in the graph process and let $\epsilon > 0$ be sufficiently small. We suspect that the changes in our tracked random variables are very close to their expected values over the ensuing $\epsilon n^{3/2}$ steps of the process and use this guess to derive our system of differential equations. We begin with $|O_i|$. Note that if $e_{i+1} = \{u,v\}$ then there is exactly one edge closed for each vertex that is partial with respect to $\{u,v\}$; in other words, if $e_{i+1} = \{u,v\}$ then $Q(i+1) = Q(i) - 1 - |Y_{u,v}(i)|$. Therefore, we should have

$$q(t+\epsilon)n^2 \approx Q\left(i + \epsilon n^{3/2}\right) \approx Q(i) - \epsilon n^{3/2} \cdot y(t)n^{1/2} \approx \left(q(t) - \epsilon y(t)\right) n^2.$$

This suggests $dq/dt = -y$. Now consider the variable $|X_{u,v}(i)|$. Consider a fixed vertex w that is open with respect to $\{u,v\}$. Note that the probability that the edge e_{i+1} closes $\{u,w\}$ (i.e. the probability of the event $\{u,w\} \in C_{i+1}$) is $|Y_{u,w}|/|O_i|$. As the probability that $e_{i+1} \in \{\{u,w\},\{v,w\}\}$ is comparatively negligible, we suspect that we have

$$\begin{aligned} x(t+\epsilon)n &\approx \left|X_{u,v}\left(i+\epsilon n^{3/2}\right)\right| \\ &\approx |X_{u,v}(i)| - \epsilon n^{3/2} \cdot x(t)n \frac{2y(t)\sqrt{n}}{q(t)n^2} \\ &\approx \left(x(t) - \epsilon \frac{2x(t)y(t)}{q(t)}\right) n. \end{aligned}$$

This suggests $dx/dt = -2xy/q$. Finally, we consider $|Y_{u,v}(i)|$. First note that a vertex that is partial with respect to $\{u,v\}$ has its one open edge closed by e_{i+1} with probability nearly $y(t)\sqrt{n}/(q(t)n^2)$. The probability that a vertex that is open with respect to $\{u,v\}$ becomes partial with respect to $\{u,v\}$ is $2/Q(i)$. So, we should have

$$\begin{aligned} y(t+\epsilon)\sqrt{n} &\approx \left|Y_{u,v}\left(i+\epsilon n^{3/2}\right)\right| \\ &\approx |Y_{u,v}(i)| - \epsilon n^{3/2} \cdot y(t)\sqrt{n} \cdot \frac{y(t)\sqrt{n}}{q(t)n^2} + \epsilon n^{3/2} \cdot \frac{2x(t)n}{q(t)n^2} \\ &\approx \left(y(t) - \epsilon \frac{y^2(t)}{q(t)} + \epsilon \frac{2x(t)}{q(t)}\right)\sqrt{n}, \end{aligned}$$

which suggests $dy/dt = -y^2/q + 2x/q$. As $|O_0| = n(n-1)/2$, $|X_{u,v}(0)| = n-2$ for all pairs $\{u,v\}$ and $|Y_{u,v}(0)| = 0$ for all pairs $\{u,v\}$, our expected value computations suggest that our random variables should follow the trajectory given by

$$\frac{dq}{dt} = -y \qquad \frac{dx}{dt} = -\frac{2xy}{q} \qquad \frac{dy}{dt} = -\frac{y^2}{q} + \frac{2x}{q} \tag{1}$$

with initial conditions $q(0) = 1/2$, $x(0) = 1$ and $y(0) = 0$. The solution to this autonomous system is

$$q(t) = \frac{e^{-4t^2}}{2} \qquad x(t) = e^{-8t^2} \qquad y(t) = 4te^{-4t^2}. \tag{2}$$

Note that if $|O_i|$ indeed follows $q(t)n^2$ then the triangle-free process will come to end at $t = \Theta(\sqrt{\log n})$; that is, the process will end with $\Theta(\sqrt{\log n} \cdot n^{3/2})$ edges. (It was Peter Keevash who pointed out that (1) has this tantalizing solution [16].) Observe that the functions $q(t), x(t), y(t)$ are the appropriate values for G chosen uniformly at random from the collection of graphs with n vertices and $tn^{3/2}$ edges. Thus, a natural interpretation of our results is that the evolution of the triangle-free process is the same as the evolution of the unconstrained random graph with respect to these variables.

We introduce absolute constants μ, β, γ and ρ. The constants μ and ρ are small, β is a large relative to μ and γ is large relative to both μ and β. (These constant can take values $\mu = \rho = 1/32$, $\beta = 1/2$ and $\gamma = 161$. No effort is made to optimize the constants, and we do not introduce the actual values in an attempt to make the paper easier to read). Set

$$m = \mu\sqrt{\log n} \cdot n^{3/2}.$$

Our first result is that our random variables indeed follow the trajectory (2) up to m random edges. In order to state this concentration result we introduce error functions that slowly deteriorate as the process evolves. Define

$$f_q(t) = \begin{cases} e^{41t^2+40t} & \text{if } t \leq 1 \\ \frac{e^{41t^2+40t}}{t} & \text{if } t > 1 \end{cases} \qquad f_x(t) = e^{37t^2+40t} \qquad f_y(t) = e^{41t^2+40t}, \tag{3}$$

and set

$$g_q(t) = f_q(t)n^{-1/6} \qquad g_x(t) = f_x(t)n^{-1/6} \qquad g_y(t) = f_y(t)n^{-1/6}.$$

Let $\mathcal{B}_j$ be the event that there exists there exists a step $i \leq j$ such that

$$\left|Q(i) - q(t)n^2\right| \geq g_q(t)n^2$$

or there exists some pair $\{u, v\} \in \binom{[n]}{2} \setminus E_i$ such that

$$||X_{u,v}(i)| - x(t)n| \geq g_x(t)n \quad \text{or} \quad \left||Y_{u,v}(i)| - y(t)\sqrt{n}\right| \geq g_y(t)\sqrt{n} \quad \text{or} \quad |Z_{u,v}(i)| \geq \log^2 n.$$

Theorem 4. *If n is sufficiently large then*

$$Pr\left(\mathcal{B}_{\mu\sqrt{\log n}\cdot n^{3/2}}\right) \leq e^{-\log^2 n}.$$

Note that Theorem 4 alone places no upper bound on the number M of edges in the graph produced by the triangle-free process. In order to achieve such a bound, we bound the independence number of G_m.

Theorem 5. *If n is sufficiently large then*

$$Pr\left(\alpha\left(G_{\mu\sqrt{\log n}\cdot n^{3/2}}\right) > \gamma\sqrt{n\log n} \mid \overline{\mathcal{B}_m}\right) < e^{-n^{1/5}}.$$

Since the neighborhood of each vertex in the triangle-free process is an independent set, it follows immediately from Theorem 5 that the maximum degree in G_M is at most $\gamma\sqrt{n\log n}$ a.a.s. Thus, we have proved Theorem 1.

The remainder of the paper is organized as follows. In the next section we establish some technical preliminaries. Theorems 4 and 5 are then proved in Sections 3 and 4, respectively. The proof of Theorem 3 is given in Section 5.

2 Preliminaries

Our probability space is the space defined naturally by the triangle-free process. Let $\Omega = \Omega_n$ be the set of all maximal sequences in $\binom{[n]}{2}^*$ with distinct entries and the property that each initial sequence gives a triangle-free graph on vertex set $[n]$. We stress that our measure is not uniform: it is the measure given by the uniform random choice at each step. We always work with the natural filtration $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots$ given by the process. Two elements x, y of Ω are in the same part of the partition that generates $\mathcal{F}_j$ iff the first j entries of x and y agree. We use the symbol ω_j to denote one of the parts in this partition (i.e. ω_j denotes a particular history of the process through j steps); in particular, if $\omega \in \Omega$ then ω_j is the part of the partition that defines $\mathcal{F}_j$ that contains ω.

For the purpose of notational convenience we use the symbol '$\pm$' in two ways: in interval arithmetic and to define pairs of random variables. The distinction between the two should be clear from context. The degree of a vertex v in G_i is denoted $d_i(v)$ and the neighborhood of v in G_i is $N_i(v)$.

Our main tool for establishing concentration is the following version of the Azuma-Hoeffding inequality. Let $\eta, N > 0$ be constants. We say that a sequence of random variables $A_0, A_1, \ldots$ is (η, N)**-bounded** if

$$A_i - \eta \leq A_{i+1} \leq A_i + N \quad \text{for all } i.$$

Lemma 6. *Suppose $\eta \leq N/2$ and $a < \eta m$. If $0 \equiv A_0, A_1, \ldots$ is an (η, N)-bounded submartingale then*

$$Pr[A_m \leq -a] \leq e^{-\frac{a^2}{3\eta m N}}.$$

Lemma 7. *Suppose $\eta \leq N/10$ and $a < m\eta$. If $0 \equiv A_0, A_1, \ldots$ is an (η, N)-bounded supermartingale then*

$$Pr[A_m \geq a] \leq e^{-\frac{a^2}{3\eta m N}}.$$

As these particular inequalities seem not to appear in the literature, proofs are given at the end of the paper, in Section 6. We often work with pairs $A_0^{\pm}, A_1^{\pm}, \ldots$ where $A_0^+, A_1^+, \ldots$ is an (η, N)-bounded submartingale and $A_0^-, A_1^-, \ldots$ is an (η, N)-bounded supermartingale. We will refer to such a pair of sequences of random variables as an (η, N)**-bounded martingale pair**.

3 Trajectory

Here we prove Theorem 4, which establishes tight concentration of the random variables $|O_i|$, $|X_{u,v}(i)|$ and $|Y_{u,v}(i)|$ around the trajectory given in (2) and bounds $|Z_{u,v}(i)|$.

Recall $t = t(i) = i/n^{3/2}$ and $m = \mu\sqrt{\log n} \cdot n^{3/2}$ and

$$g_q(t) = \begin{cases} e^{41t^2+40t} n^{-1/6} & \text{if } t \leq 1 \\ \frac{e^{41t^2+40t}}{t} n^{-1/6} & \text{if } t > 1 \end{cases} \qquad g_x(t) = e^{37t^2+40t} n^{-1/6} \qquad g_y(t) = e^{41t^2+40t} n^{-1/6}.$$

Note that

$$g_q \leq \frac{g_y}{t} \qquad \text{and} \qquad g_x = e^{-4t^2} g_y. \tag{4}$$

We define events $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$. For $\omega \in \mathcal{B}_m$ let ℓ be the smallest index such that $\omega \in \mathcal{B}_\ell$ but $\omega \notin \mathcal{B}_{\ell-1}$; in other words, the random variable ℓ is the first time that one of our tracked random variables is outside the allowable range. We define $\mathcal{X}$ to be the set of $\omega \in \mathcal{B}_m$ such that there exists a pair $\{u, v\}$ such that $\{u, v\} \notin E_\ell$ and

$$|X_{u,v}(\ell)| \notin n\left[x(t(\ell)) \pm g_x(t(\ell))\right].$$

So, an atom $\omega \in \mathcal{B}_m$ is in $\mathcal{X}$ if there is some pair of vertices $\{u, v\}$ such that the number of open vertices with respect to $\{u, v\}$ is a reason we place $\omega \in \mathcal{B}_\ell$. Define $\mathcal{Y}$ and $\mathcal{Z}$ analogously. We prove Theorem 4 by showing

$$\mathcal{B}_m = \mathcal{X} \cup \mathcal{Y} \cup \mathcal{Z} \tag{5}$$

and then bounding the probabilities of $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$. In the next subsection we show that if $|Y_{u,v}(j)|$ is in range for all $j \leq i$ and all pairs $\{u, v\}$ then $|O_i|$ is in range, thereby establishing (5). In the following three subsections we establish upper bounds on the probabilities of the events $\mathcal{X}, \mathcal{Y}$ and $\mathcal{Z}$, respectively.

3.1 Open edges

Here we simply take advantage of the strict control we enforce on the number of partial vertices at each pair; we do not invoke any concentration inequalities in this subsection. Note that if we have $e_{i+1} = \{u, v\}$ then the number of edges closed when we add e_{i+1} is simply equal to $|Y_{u,v}(i)|$, the number of partial vertices at $\{u, v\}$. Therefore, assuming $\omega \notin \mathcal{B}_{j-1}$, we have

$$\begin{aligned}
|O_j| &= \frac{n(n-1)}{2} - j - \sum_{i=0}^{j-1} \left|Y_{e_{i+1}}(i)\right| \\
&\in \frac{n^2}{2} - \frac{n}{2} - j - \sqrt{n}\left[\sum_{i=0}^{j-1} y(t) \pm g_y(t)\right] \\
&\subseteq \frac{n^2}{2} - n^2 \int_0^{t(j)} 4\tau e^{-4\tau^2} d\tau \pm n^{11/6} \int_0^{t(j)} e^{41\tau^2 + 40\tau} d\tau \pm n^{5/3} \\
&\subseteq n^2\left[q(t(j)) \pm g_q(t(j))\right].
\end{aligned}$$

Note that this establishes (5).

3.2 Open vertices

Consider a fixed $\{u, v\} \in \binom{[n]}{2}$. We write

$$|X_{u,v}(j)| = n - 2 - \sum_{i=1}^{j} A_i$$

where A_i is the number of open vertices at $\{u,v\}$ that are eliminated when e_i is added to the process. Define A_i^+ and A_i^- by

$$A_i^{\pm} = \begin{cases} A_i + \frac{1}{\sqrt{n}}\left[-\frac{2x(t)y(t)}{q(t)} \pm (17t+39)g_x(t)\right] & \text{if } \omega \notin \mathcal{B}_{i-1} \text{ and } \{u,v\} \notin E_i \\ 0 & \text{if } \omega \in \mathcal{B}_{i-1} \text{ or } \{u,v\} \in E_i \end{cases}$$

$$B_j^{\pm} = \sum_{i=1}^{j} A_i^{\pm}.$$

Note that if $\omega \notin \mathcal{B}_{j-1}$ and $\{u,v\} \notin E_j$ then we have

$$\begin{aligned} |X_{u,v}(j)| &= n - 2 - \sum_{i=1}^{j} A_i^+ - \sum_{i=1}^{j}\left[\frac{2x\,(t)\,y(t)}{q(t)} - (17t+39)g_x(t)\right]\frac{1}{\sqrt{n}} \\ &\le n - B_j^+ - n\int_0^{t(j)} \frac{2x(\tau)y(\tau)}{q(\tau)}d\tau + n^{5/6}\int_0^{t(j)} (17\tau+39)e^{37\tau^2+40\tau}d\tau \\ &\le nx(t(j)) + n^{5/6}\left(e^{37t^2(j)+40t(j)} - 1\right) - B_j^+ \\ &= n\left[x(t(j)) + g_x(t(j))\right] - \left(B_j^+ + n^{5/6}\right). \end{aligned}$$

Therefore, the event $|X_{u,v}(j)| > n\left[x(t(j)) + g_x(t(j))\right]$ is contained in the event $B_j^+ < -n^{5/6}$. Similarly, the event $|X_{u,v}(j)| < n\left[x(t(j)) - g_x(t(j))\right]$ is contained in the event $B_j^- > n^{5/6}$. We bound the probabilities of these events by application of the martingale inequalities.

Claim 8. *$B_0^{\pm}, B_1^{\pm}, \ldots$ is a $(\frac{4}{\sqrt{n}}, \sqrt{n})$-bounded martingale pair.*

Proof. We begin with the martingale condition. Of course, we can restrict our attention to ω_i such that $\omega_i \not\subseteq \mathcal{B}_i$ and $\{u,v\} \notin E_i$. Consider a vertex $w \in X_{u,v}(i)$. Note that $w \notin X_{u,v}(i+1)$ if $e_{i+1} \in \{\{u,w\},\{v,w\}\}$, e_{i+1} connects $\{u,w\}$ to one of the vertices that is partial at $\{u,w\}$ or e_{i+1} connects $\{v,w\}$ to one of the vertices that is partial at $\{v,w\}$. Note that (as we assume $\{u,v\} \notin E_i$) the edge e_{i+1} plays 2 of these roles if and only if $e_{i+1} = \{z,w\}$ where $z \in Z_{u,v}(i)$. It follows that we have

$$Pr\left(w \notin X_{u,v}(i+1)\right) = \frac{2 + |Y_{u,w}(i)| + |Y_{v,w}(i)| - |Z_{u,v}(i)|}{|O_i|},$$

and therefore

$$E[A_{i+1} \mid \mathcal{F}_i] = \frac{1}{|O_i|}\left[\sum_{w \in X_{u,v}(i)} 2 + |Y_{u,w}(i)| + |Y_{v,w}(i)| - |Z_{u,v}(i)|\right].$$

As we restrict our attention to $\omega_i \not\subseteq \mathcal{B}_i$, we have

$$\begin{aligned}
E[A_{i+1} \mid \mathcal{F}_i] &\in \frac{2n^{3/2}(x \pm g_x)(y \pm g_y)}{n^2(q \pm g_q)} + \left(-\frac{n(x+g_x)\log^2 n}{n^2(q-g_q)}, \frac{2n(x+g_x)}{n^2(q-g_q)}\right) \\
&\subseteq \frac{1}{\sqrt{n}}\left[\frac{2xy}{q} \pm \left(\frac{2g_y x + 2g_x y + 2g_x g_y}{q-g_q} + \frac{2xyg_q}{q(q-g_q)}\right)\right] + \left(-\frac{\log^2 n}{n}\left[\frac{4x}{q}\right], \frac{1}{n}\frac{4x}{q}\right) \\
&\subseteq \frac{1}{\sqrt{n}}\left[\frac{2xy}{q} \pm \left(5e^{-4t^2}g_y + 17tg_x + 5g_x g_y e^{4t^2} + 33te^{-4t^2}g_q\right)\right] \pm \frac{\log^2 n}{n}4e^{-4t^2} \\
&\subseteq \frac{1}{\sqrt{n}}\left[\frac{2xy}{q} \pm (17tg_x + 39g_x)\right].
\end{aligned}$$

(Note that we apply (4).) This establishes the martingale condition.

Now we turn to the bounds on $A_{i+1}^{\pm}$. We use the simple fact that the set of edges closed when we add e_{i+1} is determined by $Y_{e_{i+1}}(i)$; one edge in each partial triangle in $Y_{e_{i+1}}(i)$ is closed. Therefore, the maximum value of A_i is bounded above by $(y(t)+g_y(t))\sqrt{n}$, which is at most $\sqrt{n}$. Of course $A_i^{\pm}$ takes its smallest value when $A_i = 0$, and in this case we have $A_i^{\pm} > -4/\sqrt{n}$ as $2xy/q \leq 4/\sqrt{e}$. □

Applying Lemmas 6 and 7 we have

$$Pr(B_m^+ < -n^{5/6}), Pr(B_m^- > n^{5/6}) \leq e^{-\frac{n^{5/3}}{12m}}. \tag{6}$$

We claim that $\mathcal{X}$ is contained in the union, taken over all pairs $\{u,v\}$, of the events given in (6). Indeed, if $\omega \in \mathcal{X}$ on account of the the pair $\{u,v\}$ at step ℓ then either $B_\ell^+ < -n^{5/6}$ or $B_\ell^- > n^{5/6}$ and we also have $B_j^+ = B_\ell^+$ and $B_j^- = B_\ell^-$ for all $j \geq \ell$ (as we set $A_{j+1}^{\pm} = 0$ in the event $\mathcal{B}_j$). Therefore, we have

$$Pr(\mathcal{X}) \leq 2\binom{n}{2}e^{-\frac{n^{5/3}}{12m}}.$$

3.3 Partial vertices

We use the same reasoning as in the last subsection, but here we break the step by step changes in $|Y_{u,v}(i)|$ into two parts. We write $|Y_{u,v}(j)|$ as a sum

$$|Y_{u,v}(j)| = \sum_{i=1}^{j} U_i - V_i,$$

where U_i is the number of partial vertices at $\{u,v\}$ created when e_i is added and V_i is the number of partial vertices at $\{u,v\}$ eliminated when e_i is added. Note that if $\{u,v\} \in E_i$ then we set $U_i = V_i = 0$ (in order to maintain consistency with the definition of $Y_{u,v}$).

We begin with an analysis of V_i. Define $W_0^{\pm} = 0$ and

$$V_i^{\pm} = \begin{cases} V_i + \left[-\frac{y^2(t)}{q(t)} \pm (82t+1)g_y(t)\right]\frac{1}{n} & \text{if } \omega \notin \mathcal{B}_{i-1} \text{ and } \{u,v\} \notin E_i \\ 0 & \text{if } \omega \in \mathcal{B}_{i-1} \text{ or } \{u,v\} \in E_i \end{cases}$$

$$W_j^{\pm} = \sum_{i=1}^{j} V_i^{\pm}.$$

Claim 9. *$W_0^\pm, W_1^\pm, \ldots$ is a $(\frac{4}{n}, \log^2 n)$-bounded martingale pair.*

Proof. We begin with the martingale conditions. Suppose w is partial with respect to $\{u,v\}$. Let w^* be the unique vertex in $\{u,v\}$ such that $\{w^*,w\} \in O_i$. Note that w is removed from $X_{u,v}$ if either $e_{i+1} = \{w, w^*\}$ or e_{i+1} is one of the pairs in O_i that links $\{w^*, w\}$ to $Y_{w^*,w}(i)$ (other than $\{u,v\}$ itself). Therefore, restricting our attention to $\omega_i \not\subseteq \mathcal{B}_i$, we have

$$E\left[V_{i+1} \mid \mathcal{F}_i\right] = \sum_{w \in Y_{u,v}} \frac{|Y_{w^*,w}|}{|O_i|}.$$

As we restrict our attention to $\omega_i \not\subseteq \mathcal{B}_i$ and $\{u,v\} \notin E_i$ we have

$$\begin{aligned} E\left[V_{i+1} \mid \mathcal{F}_i\right] &\in \frac{\sqrt{n}(y \pm g_y)\left(\sqrt{n}(y \pm g_y)\right)}{n^2(q \pm g_q)} \\ &\subseteq \frac{1}{n}\left[\frac{y^2}{q} \pm \left(\frac{2g_y y + g_y^2}{q - g_q} + \frac{y^2 g_q}{q(q-g_q)}\right)\right] \\ &\subseteq \frac{1}{n}\left[\frac{y^2}{q} \pm \left(17tg_y + g_y + 65tg_y\right)\right], \end{aligned}$$

and the martingale conditions are established.

It remains to establish boundedness. Note that if e_{i+1} does not intersect $\{u,v\}$ then the change in $|Y_{u,v}(i)|$ is at most 2 (as all edges that are closed when we add e_{i+1} intersect e_{i+1}). So, suppose $e_{i+1} = \{u, z\}$ where $z \neq v$. If the vertex w is then removed from $Y_{u,v}$ then the edge $\{w,u\}$ must have been closed by e_{i+1}. This implies $\{w,z\} \in E_i$. Furthermore, as w is partial with respect to $\{u,v\}$, we have $\{w,v\} \in E_i$. Thus $w \in Z_{z,v}(i)$. Therefore, the change in V_i is bounded by the maximum value of $|Z_{x,y}(i)|$, which is bounded by $\log^2 n$. The lower bound follows from $y^2/q \leq 8/e$. ◻

Applying Lemmas 6 and 7 we have

$$Pr(W_m^+ < -n^{1/3}/3), Pr(W_m^- > n^{1/3}/3) \leq e^{-\frac{n^{2/3}}{108m\log^2 n/n}}.$$

Now we turn to U_i. Define $T_0^\pm = 0$ and

$$U_i^\pm = \begin{cases} U_i + \left[-\frac{2x(t)}{q(t)} \pm 14g_y(t)\right]\frac{1}{n} & \text{if } \omega \notin \mathcal{B}_{i-1} \text{ and } \{u,v\} \notin E_i \\ 0 & \text{if } \omega \in \mathcal{B}_{i-1} \text{ or } \{u,v\} \in E_i \end{cases}$$

$$T_j^\pm = \sum_{i=1}^{j} U_i^\pm.$$

Claim 10. *$T_0^\pm, T_1^\pm, \ldots$ is a $(\frac{5}{n}, 1)$-bounded martingale pair.*

Proof. We begin with the martingale conditions. As usual we restrict our attention to $\omega_i \not\subseteq \mathcal{B}_i$. We have

$$E\left[U_{i+1} \mid \mathcal{F}_i\right] = \frac{2|X_{u,v}(i)|}{|O_i|},$$

and

$$
\begin{aligned}
E\left[U_{i+1} \mid \mathcal{F}_i\right] &\in \frac{2n(x \pm g_x)}{n^2(q \pm g_q)} \\
&\subseteq \frac{1}{n}\left[\frac{2x}{q} \pm \frac{2g_q x + 2q g_x}{q(q-g_q)}\right] \\
&\subseteq \frac{1}{n}\left[\frac{2x}{q} \pm \left(9g_q + 5e^{4t^2} g_x\right)\right] \\
&\subseteq \frac{1}{n}\left[\frac{2x}{q} \pm 14 g_y\right],
\end{aligned}
$$

which establishes the martingale conditions.

As the addition of e_{i+1} to the graph can create at most one new partial vertex at $\{u, v\}$, U_i is either 1 or 0. Furthermore, $2x/q = 4e^{-4t^2} \leq 4$. These two observations establish the boundedness condition. $\square$

Applying Lemmas 6 and 7 we have

$$
Pr(T_m^+ < -n^{1/3}/3), Pr(T_m^- > n^{1/3}/3) \leq e^{-\frac{n^{2/3}}{135m/n}}
$$

Now we are ready to return to the random variable $|Y_{u,v}(i)|$ itself. We have

$$
\begin{aligned}
|Y_{u,v}(j)| &= \sum_{i=1}^{j} U_i - V_i \\
&= \sum_{i=1}^{j} U_i^+ + \frac{1}{n}\left[\frac{2x}{q} - 14g_y\right] - \left(\sum_{i=1}^{j} V_i^- + \frac{1}{n}\left[\frac{y^2}{q} + (82t+1)g_y\right]\right) \\
&= T_j^+ - W_j^- + \frac{1}{n}\sum_{i=1}^{j}\left(\frac{2x}{q} - \frac{y^2}{q}\right) - \frac{1}{n}\sum_{i=1}^{j}(82t+15)g_y \\
&\geq \sqrt{n}\int_0^{t(j)} \frac{2x}{q} - \frac{y^2}{q} d\tau - n^{1/3}\int_0^{t(j)} (82\tau+15)e^{41\tau^2+40\tau} d\tau + \left(T_j^+ - W_j^-\right) - 1/2 \\
&\geq \sqrt{n}\left[y(t(j)) - g_y(t(j))\right] + n^{1/3} + T_j^+ - W_j^- - 1/2
\end{aligned}
$$

Therefore, the event $|Y_{u,v}(j)| < \sqrt{n}\left[y(t(j)) - g_y(t(j))\right]$ is contained in

$$
\left\{T_j^+ < -n^{1/3}/3\right\} \vee \left\{W_j^- > n^{1/3}/3\right\}.
$$

We have already bounded the probabilities of these events. The analogous argument holds for the event $|Y_{u,v}(j)| > \sqrt{n}(y(t(j)) + g_y(t(j)))$, with T_j^+ and W_j^- replaced with T_j^- and W_j^+, respectively. As the random variables $T_i^{\pm}$ and $W_i^{\pm}$ are 'frozen' once one of the random variables leaves the allowable range, we have

$$
Pr(\mathcal{Y}) \leq \binom{n}{2} \cdot 2\left(e^{-\frac{n^{2/3}}{48m\log^2 n/n}} + e^{-\frac{n^{2/3}}{60m/n}}\right) < 2n^2 e^{-n^{1/6}/48}.
$$

3.4 Complete vertices

Note that the probability that e_{i+1} adds a complete vertex at $\{u,v\}$ is at most $|Y_{u,v}(i)|/|O_i|$. So, in the event $\mathcal{B}_i$, we have

$$Pr\left(|Z_{u,v}(i+1)| = |Z_{u,v}(i)|+1\right) \leq \frac{\sqrt{n}(y(t)+g_y(t))}{n^2(q(t)-g_q(t))} \leq \frac{9t}{n^{3/2}}.$$

Therefore,

$$Pr\left[|Z_{u,v}(m)| \geq \log^2 n\right] \leq \binom{\mu n^{3/2}\sqrt{\log n}}{\log^2 n}\left(\frac{9\mu\sqrt{\log n}}{n^{3/2}}\right)^{\log^2 n} \leq e^{-\frac{1}{2}(\log^2 n)\log\log n}$$

for n sufficiently large. Thus

$$Pr(\mathcal{Z}) \leq \binom{n}{2} e^{-\frac{1}{2}(\log^2 n)\log\log n}.$$

4 Independent Sets

Our goal is now to prove Theorem 5. We will bound from above the probability, conditional on $\overline{\mathcal{B}_m}$, that any fixed set K of $\gamma\sqrt{n\log n}$ vertices is independent. This bound will be so small that it remains small when multiplied by the number of such K. The conditioning on $\overline{\mathcal{B}_m}$ tells us that the variables $Q, X_{u,v}, Y_{u,v}$ all remain quite close to $q(t)n^2, x(t)n, y(t)\sqrt{n}$ throughout the process. As it happens, the strength of the error terms g_q, g_y, g_x does not play a major role in the calculations below. The reader might, at first reading, set $g_q = g_y = g_x = 0$ so as to get a less cluttered view of the techniques involved.

Recall that μ,β,γ and ρ are constants where μ and ρ are small, β is large relative to μ and γ is large relative to μ and β. Also recall $m = \mu\sqrt{\log n}\cdot n^{3/2}$. We make 2 initial observations (Claims 11 and 12). Let $\mathcal{D}_i$ be the event that G_i has a vertex of degree greater than $\beta\sqrt{n\log n}$.

Claim 11. *If n is sufficiently large then*

$$Pr(\mathcal{D}_m \wedge \overline{\mathcal{B}_m}) \leq e^{-n^{1/5}}.$$

Proof. We begin by establishing an upper bound on the number of open pairs at each vertex. For each vertex v let $W_v(i)$ be the set of pairs in O_i that contain v. Let A_i be the number of open pairs that contain v that are removed from W_v when the edge e_i is added to the process. Note that we have

$$E\left[A_{i+1} \mid \mathcal{F}_i\right] = \sum_{w\in W_v(i)} \frac{1+|Y_{v,w}(i)|}{|O_i|}.$$

Define

$$B_{i+1} = \begin{cases} A_{i+1} - \frac{1}{\sqrt{n}}\left(8te^{-4t^2} - 20g_y\right) & \text{if } W_v(i) > e^{-4t^2}n \text{ and } \omega \notin \mathcal{B}_i \\ 0 & \text{if } W_v(i) \leq e^{-4t^2}n \text{ or } \omega \in \mathcal{B}_i \end{cases}$$

Note that (restricting our attention to $\omega_i \not\subseteq \mathcal{B}_i$ and $W_v(i) > e^{-4t^2}n$)

$$\begin{aligned}
E\left[B_{i+1} \mid \mathcal{F}_i\right] &\geq e^{-4t^2} n \left(\frac{1+\sqrt{n}(y-g_y)}{n^2(q+g_q)}\right) - \frac{1}{\sqrt{n}}\left(8te^{-4t^2} - 20g_y\right) \\
&\geq \frac{e^{-4t^2}}{\sqrt{n}}\left(\frac{y}{q} - \frac{g_y}{q+g_q} - \frac{g_q y}{q(q+g_q)}\right) - \frac{1}{\sqrt{n}}\left(8te^{-4t^2} - 20g_y\right) \\
&\geq \frac{1}{\sqrt{n}}\left(-3g_y - 17tg_q + 20g_y\right) \\
&\geq 0.
\end{aligned}$$

Therefore, any sequence of the form $B_\ell, B_\ell + B_{\ell+1}, \ldots, \sum_{i=\ell}^{j} B_i, \ldots$ is a $(2/\sqrt{n}, \sqrt{n})$-bounded submartingale. Therefore, for any $\ell < j$ we have

$$Pr\left(\sum_{i=\ell}^{j} B_i \leq -n^{7/8}\right) \leq e^{-\frac{n^{7/4}}{6m}}.$$

Now consider the event $|W_v(j)| > e^{-4t(j)^2}n + 2n^{7/8}$. In this event there exists a maximum $\ell < j$ such that $|W_v(\ell)| \leq e^{-4t(\ell)^2}n$. We have

$$\sum_{i=\ell+2}^{j} A_i < \left(e^{-4t(\ell)^2} - e^{-4t(j)^2}\right)n - 2n^{7/8},$$

which implies

$$\begin{aligned}
\sum_{i=\ell+2}^{j} B_i &< \left(e^{-4t(\ell)^2} - e^{-4t(j)^2}\right)n - 2n^{7/8} - \frac{1}{\sqrt{n}}\sum_{i=\ell+1}^{j-1}\left(8te^{-4t^2} - 20g_y\right) \\
&< -\frac{3}{2}n^{7/8} + n\int_{t(\ell+1)}^{t(j)} 20g_y(\tau)d\tau \\
&< -n^{7/8}.
\end{aligned}$$

Let $\mathcal{D}'_m$ be the event that there exists a vertex v and a step $j \leq m$ such that $|W_v(j)| > e^{-4t(j)^2}n + 2n^{7/8}$. We have shown

$$Pr\left(\mathcal{D}'_m \wedge \overline{\mathcal{B}_m}\right) \leq n\binom{m}{2}\exp\left\{-n^{7/4}/(6m)\right\}.$$

So, we can restrict our attention to the event $\overline{\mathcal{D}'_j}$. Note that here we have $|W_v(j)| \leq 4|O_j|/n$ for all j, v. Now we simply use the union bound.

$$\begin{aligned}
Pr\left(\mathcal{D}_m \wedge \overline{\mathcal{B}_m} \wedge \overline{\mathcal{D}'_m}\right) &\leq n\binom{\mu\sqrt{\log n}\cdot n^{3/2}}{\beta\sqrt{n\log n}}\left(\frac{4}{n}\right)^{\beta\sqrt{\log n}\cdot n} \\
&\leq n\left(\frac{\mu\sqrt{\log n}\cdot n^{3/2}\cdot 4e}{\beta\sqrt{n\log n}\cdot n}\right)^{\beta\sqrt{n\log n}} = n\left(\frac{\mu 4e}{\beta}\right)^{\beta\sqrt{n\log n}}.
\end{aligned}$$

□

Next we consider the number of open pairs in sufficiently large bipartite subgraphs. Let A, B be disjoint subsets of $[n]$ such that

$$|A| = |B| = \left(\frac{\gamma - \beta}{2}\right)\sqrt{n \log n} = k.$$

We track the evolution of the number of pairs in O_i that intersect both A and B. Note that a vertex with large degree in either A and B can cause a large one step change in this variable. To deal with this possibility, we introduce the following definition. Let $A \times B$ be the set of pairs $\{u, v\} \in \binom{[n]}{2}$ that intersect both A and B. We say that the pair $\{u, v\} \in A \times B$ is **closed with respect to** A, B if there exists $x \notin A \cup B$ and $j \leq i$ such that

$$|N_j(x) \cap A|, |N_j(x) \cap B| \leq \frac{k}{n^\rho} \qquad \text{and} \qquad u, v \in N_j(x).$$

A pair $\{u, v\} \in A \times B$ is **open with respect to** A, B if $\{u, v\} \notin E_i$ and $\{u, v\}$ is not closed with respect to A, B. Define

$$W_{A,B}(i) = \left\{\{u, v\} \in A \times B : \{u, v\} \text{ is open with respect to } A, B \text{ in } G_i\right\}.$$

Note that a pair $\{u, v\} \in A \times B$ can be closed (i.e. in C_i) and still be in $W_{A,B}(i)$. We stop tracking $W_{A,B}$ as soon as a single edge falls in $A \cup B$; formally, if $E_i \cap \binom{A \cup B}{2} \neq \emptyset$ or $\omega_i \subseteq \mathcal{D}_i \vee \mathcal{B}_i$ then we set $W_{A,B}(i) = W_{A,B}(i-1)$.

Let $\mathcal{P}_j$ be the event there exist $A, B \in \binom{[n]}{k}$ and a step $i \leq j$ such that

$$\binom{A \cup B}{2} \cap E_i = \emptyset \quad \text{and} \quad |W_{A,B}(i)| < e^{-4t^2}k^2 - 2n^{1-\rho/3}.$$

Claim 12. *If n is sufficiently large then*

$$Pr\left(\mathcal{P}_m \wedge \overline{\mathcal{B}_m}\right) \leq e^{-n^{1/2}}.$$

Proof. Let X_i be the number of pairs that leave $W_{A,B}$ at step i of the process. We have

$$E[X_{i+1} \mid \mathcal{F}_i] \leq \sum_{\{u,v\} \in W_{A,B}} \frac{Y_{u,v}(i)}{|O_i|}.$$

Note that we only have an upper bound here as there may be edges between $\{u, v\}$ and $Y_{u,v}$ that would close $\{u, v\}$ without removing $\{u, v\}$ from $W_{A,B}$. Define

$$Y_{i+1} = \begin{cases} X_{i+1} - \frac{\log n}{\sqrt{n}}\left(\frac{(\gamma-\beta)^2}{4}\right)\left[8te^{-4t^2} + 20g_y\right] & \text{if } |W_{A,B}(i)| < e^{-4t^2}k^2 \text{ and } \omega \notin \mathcal{B}_i \\ 0 & \text{if } |W_{A,B}(i)| \geq e^{-4t^2}k^2 \text{ or } \omega \in \mathcal{B}_i. \end{cases}$$

Note that

$$\begin{aligned} E[Y_{i+1} \mid \mathcal{F}_i] &\leq e^{-4t^2}k^2\left(\frac{\sqrt{n}(y + g_y)}{n^2(q - g_q)}\right) - \frac{\log n}{\sqrt{n}}\left(\frac{(\gamma-\beta)^2}{4}\right)\left[8te^{-4t^2} + 20g_y\right] \\ &\leq \left(\frac{(\gamma-\beta)^2}{4}\right)\frac{\log n}{\sqrt{n}}\left(e^{-4t^2}\left[\frac{y}{q} + \frac{g_y}{q - g_q} + \frac{g_q y}{q(q - g_q)}\right] - \left[8te^{-4t^2} + 20g_y\right]\right) \\ &\leq \left(\frac{(\gamma-\beta)^2}{4}\right)\frac{\log n}{\sqrt{n}}\left(3g_y + 17tg_q - 20g_y\right) \\ &\leq 0. \end{aligned}$$

Therefore, $Y_\ell, Y_\ell + Y_{\ell+1}, \ldots, \sum_{i=\ell}^{j} Y_i, \ldots$ is a $(\frac{2\gamma^2 \log n}{\sqrt{n}}, kn^{-\rho})$–bounded supermartingale. It follows that we have

$$Pr\left[\sum_{i=\ell}^{j} Y_i > n^{1-\rho/3}\right] \le \exp\left\{-\frac{n^{2-2\rho/3}}{3\gamma^3 m \log^{3/2} n \cdot n^{-\rho}}\right\} = \exp\left\{-\frac{n^{1/2+\rho/3}}{3\mu\gamma^3 \log^2 n}\right\}. \qquad (7)$$

Now we turn to the event $\mathcal{P}_j \wedge \overline{\mathcal{B}_j}$, where we assume that it is step j where $|W_{A,B}(j)|$ is too small for the first time. There exists a maximum $\ell < j$ such that $|W_{A,B}(\ell)| \ge e^{-4t(\ell)^2} k^2$. Then

$$\begin{aligned}
|W_{A,B}(j)| &> e^{-4t(\ell)^2} k^2 - \sum_{i=\ell+1}^{j} X_i \\
&= e^{-4t(\ell)^2} k^2 - X_{\ell+1} - \sum_{i=\ell+2}^{j} Y_i - \frac{\log n}{\sqrt{n}} \left(\frac{(\gamma-\beta)^2}{4}\right) \sum_{i=\ell+2}^{j} \left[8te^{-4t^2} + 20g_y\right] \\
&> e^{-4t(j)^2} k^2 - \sum_{i=\ell+1}^{j} Y_i - k^2 \int_{t(\ell)}^{t(j)} 20 g_y(\tau) d\tau - \sqrt{n}.
\end{aligned}$$

Therefore, applying (7), we have

$$Pr\left(\mathcal{P}_m \wedge \overline{\mathcal{B}_m}\right) \le \binom{n}{k}^2 \cdot \left(n^{3/2}\sqrt{\log n}\right)^2 \cdot \exp\left\{-\frac{n^{1/2+\rho/3}}{3\mu\gamma^3 \log^2 n}\right\}.$$

□

Consider a fixed set K of $\gamma\sqrt{n \log n}$ vertices. We bound the probability that K is independent in G_m by first showing that if K is independent in G_i (and we are not in the 'bad' event $\mathcal{B}_i \vee \mathcal{D}_i \vee \mathcal{P}_i$) then the number of pairs in $\binom{K}{2} \cap O_i$ is at least a constant time $e^{-4t^2}|K|^2$. This implies that the edge e_{i+1} has a reasonably good chance of falling in K.

We restrict our attention to $\overline{\mathcal{B}_m} \wedge \overline{\mathcal{D}_m} \wedge \overline{\mathcal{P}_m}$. For each step i of the process such that $\binom{K}{2} \cap E_i = \emptyset$ let L_i be the set of vertices x such that $x \notin K$ and $|N_i(x) \cap K| > k/n^\rho$. Set

$$\mathcal{N}_i = \{N_i(x) \cap K : x \in L_i\}.$$

We first note that, since co-degrees are bounded when we are not in the event $\mathcal{B}_i$, we have

$$X, Y \in \mathcal{N}_i \quad \Rightarrow \quad |X \cap Y| \le \log^2 n.$$

It follows that the cardinality of the union of f sets in $\mathcal{N}_i$ is at least $fk/n^\rho - f^2 \log^2 n$, and therefore

$$|L_i| \le 2n^\rho.$$

Furthermore, as we restrict our attention to $\overline{\mathcal{D}_i}$, we have

$$X \in \mathcal{N}_i \quad \Rightarrow \quad |X| \le \beta\sqrt{n \log n}.$$

Now, we identify disjoint sets A, B such that the set of pairs $A \times B$ is essentially disjoint from $\binom{X}{2}$ for all $X \in \mathcal{N}_i$. Form $A \subseteq K$ such that $|A| = k$ by iteratively adding sets from $\mathcal{N}_i$ for as long as possible. Let $B \subseteq K \setminus A$ have the property that $|B| = k$ and $B \cap X = \emptyset$ for all $X \in \mathcal{N}_i$ that are used to form A. Note that we have

$$X \in \mathcal{N}_i \qquad \Rightarrow \qquad |X \cap A| \leq \log^2 n \cdot 2n^\rho \;\; \text{or} \;\; |X \cap B| = 0.$$

Note that the number of edges in $W_{A,B}(i)$ that are in C_i is at most

$$|L_i| \left(2 \log^2 n \cdot n^\rho\right) \cdot \beta\sqrt{n \log n} \leq 4\beta \log^{5/2} n \cdot n^{1/2+2\rho}.$$

Therefore, since $\omega_i \not\subseteq \mathcal{P}_i$,

$$\begin{aligned} \left| O_i \cap \binom{K}{2} \right| &\geq e^{-4t^2} \left(\frac{(\gamma-\beta)^2}{4} \right) n \log n - 2n^{1-\rho/3} - 4\beta \log^{5/2} n \cdot n^{1/2+2\rho} \\ &\geq e^{-4t^2} \cdot \frac{(\gamma-\beta)^2}{5} \cdot n \log n. \end{aligned}$$

Thus, since $\omega_i \not\subseteq \mathcal{B}_i$,

$$Pr\left(e_{i+1} \in \binom{K}{2}\right) \geq \frac{(\gamma-\beta)^2 \log n}{5n},$$

and the probability that K remains independent is at most

$$\left(1 - \frac{(\gamma-\beta)^2 \log n}{5n}\right)^{\mu\sqrt{\log n} \cdot n^{3/2}} \leq \exp\left\{ -\frac{(\gamma-\beta)^2}{5} \mu \log^{3/2} n \cdot \sqrt{n} \right\}.$$

On the other hand, the number of $\gamma\sqrt{n \log n}$-element sets of vertices is

$$\binom{n}{\gamma\sqrt{n \log n}} \leq \left(\frac{ne}{\gamma\sqrt{n \log n}} \right)^{\gamma\sqrt{\log n \cdot n}} \leq \exp\left\{ \frac{\gamma}{2} \log^{3/2} n \cdot \sqrt{n} \right\}.$$

Theorem 5 now follows from the union bound.

5 The K_4-free Process

We prove Theorem 3 by analyzing the K_4-free process on n vertices, showing that it produces a graph with independence number $O(\log^{4/5} n \cdot n^{2/5})$.

As in the analysis of the K_3-free process, we let E_i be the set of edges chosen through the first i steps in the process, $C_i \subseteq \binom{[n]}{2}$ be the set of forbidden pairs in G_i and $O_i \subseteq \binom{[n]}{2}$ be the set of available pairs in G_i.

We track the following random variables through the evolution of the K_4-free process. Let $Q(i)$ be $|O_i|$, the number of open pairs in $\binom{[n]}{2}$ after i steps of the process. For $A \in \binom{[n]}{2}$ and $f \in \{0, 1, 2, 3, 4\}$ let $X_{A,f}(i)$ be the collection of sets $B \in \binom{[n]}{2}$ such that

$$\left| E_i \cap \binom{A \cup B}{2} \right| = f \qquad \text{and} \qquad C_i \cap \binom{A \cup B}{2} \subseteq \binom{A}{2}.$$

Furthermore, for $f \in \{0,1,2,3\}$ and $A \in \binom{[n]}{3}$ let $Y_{A,f}(i)$ be the set of vertices v such that

$$|E_i \cap (A \times \{v\})| = f \qquad \text{and} \qquad C_i \cap \binom{A \cup B}{2} \subseteq \binom{A}{2}.$$

Of course, the random variables $X_{A,f}$ are the variables we are most interested in tracking; the variables $Y_{A,f}$ are introduced in order to maintain bounds on the one-step changes in the variables that comprise $X_{A,f}$. We stop tracking the variables once $\binom{A}{2} \subseteq E_i$, formally setting $X_{A,f}(i) = X_{A,f}(i-1)$ and $Y_{A,f}(i) = Y_{A,f}(i-1)$ in this situation. Our scaling is given by $t = t(i) = i/n^{8/5}$.

We introduce functions $q(t)$, $x_f(t)$ for $f = 0,1,2,3,4$, and $y_f(t)$ for $f = 0,1,2$. Our guess for the purpose of setting up the differential equations is the following

$$Q(i) \approx q(t)n^2 \qquad |X_{A,f}(i)| \approx x_f(t)n^{2-\frac{2f}{5}} \qquad |Y_{A,f}(i)| \approx y_f(t)n^{1-\frac{2f}{5}}.$$

This leads to the system of differential equations

$$\frac{dq}{dt} = -x_4 \qquad \frac{dx_0}{dt} = -\frac{5x_0x_4}{q} \qquad \frac{dx_f}{dt} = \frac{(6-f)x_{f-1}}{q} - \frac{(5-f)x_f x_4}{q} \quad \text{for } f = 1,2,3,4$$

with initial condition $q(0) = 1/2$, $x_0(0) = 1/2$ and $x_1(0) = \cdots = x_4(0) = 0$. This has solution

$$q(t) = \frac{1}{2}e^{-16t^5} \qquad x_f(t) = 2^{f-1}\binom{5}{f}t^f e^{-16(5-f)t^5} \quad \text{for } f = 0,1,2,3,4$$

With this solution in hand, we turn to $y_f(t)$. Here we have the equations

$$\frac{dy_0}{dt} = -\frac{3y_0x_4}{q} \qquad \frac{dy_f}{dt} = \frac{(4-f)y_{f-1}}{q} - \frac{(3-f)y_f x_4}{q} \quad \text{for } f = 1,2$$

with initial condition $y_0(0) = 1, y_1(0) = 0$ and $y_2(0) = 0$. This has solution

$$y_f(t) = 2^f\binom{3}{f}t^f e^{-16(3-f)t^5}.$$

Note that this suggests that the K_4-free process terminates with $\Theta\left(n^{8/5} \cdot \log^{1/5} n\right)$ edges.

In order to state our stability results we introduce error functions that slowly decay as the process evolves. The polynomial $p(t)$ has degree 5 and positive coefficients. We do not explicitly define this polynomial; it suffices that its coefficients are sufficiently large. Define

$$f_q = \begin{cases} e^{p(t)} & \text{if } t \le 1 \\ \frac{e^{p(t)}}{t^4} & \text{if } t > 1 \end{cases} \qquad f_f = e^{p(t)-16(4-f)t^5} \quad \text{for } f = 0,1,2,3,4$$

$$h_f = e^{p(t)-16(2-f)t^5} \quad \text{for } f = 0,1,2.$$

Define $\mathcal{B}_i$ to be the event that there exists $j \le i$ such that

$$\left|Q(j) - q(t(j))n^2\right| \ge f_q(t(j))n^{29/15}$$

or there is a set $A \in \binom{[n]}{2}$ and $f \in \{0, 1, \ldots, 4\}$ such that $\binom{A}{2} \not\subseteq E_j$ and

$$\left| |X_{A,f}(j)| - x_f(t(j))n^{2-\frac{2f}{5}} \right| \geq f_f(t(j))n^{2-\frac{2f}{5}-\frac{1}{15}}$$

or there is a set $A \in \binom{[n]}{3}$ and $f \in \{0, 1, 2\}$ such that $\binom{A}{2} \not\subseteq E_j$ and

$$|Y_{A,f}(j)| > y_f(t(j))n^{1-\frac{2f}{5}} + h_f(t(j))n^{1-\frac{2f}{5}-\frac{1}{15}}$$

or there is a set $A \in \binom{[n]}{3}$ such that $\binom{A}{2} \not\subseteq E_j$ and

$$|Y_{A,3}(j)| > 15.$$

We introduce absolute constants μ, ρ and γ. As in our analysis of the K_4-free process, μ and ρ are small relative to $p(t)$ and γ is large with respect to μ. Define $m = \mu n^{8/5} \log^{1/5} n$.

Theorem 13. *If n is sufficiently large then*

$$Pr\left(\mathcal{B}_{\mu n^{8/5} \log^{1/5} n}\right) \leq n^{-1/6}.$$

Theorem 14. *If n is sufficiently large then*

$$Pr\left(\alpha\left(G_{\mu n^{8/5} \log^{1/5} n}\right) > \gamma n^{2/5} \log^{4/5} n \mid \overline{\mathcal{B}_m}\right) < e^{-n^{1/15}}.$$

The methods introduced in Sections 3 and 4 can be used to prove Theorems 13 and 14. This is more or less straightforward and is mostly left to the reader; we conclude this section with the details that do not follow immediately as above.

Proof of Theorem 13. There is one significant difference between the triangle-free process and the K_4-free process that must be dealt with here. In the case of the triangle-free process, there is a one-to-one correspondence between edges closed when e_i is added and vertices that are partial with respect to e_i in G_{i-1}. The analogous correspondence does not hold for the K_4-free process: Since a pair $\{u, v\}$ that intersects e_i could be a subset of $B \cup e_i$ for many sets $B \in X_{e_i,4}(i-1)$, there is not a one-to-one correspondence between pairs $\{u, v\}$ closed by the addition of e_i to the graph and $X_{e_i,4}(i-1)$. In order to overcome this problem we note that, based on simple density considerations, the difference between these two quantities is bounded by $n^{4/15}$ in the event $\overline{\mathcal{B}_m}$.

Let ϵ be a sufficiently small constant. (This constant is chosen so that $|O_i| \geq n^{2-\epsilon}$ for all $i \leq m$ in the event $\overline{\mathcal{B}_m}$). Set $k = n^{1/5+8\epsilon}$ and let $\mathcal{M}_i$ be the event that there exists $\{u, v\} \in \binom{[n]}{2}$, distinct vertices $w_1, \ldots, w_k \in [n] \setminus \{u, v\}$ and distinct $z_1, z_2, \ldots, z_{2k} \in [n] \setminus \{u, v, w_1, w_2, \ldots, w_k\}$ such that

$$\{u, w_j\}, \{u, z_{2j-1}\}, \{u, z_{2j}\}, \{v, z_{2j-1}\}, \{v, z_{2j}\}, \{w_j, z_{2j-1}\}, \{w_j, z_{2j}\} \in E_i \text{ for } j = 1, \ldots, k.$$

Claim 15. *If n is sufficiently large*

$$Pr\left(\mathcal{M}_m \wedge \overline{\mathcal{B}_m}\right) \leq e^{-n^{1/5}}.$$

Proof.

$$\begin{aligned}
Pr\left(\mathcal{M}_m \wedge \overline{\mathcal{B}_m}\right) &\leq n^2 \cdot \binom{n}{k} \cdot n^{2k} \cdot m^{7k} \left(\frac{1}{n^{2-\epsilon}}\right)^{7k} \\
&\leq n^2 \left(\frac{en^3 \cdot (\mu n^{8/5} \log^{1/5} n)^7}{k \cdot n^{14-7\epsilon}}\right)^k \\
&= n^2 \left(\frac{e\mu^7 n^{1/5+7\epsilon} \log^{7/5} n}{k}\right)^k .
\end{aligned}$$

□

Now let $\ell = n^{6\epsilon}$ and let $\mathcal{N}_i$ be the event that there exist vertices u, v, z and disjoint sets $A, B \in \binom{[n]}{\ell}$ such that $A \subseteq N_i(u) \cap N_i(z)$, $B \subseteq N_i(u) \cap N_i(v)$ and G_i has a matching of ℓ edges in $A \times B$.

Claim 16. *If n is sufficiently large then*

$$Pr(\mathcal{N}_m \wedge \overline{\mathcal{B}_m}) \leq e^{-n^{5\epsilon}}.$$

Proof.

$$\begin{aligned}
Pr\left(\mathcal{N}_m\right) &\leq n^3 \cdot \binom{n}{\ell} \cdot n^{\ell} \cdot m^{5\ell} \left(\frac{1}{n^{2-\epsilon}}\right)^{5\ell} \\
&\leq n^3 \left(\frac{en^2 \cdot (\mu n^{8/5} \log^{4/5} n)^5}{\ell \cdot n^{10-5\epsilon}}\right)^{\ell} .
\end{aligned}$$

□

Now suppose $\omega_{j-1} \not\subseteq \mathcal{B}_{j-1} \vee \mathcal{M}_{j-1} \vee \mathcal{N}_{j-1}$. Let W be the set of vertices w such that $\{v, w\}$ is closed by the addition of the edge $e_j = \{u, v\}$ and there exist distinct vertices z_w, z'_w such that $\{w, z_w\}, \{w, z'_w\} \in X_{e_j,4}(j-1)$. Note that $\omega_{j-1} \not\subseteq \mathcal{B}_{j-1}$ implies that the number of pairs $B \in X_{e_j,4}(j-1)$ that correspond to a particular vertex $w \in W$ is at most 16. Thus the difference between $|X_{e_j,4}(j-1)|$ and the number of pairs closed by the addition of edges e_j is at most 32 any bound we establish on W (taking into account pairs closed by the addition of $\{u, v\}$ that contain u). It remains to argue that W is small. First note that $\omega_{j-1} \not\subseteq \mathcal{N}_{j-1}$ implies that each vertex z is in the set $\{z_w, z'_w\}$ for at most $16n^{6\epsilon}$ vertices $w \in W$. Let $W' \subseteq W$ be a maximum set such that $a, b \in W'$ implies $\{z_a, z'_a\} \cap \{z_b, z'_b\} = \emptyset$. By the previous observation (using $\omega_{j-1} \not\subseteq \mathcal{N}_{j-1}$) we have $|W'| \geq |W|/(16n^{6\epsilon})$. Furthermore, $\omega_{j-1} \not\subseteq \mathcal{M}_{j-1}$ implies that $|W'| < n^{1/5+8\epsilon}$. Thus, the number of pairs closed by the addition of e_j is in the interval

$$\left[\left|X_{e_j,4}(j-1)\right| - 32 \cdot 16n^{1/5+14\epsilon}, \left|X_{e_j,4}(j-1)\right|\right],$$

which is sufficient for the proof.

□

Proof of Theorem 14. As in the proof of Theorem 5, we fix a set K of $\gamma n^{2/5}\log^{4/5} n$ vertices and show that the probability that K remains independent is small even when compared with the number of such sets. We condition on $\overline{\mathcal{B}_m}$ and a bound of $n^{1/5+3\epsilon}$ on all co-degrees. (This bound on the co-degrees follows from a very simple first moment calculation. We could establish a tighter bound using martingale inequalities, but that is not necessary for this argument.)

There are two significant differences between the triangle-free process and the K_4-free process here: the fact that in the latter the addition of an edge e_i that is disjoint from K could close many pairs within K and the fact that the neighborhood of a single vertex could include K as a subset.

We track the number of open pairs within two kinds of subgraphs. Set

$$k = \frac{\gamma}{3} n^{2/5}\log^{4/5} n.$$

Let $A, B \in \binom{[n]}{k}$. We say that a pair $\{u,v\} \in A \times B$ is **closed with respect to** $A \times B$ at step j if there exists a step $i \le j$ such that $\{u,v\}$ is among the edges closed by $e_i = \{x,y\}$ and either

(i) $e_i \cap (A \cup B) = \{y\} = \{u\}$ and there exists $z \notin A \cup B$ such that $\{v,z\} \in X_{\{x,y\},4}$ and $|N_{i-1}(z) \cap N_{i-1}(x) \cap (A \cup B)| < n^{1/5-\rho-3\epsilon}$ or

(ii) $e_i \cap (A \cup B) = \emptyset$ and ($|N_{i-1}(e_i) \cap A| \le n^{1/5-\rho-3\epsilon}$ or $|N_{i-1}(e_i) \cap B| \le n^{1/5-\rho-3\epsilon}$).

If the pair $\{u,v\} \in A \times B$ is neither closed with respect to $\{u,v\}$ nor in the edge set E_j then it is **open with respect to** $A \times B$. Note that, since we assume co-degrees are bounded by $n^{1/5+3\epsilon}$, the change in the number of pairs closed with respect to $A \times B$ that results from the addition of an edge e_i is at most $n^{2/5-\rho}$. It follows from the techniques in Section 4 that with high probability we have the following: For all steps $j \le m$ and all pairs $A, B \in \binom{[n]}{k}$ such that $(A \times B) \cap E_j = \emptyset$ the number of edges in $A \times B$ that are open with respect to $A \times B$ is at least $\frac{k^2}{2} e^{-16t(j)^5}$.

We also track the number of open pairs within sets D consisting of k vertices. We say that a pair $\{u,v\} \in \binom{D}{2}$ is **closed with respect to** D at step j if there exists a step $i \le j$ such that $\{u,v\}$ is among the edges closed by $e_i = \{x,y\}$ and either

(i) $e_i \cap D = \{y\} = \{u\}$ and there exists $z \notin D$ such that $|N_{i-1}(z) \cap N_{i-1}(x) \cap D| < n^{1/5-\rho-3\epsilon}$ and $\{v,z\} \in X_{\{x,y\},4}$.

(ii) $e_i \cap D = \emptyset$ and $|N_{i-1}(x) \cap N_{i-1}(y) \cap D| < n^{1/5-\rho/2}$.

If the pair $\{u,v\} \in \binom{D}{2}$ is neither closed with respect to $\{u,v\}$ nor in the edge set E_j then it is **open with respect to** D . Again following the techniques in Section 4, we see that with high probability we have the following: For all steps $j \le m$ and all sets $D \in \binom{[n]}{k}$ such that $\binom{D}{2} \cap E_j = \emptyset$ the number of edges in $\binom{D}{2}$ that are open with respect to D is at least $\frac{k^2}{4} e^{-16t(j)^5}$.

It remains to show that every set K of $\gamma n^{2/5}\log^{4/5} n$ vertices contains:

(a) Disjoint sets A, B of k vertices such the difference between the number of pairs in $A \times B$ that are open and the number that are open with respect to $A \times B$ is less than, say, $n^{23/30}$, or

(b) A set D of k vertices such the difference between the number of pairs within D that are open and the number that are open with respect to D is less than $n^{23/30}$.

A main tool here is the following observation which follows from a simple first moment calculation. Let $\mathcal{M}_i$ be the event that there exist integers r, s such that $s \geq n^{2\epsilon}$, $r \cdot s \geq n^{2/5+\epsilon}$ and disjoint sets $X \in \binom{[n]}{k}$ and $Y \in \binom{[n]}{r}$ such that

$$|N_m(y) \cap X| \geq s \quad \text{for all} \quad y \in Y.$$

Claim 17. *If n is sufficiently large then $Pr\left(\mathcal{M}_m\right) \leq e^{-n^{2/5}}$.*

Now, let L_j be the set of vertices $x \notin K$ such that

$$|N_j(x) \cap K| \geq n^{1/5-\rho-3\epsilon}.$$

Let $L_j = \{x_1, x_2, \dots\}$ be arranged in decreasing order of $|N_j(x_\ell) \cap K|$. A simple case analysis in conjuction with Claim 17 now establishes the desired property.

Case 1. $|N_j(x_1) \cap K| \geq k$.

Consider $D \subseteq N_j(x_1) \cap K$ such that $|D| = k$. Note that, appealing to the bound on common neighbors of triples of vertices given by conditioning on $\overline{\mathcal{B}_m}$, all pairs within D that are closed but not closed with respect to D are contained in $N_j(x_1) \cap N_j(y) \cap D$ where $y \in L_j$ and $\{x_1, y\} \in E_j$. The number of such vertices y is at most $n^{1/5+\rho+4\epsilon}$ by Claim 17. Each such neighborhood includes less than $n^{2/5+6\epsilon}$ edges because of the bound on the co-degrees. Therefore, the number of spoiled pairs within D is at most $n^{3/5+\rho+10\epsilon}$.

Case 2. $|N_j(x_1) \cap K| < k$.

Choose

$$A \subseteq \bigcup_{\ell=1}^{\ell'} N_j(x_\ell) \cap K \qquad B \subseteq K \setminus \left(\bigcup_{\ell=1}^{\ell'} N_j(x_\ell) \right)$$

such that $|A| = |B| = k$ where ℓ' is the smallest index such that the cardinality of this union is at least k.

First suppose $\ell' < n^{2/15}$. Note that no pairs in $A \times B$ are spoiled in this case: If $x, y \notin K$ then either $N_j(x) \cap N_j(y) \subseteq A$ or $|N_j(x) \cap N_j(y) \cap A| \leq 16n^{2/15}$ (using the bound on common neighbors of triples of vertices).

Finally, suppose $\ell' \geq n^{2/15}$. Note that, by Claim 17, we have $|N_j(x_{\ell'}) \cap K| \leq n^{4/15+\epsilon}$ and $|L_j| \leq n^{1/5+\rho+4\epsilon}$. Thus, the number of spoiled pairs is at most $n^{8/15+2\epsilon} n^{1/5+\rho+4\epsilon} = n^{11/15+\rho+6\epsilon}$.

□

6 Martingale Inequalities

Lemmas 6 and 7 follow from the original martingale inequality of Hoeffding.

Theorem 18 (Hoeffding [15]). *Let $0 \equiv X_0, X_1, \ldots$ be a sequence of random variables such that*

$$X_{k-1} - \mu_k \leq X_k \leq X_{k-1} + 1 - \mu_k$$

for some constant $0 < \mu_k < 1$ for k =1, ..., m. Set $\mu = \frac{1}{m}\sum_{k=1}^{m} \mu_k$ and $\overline{\mu} = 1 - \mu$. If $X_0, X_1, \ldots$ is a supermartingale and $0 < t < \overline{\mu}$ then

$$Pr\left(X_m \geq mt\right) \leq \left[\left[\frac{\mu}{\mu + t}\right]^{\mu+t}\left[\frac{\overline{\mu}}{\overline{\mu} - t}\right]^{\overline{\mu}-t}\right]^m. \tag{8}$$

Hoeffding's result was for martingales, but the extension to supermartingales is straightforward. For a survey of applications of this and similar results, see McDiarmid [20].

In order to apply Theorem 18 to the martingales considered in this paper, we introduce the following function. For $0 < v < 1/2$ set $\overline{v} = 1 - v$ and define

$$g(x) = g(x, v) = (v + xv)\log\left(\frac{v}{v + xv}\right) + (\overline{v} - xv)\log\left(\frac{\overline{v}}{\overline{v} - xv}\right) \qquad \text{for } -1 < x < 1.$$

Note that, under the conditions of Theorem 18, we have

$$Pr(X_m \geq mx\mu) \leq e^{g(x,\mu)m} \qquad \text{and} \qquad Pr(X_m \geq mx\overline{\mu}) \leq e^{g(-x,\overline{\mu})m}.$$

Note further

$$g''(x) = -\frac{v}{(1 + x)(\overline{v} - xv)}.$$

Proof of Lemma 6. Let $0 \equiv A_0, A_1, \ldots$ be a (η, N)-bounded submartingale with $N \geq 2\eta$. Let $a \leq m\eta$. Define $X_i = -A_i/(\eta + N)$. Note that Theorem 18 applies to $X_0, X_1, \ldots$ with $\mu = N/(\eta + N)$. Thus

$$Pr(A_m \leq -a) = Pr\left(X_m \geq \frac{a}{\eta + N}\right) \leq \exp\left\{g\left(-\frac{a}{m\eta}, \overline{\mu}\right)m\right\}.$$

It remains to bound $g(x)$. Note that if $-1 < x \leq 0$ then $g''(x) \leq -v$. As $g(0) = g'(0) = 0$, it follows that $g(x) \leq -vx^2/2$ for $-1 \leq x \leq 0$. Therefore,

$$Pr(A_m \leq -a) \leq \exp\left\{-\frac{\eta}{N + \eta}\frac{a^2 m}{2m^2\eta^2}\right\} \leq \exp\left\{-\frac{a^2}{2m\eta(N + \eta)}\right\}.$$

□

Proof of Lemma 7. Let $0 \equiv A_0, A_1, \ldots$ be a (η, N)-bounded supermartingale with $N \geq 10\eta$. Let $a < \eta m$. Define $X_i = A_i/(\eta+N)$. Theorem 18 applies to $X_0, X_1, \ldots$, with $\mu = \eta/(\eta+N)$. We have

$$Pr(A_m \geq a) = Pr\left(X_m \geq \frac{a}{\eta + N}\right) \leq \exp\left\{g\left(\frac{a}{m\eta}, \mu\right)m\right\}.$$

It remains to bound $g(x)$. Note that for $x \geq 0$ we have $g''(x) \leq -v/(1 + x)$. Since $g(0) = g'(0) = 0$, this implies

$$g(x) \leq -v\left[(1 + x)\log(1 + x) - x\right] \leq v\left[-\frac{x^2}{2} + \frac{x^3}{6} - \frac{x^4}{12} + \frac{x^5}{20}\right] \leq -\frac{11}{30}vx^2.$$

Thus

$$Pr(A_m \geq a) \leq \exp\left\{-\frac{11}{30}\frac{a^2}{m\eta(N+\eta)}\right\} \leq \exp\left\{-\frac{a^2}{3m\eta N}\right\}.$$

□

Acknowledgment. The author thanks Alan Frieze, Jeff Kahn, Peter Keevash and Joel Spencer for many useful comments on earlier drafts of this paper.

References

[1] M. Ajtai, J. Komlós and E. Szemerédi, A note on Ramsey numbers. *Journal of Combinatorial Theory A* **29** (1980) 354-360.

[2] N. Alon, Explicit ramsey graphs and orthonormal labelings. *Electronic Journal of Combinatorics* **1** (1994) R12, 8pp.

[3] N. Alon, S. Ben-Shimon and M. Krivelevich, A note on regular Ramsey Graphs. manuscript, 2008.

[4] A. Beveridge, T. Bohman, A. Frieze and O. Pikhurko, Product rule wins a competitive game. *Proceedings of the American Mathematical Society* **135** (2007) 3061-3071.

[5] T. Bohman and A. Frieze, Karp-Sipser on random graphs with a fixed degree sequence, submitted.

[6] T. Bohman and D. Kravitz, Creating a giant component. *Combinatorics, Probability and Computing* **15** (2006) 489-511.

[7] B. Bollobás, personal communication.

[8] B. Bollobás and O. Riordan, Constrained graph processes, *Electronic Journal of Combinatorics* **7(1)** (2000) R18.

[9] F. Chung and R. Graham, *Erdős on Graphs: His Legacy of Unsolved Problems*. A.K. Peters 1999.

[10] P. Erdős, Graph theory and probability II. *Canadian Journal of Mathematics* **13** (1961) 346-352.

[11] P. Erdős and L. Lovász, Problems and results on 3-chromatic hypergraphs and some related questions, in *Infinite and Finite Sets*. North Holland 1975.

[12] P. Erdős, S. Suen and P. Winkler, On the size of a random maximal graph. *Random Structures and Algorithms* **6** (1995) 309-318.

[13] R. Graham, B. Rothschild and J. Spencer, *Ramsey Theory*. Wiley, 1990.

[14] J. Graver and J. Yackel, Some graph theoretic results associated with Ramsey's theorem. *Journal of Combinatorial Theory* **4** (1968) 125-175.

[15] W. Hoeffding, Probability inequalities for sums of bounded variables. *Journal of the American Statistical Association* **58** (1963) 13-30.

[16] P. Keevash, personal communication.

[17] J. H. Kim, The ramsey number $R(3,t)$ has order of magnitude $t^2/\log t$. *Random Structures and Algorithms* **7** (1995) 173-207.

[18] M. Krivelevich, Bounding Ramsey numbers through large deviation inequalities, *Random Structures and Algorithms* **7** (1995) 145-155.

[19] D. Osthus and A. Taraz, Random maximal H-free greaphs. *Random Structures and Algorithms* **18** (2001) 61-82.

[20] C. McDiarmid, On the method of bounded differences, in *Surveys in Combinatorics, London Mathematics Society Lecture Note Series 141.* Cambridge University Press 1989.

[21] A. Ruciński and N. Wormald, Random graph processes with degree restrictions. *Combinatorics, Probability and Computing* **1** (1992) 169-180.

[22] J. Shearer, A note on the independence number of triangle-free graphs II. *Journal of Combinatorial Theory B* **53** (1991) 300-307.

[23] J. Spencer, Asymptotic lower bounds for Ramsey functions. *Discrete Mathematics* **20** (1977) 69-76.

[24] J. Spencer, Maximal triangle-free graphs and Ramsey $R(3,t)$, unpublished manuscript, 1995.

[25] J. Spencer and N. Wormald, Birth control for giants. *Combinatorica* **27** (2007) 587-628.

[26] N. Wormald, The differential equations method for random graph processes and greedy algorithms. pp. 73-155 in *Lectures on Approximation and Randomized Algorithms*, Karonski and Prömel eds. PWN, Warsaw 1999.