Internet Mathematics Vol. 1, No. 4: 409-461

Coupling Online and Offline Analyses for Random Power Law Graphs

Fan Chung and Linyuan Lu

Abstract. We develop a coupling technique for analyzing online models by using offline models. This method is especially effective for a growth-deletion model that generalizes and includes the preferential attachment model for generating large complex networks which simulate numerous realistic networks. By coupling the online model with the offline model for random power law graphs, we derive strong bounds for a number of graph properties including diameter, average distances, connected components, and spectral bounds. For example, we prove that a power law graph generated by the growth-deletion model almost surely has diameter $O(\log n)$ and average distance $O(\log \log n)$.

1. Introduction

In the past few years, it has been observed that a variety of information networks, including Internet graphs, social networks, and biological networks among others [Aiello et al. 00, Aiello et al. 02, Barabási and Albert 99, Barabási et al. 00, Jeong et al. 00, Kleinberg et al. 99, Lu 01], have the so-called *power law* degree distribution. A graph is called a power law graph if the fraction of vertices with degree k is proportional to $\frac{1}{k^\beta}$ for some constant $\beta > 0$. There are basically two different models for random power law graphs.

The first model is an "online" model that mimics the growth of a network. Starting from a vertex (or some small initial graph), a new node and/or new edge is added at each unit of time following the so-called *preferential attachment*

1542-7951/04 $0.50 per page

scheme [Aiello et al. 02, Barabási and Albert 99, Kleinberg et al. 99]. The endpoint of a new edge is chosen with the probability proportional to their (current) degrees. By using a combination of adding new nodes and new edges with given respective probabilities, one can generate large power law graphs with exponents β greater than 2 (see [Aiello et al. 02, Bollabás and Riordan 03] for rigorous proofs). Since realistic networks encounter both growth and deletion of vertices and edges, we consider a growth-deletion online model that generalizes and includes the preferential attachment model. Detailed definitions will be given in Section 3.

The second model is an "offline" model of random graphs with given expected degrees. For a given sequence $\mathbf{w}$ of weights w_i, a random graph in $G(\mathbf{w})$ is formed by choosing the edge between u and v with probability proportional to the product of w_u and w_v. The Erdős-Rényi model $G(n,p)$ can be viewed as a special case of $G(\mathbf{w})$ with all w_i equal. Because of the independence in the choices of edges, the model $G(\mathbf{w})$ is amenable to a rigorous analysis of various graph properties and structures. In a series of papers [Chung and Lu 02a, Chung and Lu 02b, Chung et al. 03, Lu 01], various graph invariants have been examined and sharp bounds have been derived for diameter, average distance, connected components, and spectra for random power law graphs and, in general, random graphs with given expected degrees.

The online model is obviously much harder to analyze than the offline model. There has been some recent work on the online model beyond showing that the generated graph has a power law degree distribution. Bollobás and Riordan [Bollabás and Riordan 03] have derived a number of graph properties for the online model by "coupling" with $G(n,p)$, namely, identifying (almost regular) subgraphs whose behavior can be captured in a similar way as graphs from $G(n,p)$ for some appropriate p.

In this paper, our goal is to couple the online model with the offline model of random graphs with a similar power law degree distribution so that we can apply the techniques from the offline model to the online model. The basic idea is similar to the martingale method but with substantial differences. Although a martingale involves a sequence of functions with consecutive functions having small bounded differences, each function is defined on a fixed probability space Ω. For the online model, the probability space for the random graph generated at each time instance is *different* in general. We have a sequence of probability spaces where two consecutive ones have "small" differences. To analyze this, we need to examine the relationship of two distinct random graph models, each of which can be viewed as a probability space. In order to do so, we shall describe two basic methods that are not only useful for our proofs here but also interesting in their own right.

- **Comparing two random graph models.** We define the *dominance* of one random graph model over another in Section 4. Several key lemmas for controlling the differences are also given there.
- **A general Azuma inequality.** A concentration inequality is derived for martingales that are almost Lipschitz. A complete proof is given in Section 5.

The main goal of this paper is to show the following results for the random graph G generated by the online model $G(p_1, p_2, p_3, p_4, m)$ with $p_1 > p_3, p_2 > p_4$, as defined in Section 5:

1. Almost surely the degree sequence of the random graph generated by growth-deletion model $G(p_1, p_2, p_3, p_4, m)$ follows the power law distribution with exponent $\beta = 2 + (p_1 + p_3)/(p_1 + 2p_2 - p_3 - 2p_4)$.
2. Suppose $m > \log^{1+\epsilon} n$. For $p_2 < p_3 + p_4$, we have $2 < \beta < 3$. Almost surely a random graph in $G(p_1, p_2, p_3, p_4, m)$ has diameter $\Theta(\log n)$ and average distance $O(\frac{\log\log n}{\log(1/(\beta-2)})$. We note that the average distance is defined to be the average over all distances among pairs of vertices in the same connected component.
3. Suppose $m > \log^{1+\epsilon} n$. For $p_2 \geq p_3 + p_4$, we have $\beta > 3$. Almost surely a random graph in $G(p_1, p_2, p_3, p_4, m)$ has diameter $\Theta(\log n)$ and average distance $O(\frac{\log n}{\log d})$ where d is the average degree.
4. Suppose $m > \log^{1+\epsilon} n$. Almost surely a random graph in $G(p_1, p_2, p_3, p_4, m)$ has Cheeger constant at least $1/2 + o(1)$.
5. Suppose $m > \log^{1+\epsilon} n$. Almost surely a random graph in $G(p_1, p_2, p_3, p_4, m)$ has spectral gap λ at least $1/8 + o(1)$.

We note that the *Cheeger constant* h_G of a graph G, which is sometimes called the *conductance*, is defined by

$$h_G = \frac{|E(A, \bar{A})|}{\min\{\mathrm{vol}(A), \mathrm{vol}(\bar{A})\}},$$

where $\mathrm{vol}(A) = \sum_{x \in A} deg(x)$. The Cheeger constant is closely related to the spectral gap λ of the Laplacian of a graph by the Cheeger inequality

$$2h_G \geq \lambda \geq h_G^2/2.$$

Thus, both h_G and λ are key invariants for controlling the rate of convergence of random walks on G.

2. Strong Properties of Offline Random Power Law Graphs

For random graphs with given expected degree sequences satisfying a power law distribution with exponent β, we may assume that the expected degrees are $w_i = ci^{-\frac{1}{\beta-1}}$ for i satisfying $i_0 \le i < n + i_0$. Here c depends on the average degree, and i_0 depends on the maximum degree m, namely, $c = \frac{\beta-2}{\beta-1} dn^{\frac{1}{\beta-1}}$ and

$$i_0 = n \left(\frac{d(\beta-2)}{m(\beta-1)} \right)^{\beta-1}.$$

2.1. Average Distance and Diameter

Fact 2.1. ([Chung and Lu 02b]) *For a power law random graph with exponent $\beta > 3$ and average degree d strictly greater than 1, almost surely the average distance is $(1+o(1))\frac{\log n}{\log \tilde{d}}$ and the diameter is $\Theta(\log n)$.*

Fact 2.2. ([Chung and Lu 02b]) *Suppose a power law random graph with exponent β has average degree d strictly greater than 1 and maximum degree m satisfying $\log m \gg \log n / \log \log n$. If $2 < \beta < 3$, almost surely the diameter is $\Theta(\log n)$ and the average distance is at most $(2+o(1))\frac{\log \log n}{\log(1/(\beta-2))}$.*

For the case of $\beta = 3$, the power law random graph has diameter almost surely $\Theta(\log n)$ and has average distance $\Theta(\log n / \log \log n)$.

2.2. Connected Components

Fact 2.3. ([Chung and Lu 02a]) *Suppose that G is a random graph in $G(\mathbf{w})$ with given expected degree sequence $\mathbf{w}$. If the expected average degree d is strictly greater than 1, then the following hold:*

1. *Almost surely G has a unique giant component. Furthermore, the volume of the giant component is at least $(1 - \frac{2}{\sqrt{de}} + o(1))\mathrm{Vol}(G)$ if $d \ge \frac{4}{e} = 1.4715\ldots$ and is at least $(1 - \frac{1+\log d}{d} + o(1))\mathrm{Vol}(G)$ if $d < 2$.*

2. *The second largest component almost surely has size $O(\frac{\log n}{\log d})$.*

2.3. Spectra of the Adjacency Matrix and the Laplacian

The spectra of the adjacency matrix and the Laplacian of a non-regular graph can have quite different distribution. The definition for the Laplacian can be found in [Chung 97].

Fact 2.4. ([Chung et al. 03])

1. *The largest eigenvalue of the adjacency matrix of a random graph with a given expected degree sequence is determined by m, the maximum degree, and $\tilde{d}$, the weighted average of the squares of the expected degrees. We show that the largest eigenvalue of the adjacency matrix is almost surely $(1+o(1))\max\{\tilde{d},\sqrt{m}\}$ provided that some minor conditions are satisfied. In addition, suppose that the kth largest expected degree m_k is significantly larger than $\tilde{d}^2$. Then the kth largest eigenvalue of the adjacency matrix is almost surely $(1+o(1))\sqrt{m_k}$.*

2. *For a random power law graph with exponent $\beta > 2.5$, the largest eigenvalue of a random power law graph is almost surely $(1+o(1))\sqrt{m}$, where m is the maximum degree. Moreover, the k largest eigenvalues of a random power law graph with exponent β have power law distribution with exponent $2\beta-1$ if the maximum degree is sufficiently large and k is bounded above by a function depending on β, m, and d, the average degree. When $2 < \beta < 2.5$, the largest eigenvalue is heavily concentrated at $cm^{3-\beta}$ for some constant c depending on β and the average degree.*

3. *We will show that the eigenvalues of the Laplacian satisfy the semicircle law under the condition that the minimum expected degree is relatively large ($\gg$ the square root of the expected average degree). This condition contains the basic case when all degrees are equal (the Erdös-Rényi model). If we weaken the condition on the minimum expected degree, we can still have the following strong bound for the eigenvalues of the Laplacian which implies strong expansion rates for rapidly mixing:*

$$\max_{i\neq 0}|1-\lambda_i| \leq (1+o(1))\frac{4}{\sqrt{\bar{w}}} + \frac{g(n)\log^2 n}{w_{\min}},$$

where $\bar{w}$ is the expected average degree, $w_{\min}$ is the minimum expected degree, and $g(n)$ is any slow growing function of n.

3. A Growth-Deletion Model for Generating Random Power Law Graphs

One explanation for the ubiquitous occurrence of power laws is the simple growth rules that can result in a power law distribution (see [Aiello et al. 02, Barabási and Albert 99]). Nevertheless, realistic networks usually encounter both the growth and deletion of vertices and edges. Here we consider a general online model that combine deletion steps with the *preferential attachment model.*

Vertex-growth step. Add a new vertex v and form a new edge from v to an existing vertex u chosen with probability proportional to d_u.

Edge-growth step. Add a new edge with endpoints to be chosen among existing vertices with probability proportional to the degrees. If existing in the current graph, the generated edge is discarded. The edge-growth step is repeated until a new edge is successfully added.

Vertex-deletion step. Delete a vertex randomly.

Edge-deletion step. Delete an edge randomly.

For nonnegative values p_1, p_2, p_3, p_4 summing to 1, we consider the following growth-deletion model $G(p_1, p_2, p_3, p_4)$:

At each step,
with probability p_1, take a vertex-growth step;
with probability p_2, take an edge-growth step;
with probability p_3, take a vertex-deletion step;
with probability $p_4 = 1 - p_1 - p_2 - p_3$, take an edge-deletion step.

Here we assume that $p_3 < p_1$ and $p_4 < p_2$ so that the number of vertices and edge grows as t goes to infinity. If $p_3 = p_4 = 0$, the model is just the usual preferential attachment model that generates power law graphs with exponent $\beta = 2 + \frac{p_1}{p_1+2p_2}$. An extensive survey on the preferential attachment model is given in [Mitzenmacher 05] and rigorous proofs can be found in [Aiello et al. 02, Cooper and Frieze 03].

This growth-deletion model generates only simple graphs because the multiple edges are disallowed at the edge-growth step. The drawback is that the edge-growth step could run in a loop. It only happens if the current graph is a completed graph. If this happens, we simply restart the whole procedure from the same initial graph. With high probability, the model generates sparse graphs so that we could omit the analysis of this extreme case.

Previously, Bollobás considered edge deletion after the power law graph is generated [Bollabás and Riordan 03]. Very recently, Cooper, Frieze, and Vera [Cooper et al. 04] independently consider the growth-deletion model with vertex deletion only. We will show (see Section 6) the following.

Suppose that $p_3 < p_1$ and $p_4 < p_2$. Then almost surely the degree sequence of the growth-deletion model $G(p_1, p_2, p_3, p_4)$ follows the power law distribution with the exponent

$$\beta = 2 + \frac{p_1 + p_3}{p_1 + 2p_2 - p_3 - 2p_4}.$$

We note that a random graph in $G(p_1, p_2, p_3, p_4)$ almost surely has expected average degree $(p_1 + p_2 - p_4)/(p_1 + p_3)$. For of p_is in certain ranges, this value can be below 1 and the random graph is not connected. To simulate graphs with specified degrees, we consider the following modified model $G(p_1, p_2, p_3, p_4, m)$, for some integer m that generates random graphs with the expected degree $m(p_1 + p_2 - p_4)/(p_1 + p_3)$:

At each step,
with probability p_1, add a new vertex v and form m new edges from v to existing vertices u chosen with probability proportional to d_u;
with probability p_2, take m edge-growth steps;
with probability p_3, take a vertex-deletion step;
with probability $p_4 = 1 - p_1 - p_2 - p_3$, take m edge-deletion steps.

Suppose that $p_3 < p_1$ and $p_4 < p_2$. Then almost surely the degree sequence of the growth-deletion model $G(p_1, p_2, p_3, p_4, m)$ follows the power law distribution with the exponent β the same as the exponent for the model $G(p_1, p_2, p_3, p_4)$:

$$\beta = 2 + \frac{p_1 + p_3}{p_1 + 2p_2 - p_3 - 2p_4}.$$

Many results for $G(p_1, p_2, p_3, p_4, m)$ can be derived in the same fashion as for $G(p_1, p_2, p_3, p_4)$. Indeed, $G(p_1, p_2, p_3, p_4) = G(p_1, p_2, p_3, p_4, 1)$ is usually the hardest case because of the sparseness of the graphs.

4. Comparing Random Graphs

In the early work of Erdős and Rényi on random graphs, they first used the model $F(n, m)$ that each graph on n vertices and m edges is chosen randomly with equal probability, where n and m are given fixed numbers. This model is apparently different from the later model $G(n, p)$, for which a random graph is formed by choosing independently each of the $\binom{n}{2}$ pairs of vertices to be an edge with probability p. Because of the simplicity and ease to use, $G(n, p)$ is the model for the seminar work of Erdős and Rényi. Since then, $G(n, p)$ has been widely used and often been referred to as the Erdős-Rényi model. For $m = p\binom{n}{2}$, the two models are apparently correlated in the sense that many graph properties are satisfied by both random graph models. To precisely define the relationship of two random graph models, we need some definitions.

A graph property P can be viewed as a set of graphs. We say that a graph G satisfies property P if G is a member of P. A graph property is said to be *monotone* if whenever a graph H satisfies A, then any graph containing H must also satisfy A. For example, the property A of containing a specified subgraph, say, the Peterson graph, is a monotone property. A random graph G

is a probability distribution $\Pr(G = \cdot)$. Given two random graphs G_1 and G_2 on n vertices, we say that G_1 *dominates* G_2 if, for any monotone graph property A, the probability that a random graph from G_1 satisfies A is greater than or equal to the probability that a random graph from G_2 satisfies A, i.e.,

$$\Pr(G_1 \text{ satisfies } A) \geq \Pr(G_2 \text{ satisfies } A).$$

In this case, we write $G_1 \geq G_2$ and $G_2 \leq G_1$. For example, for any $p_1 \leq p_2$, we have $G(n, p_1) \leq G(n, p_2)$.

For any $\epsilon > 0$, we say that G_1 *dominates* G_2 with an error estimate ϵ if, for any monotone graph property A, the probability that a random graph from G_1 satisfies A is greater than or equal to the probability that a random graph from G_2 satisfies A up to an ϵ error term, i.e.,

$$\Pr(G_1 \text{ satisfies } A) + \epsilon \geq \Pr(G_2 \text{ satisfies } A).$$

If G_1 *dominates* G_2 with an error estimate $\epsilon = \epsilon_n$, which goes to zero as n approaches infinity, we say that G_1 almost surely dominates G_2. In this case, we write almost surely $G_1 \succeq G_2$ and $G_2 \preceq G_1$.

For example, for any $\delta > 0$, we have almost surely

$$G\left(n, (1-\delta)\frac{m}{\binom{n}{2}}\right) \preceq F(n, m) \preceq G\left(n, (1+\delta)\frac{m}{\binom{n}{2}}\right).$$

We can extend the definition of domination to graphs with different sizes in the following sense. Suppose that the random graph G_i has n_i vertices for $i = 1, 2$, and $n_1 < n_2$. By adding $n_2 - n_1$ isolated vertices, the random graph G_1 is extended to the random graph G_1' with the same size as G_2. We say that G_2 dominates G_1 if G_2 dominates G_1'.

We consider random graphs that are constructed inductively by *pivoting* at one edge at a time. Here we assume the number of vertices is n.

Edge-pivoting. For an edge $e \in K_n$, a probability q $(0 \leq q \leq 1)$, and a random graph G, a new random graph G' can be constructed in the following way. For any graph H, we define

$$\Pr(G' = H) \quad = \quad \begin{cases} (1-q)\Pr(G = H) & \text{if } e \notin E(H), \\ \Pr(G = H) + q\Pr(G = H \setminus \{e\}) & \text{if } e \in E(H). \end{cases}$$

It is easy to check that $\Pr(G' = \cdot)$ is a probability distribution. We say that G' is constructed from G by pivoting at the edge e with probability q.

For any graph property A, we define the set A_e to be

$$A_e = \{H \cup \{e\} | H \in A\}.$$

Further, we define the set $A_{\bar{e}}$ to be

$$A_{\bar{e}} = \{H \setminus \{e\} | H \in A\}.$$

In other words, A_e consists of the graphs obtained by adding the edge e to the graphs in A; $A_{\bar{e}}$ consists of the graphs obtained by deleting the edge e from the graphs in A. We have the following useful lemma.

Lemma 4.1. *Suppose that G' is constructed from G by pivoting at the edge e with probability q. Then for any property A, we have*

$$\Pr(G' \in A) = \Pr(G \in A) + q[\Pr((A \cap A_e)_{\bar{e}}) - \Pr(A \cap A_{\bar{e}})].$$

In particular, if A is a monotone property, we have

$$\Pr(G' \in A) \geq \Pr(G \in A).$$

Thus, G' dominates G.

Proof. The set associated with a property A can be partitioned into the following subsets. Let $A_1 = A \cap A_e$ be the graphs of A containing the edge e, and let $A_2 = A \cap A_{\bar{e}}$ be the graphs of A not containing the edge e. We have

$$\begin{aligned}
\Pr(G' \in A) &= \Pr(G' \in A_1) + \Pr(G' \in A_2) \\
&= \sum_{H \in A_1} \Pr(G' = H) + \sum_{H \in A_2} \Pr(G' = H) \\
&= \sum_{H \in A_1} (\Pr(G = H) + q \Pr(G = H \setminus \{e\})) \\
&\quad + \sum_{H \in A_2} (1 - q) \Pr(G = H) \\
&= \Pr(G \in A_1) + \Pr(G \in A_2) + q \Pr(G \in (A_1)_{\bar{e}}) - q \Pr(A_2) \\
&= \Pr(G \in A) + q[\Pr((A \cap A_e)_{\bar{e}}) - \Pr(A \cap A_{\bar{e}})].
\end{aligned}$$

If A is monotone, we have $A_2 \subset (A_1)_{\bar{e}}$. Thus,

$$\Pr(G' \in A) \geq \Pr(G \in A).$$

Lemma 4.1 is proved. □

Lemma 4.2. *Suppose that G'_i is constructed from G_i by pivoting the edge e with probability q_i, for $i = 1, 2$. If $q_1 \geq q_2$ and G_1 dominates G_2, then G'_1 dominates G'_2.*

Proof. Following the definitions of A, and letting A_1 and A_2 be as in the proof of Lemma 4.1, we have

$$\begin{aligned}
\Pr(G'_2 \in A) &= \Pr(G_2 \in A) + q_2[\Pr(G_2 \in (A_1)_{\bar{e}}) - \Pr(G_2 \in A_2)] \\
&= \Pr(G_2 \in A) + q_2 \Pr(G_2 \in ((A_1)_{\bar{e}} \setminus A_2)) \\
&\geq \Pr(G_1 \in A) + q_1 \Pr(G_1 \in ((A_1)_{\bar{e}} \setminus A_2)) \\
&= \Pr(G_1 \in A) + q_1[\Pr(G_1 \in (A_1)_{\bar{e}}) - \Pr(G_1 \in A_2)] \\
&= \Pr(G'_1 \in A).
\end{aligned}$$

The proof of Lemma 4.2 is complete. □

Let G_1 and G_2 be the random graphs on n vertices. We define $G_1 \cup G_2$ to be the random graph as follows:

$$\Pr(G_1 \cup G_2 = H) = \sum_{H_1 \cup H_2 = H} \Pr(G_1 = H_1) \Pr(G_2 = H_2)$$

where H_1, H_2 range over all possible pairs of subgraphs that are not necessarily disjoint.

The following lemma is a generalization of Lemma 4.2.

Lemma 4.3. *If G_1 dominates G_3 with an error estimate ϵ_1 and G_2 dominates G_4 with an error estimate ϵ_2, then $G_1 \cup G_2$ dominates $G_3 \cup G_4$ with an error estimate $\epsilon_1 + \epsilon_2$.*

Proof. For any monotone property A and any graph H, we define the set $f(A, H)$ to be

$$f(A, H) = \{G | G \cup H \in A\}.$$

We observe that $f(A, H)$ is also a monotone property. Therefore,

$$\begin{aligned}
\Pr(G_1 \cup G_2 \in A) &= \sum_{H \in A} \sum_{H_1 \cup H_2 = H} \Pr(G_1 = H_1) \Pr(G_2 = H_2) \\
&= \sum_{H_1} \Pr(G_1 = H_1) \Pr(G_2 \in f(A, H_1)) \\
&\geq \sum_{H_1} \Pr(G_1 = H_1)(\Pr(G_4 \in f(A, H_1)) - \epsilon_2) \\
&\geq \Pr(G_1 \cup G_4 \in A) - \epsilon_2.
\end{aligned}$$

Similarly, we have

$$\Pr(G_1 \cup G_4 \in A) \geq \Pr(G_3 \cup G_4 \in A) - \epsilon_1.$$

Thus, we get

$$\Pr(G_1 \cup G_2 \in A) \geq \Pr(G_3 \cup G_4 \in A) - (\epsilon_1 + \epsilon_2),$$

as desired. □

Suppose that ϕ is a sequence of random graphs $\phi(G_1), \phi(G_2), \ldots$, where the indices of ϕ range over all graphs on n vertices. Recall that a random graph G is a probability distribution $\Pr(G = \cdot)$ over the space of all graphs on n vertices. For any random graph G, we define $\phi(G)$ to be the random graph defined as follows:

$$\Pr(\phi(G) = H) = \sum_{H_1 \cup H_2 = H} \Pr(G = H_1) \Pr(\phi(H_1) = H_2).$$

We have the following lemmas.

Lemma 4.4. *Let ϕ_1 and ϕ_2 be two sequences of random graphs where the indices of ϕ_1 and ϕ_2 range over all graphs on n vertices. Let G be any random graph. If*

$$\Pr(G \in \{H | \phi_1(H) \textit{dominates } \phi_2(H) \textit{with an errorestimate } \epsilon_1\}) \geq 1 - \epsilon_2,$$

then $\phi_1(G)$ dominates $\phi_2(G)$ with an error estimate $\epsilon_1 + \epsilon_2$.

Proof. For any monotone property A and any graph H, we have

$$\begin{aligned}
\Pr(\phi_1(G) \in A) &= \sum_{H \in A} \sum_{H_1 \cup H_2 = H} \Pr(G = H_1) \Pr(\phi_1(H_1) = H_2) \\
&= \sum_{H_1} \Pr(G = H_1) \Pr(\phi_1(H_1) \in f(A, H_1)) \\
&\geq \sum_{H_1} \Pr(G = H_1) \Pr(\phi_2(H_1) \in f(A, H_1)) - \epsilon_1 - \epsilon_2 \\
&\geq \Pr(\phi_2(G) \in A) - (\epsilon_1 + \epsilon_2),
\end{aligned}$$

as desired, since $f(A, H) = \{G | G \cup H \in A\}$ is also a monotone property. □

Let G_1 and G_2 be the random graphs on n vertices. We define $G_1 \setminus G_2$ to be the random graph as follows:

$$\Pr(G_1 \setminus G_2 = H) = \sum_{H_1 \setminus H_2 = H} \Pr(G_1 = H_1) \Pr(G_2 = H_2),$$

where H_1 and H_2 range over all pairs of graphs.

Lemma 4.5. *If G_1 dominates G_3 with an error estimate ϵ_1 and G_2 is dominated by G_4 with an error estimate ϵ_2, then $G_1 \setminus G_2$ dominates $G_3 \setminus G_4$ with an error estimate $\epsilon_1 + \epsilon_2$.*

Proof. For any monotone property A and any graph H, we define the set $\psi(A,H)$ to be

$$\psi(A,H) = \{G | G \setminus H \in A\}.$$

We observe that $\psi(A,H)$ is also a monotone property. Therefore,

$$\begin{aligned}
\Pr(G_1 \setminus G_2 \in A) &= \sum_{H \in A} \sum_{H_1 \setminus H_2 = H} \Pr(G_1 = H_1) \Pr(G_2 = H_2) \\
&= \sum_{H_2} \Pr(G_2 = H_2) \Pr(G_1 \in \psi(A, H_2)) \\
&\geq \sum_{H_2} \Pr(G_2 = H_2)(\Pr(G_3 \in \psi(A, H_2)) - \epsilon_1) \\
&\geq \Pr(G_3 \setminus G_2 \in A) - \epsilon_1.
\end{aligned}$$

Similarly, we define the set $\theta(A,H)$ to be

$$\theta(A,H) = \{G | H \setminus G \in A\}.$$

We observe that the complement of the set $\theta(A,H)$ is a monotone property. We have

$$\begin{aligned}
\Pr(G_3 \setminus G_2 \in A) &= \sum_{H \in A} \sum_{H_1 \setminus H_2 = H} \Pr(G_3 = H_1) \Pr(G_2 = H_2) \\
&= \sum_{H_1} \Pr(G_3 = H_1) \Pr(G_2 \in \theta(A, H_1)) \\
&\geq \sum_{H_1} \Pr(G_3 = H_1)(\Pr(G_4 \in \theta(A, H_1)) - \epsilon_2) \\
&\geq \Pr(G_3 \setminus G_4 \in A) - \epsilon_2.
\end{aligned}$$

Thus, we get

$$\Pr(G_1 \cup G_2 \in A) \geq \Pr(G_3 \cup G_4 \in A) - (\epsilon_1 + \epsilon_2),$$

as desired. □

A random graph G is called *edge-independent* (or independent, for short) if there is an edge-weighted function $p \colon E(K_n) \to [0,1]$ satisfying

$$\Pr(G = H) = \prod_{e \in H} p_e \times \prod_{e \notin H} (1 - p_e).$$

For example, a random graph with a given expected degree sequence is edge-independent. Edge-independent random graphs have many nice properties, several of which we derive here.

Lemma 4.6. *Suppose that G and G' are independent random graph with edge-weighted functions p and p'; then, $G \cup G'$ is edge-independent with the edge-weighted function p'' satisfying*

$$p''_e = p_e + p'_e - p_e p'_e.$$

Proof. For any graph H, we have

$$\begin{aligned}
\Pr(G \cup G' = H) &= \sum_{H_1 \cup H_2 = H} \Pr(G = H_1)\Pr(G' = H_2) \\
&= \sum_{H_1 \cup H_2 = H} \prod_{e_1 \in H_1} p_{e_1} \prod_{e_2 \in H_2} p'_{e_2} \prod_{e_3 \notin H_1} (1 - p_{e_3}) \prod_{e_4 \notin H_2} (1 - p'_{e_4}) \\
&= \prod_{e \notin H} (1 - p_e)(1 - p'_e) \prod_{e \in H} (p_e(1 - p'_e) + (1 - p_e)p'_e + p_e p'_e) \\
&= \prod_{e \in H} p''_e \times \prod_{e \notin H} (1 - p''_e).
\end{aligned}$$

□

Lemma 4.7. *Suppose that G and G' are independent random graph with edge-weighted functions p and p'; then, $G \setminus G'$ is independent with the edge-weighted function p'' satisfying*

$$p''_e = p_e(1 - p'_e).$$

Proof. For any graph H, we have

$$\begin{aligned}
\Pr(G \setminus G' = H) &= \sum_{H_1 \setminus H_2 = H} \Pr(G = H_1)\Pr(G' = H_2) \\
&= \sum_{H_1 \setminus H_2 = H} \prod_{e_1 \in H_1} p_{e_1} \prod_{e_2 \in H_2} p'_{e_2} \prod_{e_3 \notin H_1} (1 - p_{e_3}) \prod_{e_4 \notin H_2} (1 - p'_{e_4}) \\
&= \prod_{e \in H} (p_e(1 - p'_e)) \prod_{e \notin H} (1 - p_e - p_e p'_e) \\
&= \prod_{e \in H} p''_e \times \prod_{e \notin H} (1 - p''_e).
\end{aligned}$$

□

Let $\{p_e\}_{e \in E(K_n)}$ be a probability distribution over all pairs of vertices. Let G_1 be the random graph of one edge, where a pair e of vertices is chosen with probability p_e. Inductively, we can define the random graph G_m by adding one more random edge to G_{m-1}, where a pair e of vertices is chosen (as the new edge) with probability p_e. (There is a small probability of having the same edges chosen more than once. In such cases, we will keep on sampling until we have exactly m different edges.) Hence, G_m has exactly m edges. The probability that G_m has edges $e_1, \ldots, e_m$ is proportional to $p_{e_1} p_{e_2} \cdots p_{e_m}$. The following lemma states that G_m can be sandwiched by two independent random graphs with exponentially small errors if m is large enough.

Lemma 4.8. *Assume that $p_e = o(\frac{1}{m})$ for all $e \in E(K_n)$. Let G' be the independent random graph with edge-weighted function $p'_e = (1-\delta)mp_e$. Let G'' be the independent random graph with edge-weighted function $p''_e = (1+\delta)mp_e$. Then, G_m dominates G' with error $e^{-\delta^2 m/4}$, and G_m is also dominated by G'' within an error estimate $e^{-\delta^2 m/4}$.*

Proof. For any Graph H, we define

$$f(H) = \prod_{e \in H} p_e.$$

For any graph property B, we define

$$f(B) = \sum_{H \in B} f(H).$$

Let C_k be the set of all graphs with exact k edges.

Claim 4.9. *For a graph monotone property A and an integer k, we have*

$$\frac{f(A \cap C_k)}{f(C_k)} \leq \frac{f(A \cap C_{k+1})}{f(C_{k+1})}.$$

Proof of Claim 4.9. Both $f(A \cap C_k) f(C_{k+1})$ and $f(A \cap C_{k+1}) f(C_k)$ are homogeneous polynomials on $\{p_e\}$ of degree $2k+1$. We compare the coefficients of a general monomial

$$p_{e_1}^2 \cdots p_{e_r}^2 p_{e_{r+1}} \cdots p_{e_{2k-r+1}}$$

in $f(A \cap C_k) f(C_{k+1})$ and $f(A \cap C_{k+1}) f(C_k)$. The coefficient c_1 of the monomial in $f(A \cap C_k) f(C_{k+1})$ is the number of $(k-r)$-subsets $\{e_{i_1}, e_{i_2}, \ldots, e_{i_{k-r}}\}$ of $e_{r+1}, \ldots, e_{2k-r+1}$ satisfying that the graph with edges

$$\{e_1, \ldots, e_r, e_{i_1}, e_{i_2}, \ldots, e_{i_{k-r}}\}$$

belongs to A_k. The coefficient c_2 of the monomial in $f(A \cap C_k) f(C_{k+1})$ is the number of $(k-r+1)$-subset $\{e_{i_1}, e_{i_2}, \ldots, e_{i_{k-r+1}}\}$ of $e_{r+1}, \ldots, e_{2k-r+1}$ satisfying that the graph with edges $\{e_1, \ldots, e_r, e_{i_1}, e_{i_2}, \ldots, e_{i_{k-r+1}}\}$ belongs to A_{k+1}. Since A is monotone, if the graph with edges $\{e_1, \ldots, e_r, e_{i_1}, e_{i_2}, \ldots, e_{i_{k-r}}\}$ belongs to A_k, then the graph with edges $\{e_1, \ldots, e_r, e_{i_1}, e_{i_2}, \ldots, e_{i_{k-r+1}}\}$ must belong to A_{k+1}. Hence, c_1 is always less than or equal to c_2. Thus, we have

$$f(A \cap C_k)f(C_{k+1}) \leq f(A \cap C_{k+1})f(C_k).$$

The claim is proved.

Now let $p'_e = \frac{(1-\delta)mp_e}{1+(1-\delta)mp_e} = (1+o(1))(1-\delta)mp_e$, or equivalently, $\frac{p'_e}{1-p'_e} = (1-\delta)mp_e$.

$$\begin{aligned}
\Pr(G' \in A) &= \sum_{k=0}^{n} \Pr(G' \in A \cap C_k) \\
&\leq \sum_{k=0}^{m} \Pr(G' \in A \cap C_k) + \sum_{k=m+1}^{n} \Pr(G' \in C_k) \\
&= \prod_{e \in E(K_n)} (1-p'_e) \sum_{k=0}^{m} ((1-\delta)m)^k f(A \cap C_k) \\
&\quad + \Pr(G' \text{ has more than } m \text{ edges}) \\
&\leq \prod_{e \in E(K_n)} (1-p'_e) \sum_{k=0}^{m} ((1-\delta)m)^k f(C_k) \frac{f(A \cap C_m)}{f(C_m)} \\
&\quad + \Pr(G' \text{ has more than } m \text{ edges}) \\
&\leq \frac{f(A \cap C_m)}{f(C_m)} \prod_{e \in E(K_n)} (1-p'_e) \sum_{k=0}^{m} ((1-\delta)m)^k f(C_k) \\
&\quad + \Pr(G' \text{ has more than } m \text{ edges}) \\
&= \frac{f(A \cap C_m)}{f(C_m)} \sum_{k=0}^{m} \Pr(G' \in C_k) + \Pr(G' \text{ has more than } m \text{ edges}) \\
&\leq \frac{f(A \cap C_m)}{f(C_m)} + \Pr(G' \text{ has more than } m \text{ edges}) \\
&= \Pr(G_m \in A) + \Pr(G' \text{ has more than } m \text{ edges}).
\end{aligned}$$

Now we estimate the probability that G' has more than m edges. Let X_e be the 0-1 random variable with $\Pr(X_e = 1) = p'_e$. Let $X = \sum_e X_e$. Then, $E(X) = (1+o(1))m(1-\delta)$. Now we apply the following large deviation inequality:

$$\Pr(X - E(X) > a) \leq e^{-\frac{a^2}{2(E(X)+a/3)}}.$$

We have

$$\begin{aligned}\Pr(X > m) &= \Pr(X - E(X) > (1+o(1))\delta m)\\ &\leq e^{-(1+o(1))\frac{\delta^2 m^2}{2(1-\delta)m + 2\delta m/3}}\\ &\leq e^{-\delta^2 m/2}.\end{aligned}$$

For the other direction, let $p''_e = \frac{(1+\delta)mp_e}{1+(1+\delta)mp_e} = (1+o(1))(1+\delta)mp_e$, which implies that $\frac{p''_e}{1-p''_e} = (1+\delta)mp_e$.

$$\begin{aligned}\Pr(G'' \in A) &= \sum_{k=0}^{n} \Pr(G'' \in A \cap C_k)\\ &\geq \sum_{k=m}^{n} \Pr(G' \in A \cap C_k)\\ &= \prod_e (1-p''_e) \sum_{k=m}^{n} ((1+\delta)m)^k f(A \cap C_k)\\ &\geq \prod_e (1-p'_e) \sum_{k=m}^{n} ((1+\delta)m)^k f(C_k) \frac{f(A \cap C_m)}{f(C_m)}\\ &\geq \frac{f(A \cap C_m)}{f(C_m)} \prod_e (1-p'_e) \sum_{k=m}^{n} ((1+\delta)m)^k f(C_k)\\ &= \frac{f(A \cap C_m)}{f(C_m)} \left(1 - \sum_{k=0}^{m-1} \Pr(G' \in C_k)\right)\\ &\geq \frac{f(A \cap C_m)}{f(C_m)} - \Pr(G'' \text{ has less than } m \text{ edges})\\ &= \Pr(G_m \in A) - \Pr(G'' \text{ has less than } m \text{ edges})\end{aligned}$$

Now we estimate the probability that G'' has less than m edges. Let X_e be the 0-1 random variable with $\Pr(X_e = 1) = p''_e$. Let $X = \sum_e X_e$. Then $E(X) = (1+o(1))m(1+\delta)$. Now we apply the following large deviation inequality:

$$\Pr(X - E(X) < a) \leq e^{-\frac{a^2}{2E(X)}}.$$

We have

$$\begin{aligned}\Pr(X < m) &= \Pr(X - E(X) < (1+o(1))\delta m)\\ &\leq e^{-(1+o(1))\frac{\delta^2 m^2}{2(1+\delta)m}}\\ &\leq e^{-\delta^2 m/3}.\end{aligned}$$

The proof of Lemma 4.8 is completed. □

5. General Martingale Inequalities

In this subsection, we will extend and generalize the Azuma inequality to a martingale that is not strictly Lipschitz but is nearly Lipschitz. Similar techniques have been introduced by Kim and Vu [Kim and Vu 00] in their important work on deriving concentration inequalities for multivariate polynomials. Here we use a rather general setting, and we shall give a complete proof.

Suppose that Ω is a probability space and $\mathcal{F}$ is a σ-field; X is a random variable that is $\mathcal{F}$-measurable. (The reader is referred to [Janson et al. 00] for the terminology on martingales.) A *filter* $\mathbf{F}$ is an increasing chain of σ-subfields

$$\{0, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n = \mathcal{F}.$$

A martingale (obtained from) X associated with a filter $\mathbf{F}$ is a sequence of random variables $X_0, X_1, \ldots, X_n$ with $X_i = E(X \mid \mathcal{F}_i)$ and, in particular, $X_0 = E(X)$ and $X_n = X$.

For $\mathbf{c} = (c_1, c_2, \ldots, c_n)$ a positive vector, the martingale X is said to be c-Lipschitz if $|X_i - X_{i-1}| \leq c_i$ for $i = 1, 2, \ldots, n$. A powerful tool for controlling martingales is the following:

Azuma's inequality. If a martingale X is c-Lipschitz, then

$$\Pr(|X - E(X)| < a) \leq 2e^{-\frac{a^2}{2\sum_{i=1}^{n} c_i^2}},$$

where $c = (c_1, \ldots, c_n)$.

Here we are only interested in finite probability spaces, and we use the following computational model. The random variable X can be evaluated by a sequence of decisions $Y_1, Y_2, \ldots, Y_n$. Each decision has no more than r outputs. The probability that an output is chosen depends on the previous history. We can describe the process by a decision tree T; T is a complete rooted r-tree with depth n. Each edge uv of T is associated with a probability p_{uv} depending on the decision made from u to v. We allow p_{uv} to be zero and thus include the case of having fewer than r outputs. Let Ω_i denote the probability space obtained after the first i decisions. Suppose that $\Omega = \Omega_n$ and X is the random variable on Ω. Let $\pi_i \colon \Omega \to \Omega_i$ be the projection mapping each point to its first i coordinates. Let $\mathcal{F}_i$ be the σ-field generated by $Y_1, Y_2, \ldots, Y_i$. (In fact, $\mathcal{F}_i = \pi^{-1}(2^{\Omega_i})$ is the full σ-field via the projection π_i.) $\mathcal{F}_i$ forms a natural filter:

$$\{0, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n = \mathcal{F}.$$

Any vertex u of T is associated with a real value $f(u)$. If u is a leaf, we define $f(u) = X(u)$. For a general u, here are several equivalent definitions for $f(u)$.

1. For any non-leaf node u, $f(u)$ is the weighted average over the f-values of the children of u:
$$f(u) = \sum_{i=1}^{r} p_{uv_i} f(v_i),$$
where $v_1, v_2, \ldots, v_r$ are the children of u.

2. For a non-leaf node u, $f(u)$ is the weighted average over all leaves in the sub-tree T_u rooted at u:
$$f(u) = \sum_{v \text{ leaf } in \ T_u} p_u(v) f(v),$$
where $p_u(v)$ denotes the product of edge-weights over edges in the unique path from u to v.

3. Let X_i be a random variable of Ω, which for each node u of depth i assumes the value $f(u)$ for every leaf in the subtree T_u. Then, $X_0, X_1, \ldots, X_n$ form a martingale, i.e., $X_i = E(X_n \mid \mathcal{F}_i)$. In particular, $X = X_n$ is the restriction of f to leaves of T.

We note that the Lipschitz condition $|X_i - X_{i-1}| \leq c_i$ is equivalent to
$$|f(u) - f(v)| \leq c_i$$
for any edge uv from a vertex u with depth $i-1$ to a vertex v with depth i.

We say an edge uv is *bad* if $|f(u) - f(v)| > c_i$. We say a node u is *good* if the path from the root to u does not contain any node of a bad edge.

The following theorem further generalizes the Azuma's Inequality. A similar but more restricted version can be found in [Kim and Vu 00].

Theorem 5.1. *For any $c_1, c_2, \ldots, c_n$, a martingale X satisfies*
$$\Pr(|X - E(X)| < a) \leq 2e^{-\frac{a^2}{2\sum_{i=1}^{n} c_i^2}} + \Pr(B),$$
where B is the set of all bad leaves of the decision tree associated with X.

Proof. We define a modified labeling f' on T so that $f'(u) = f(u)$ if u is a good node in T. For each bad node u, let xy be the first bad edge that intersects the path from the root to u at x. We define $f'(u) = f(x)$.

Claim 5.2. $f'(u) = \sum_{i=1}^{r} p_{uv_i} f'(v_i)$, for any u with children $v_1, \ldots, v_r$.

If u is a good vertex, we always have $f'(v_i) = f(v_i)$ whether v_i is good or not. Since $f(u) = \sum_{i=1}^{r} p_{uv_i} f(v_i)$, we have $f'(u) = \sum_{i=1}^{r} p_{uv_i} f'(v_i)$.

If u is a bad vertex, $v_1, \ldots, v_r$ are all bad by the definition. We have $f'(u) = f'(v_1) = \cdots = f'(v_r)$. Hence, $\sum_{i=1}^r p_{uv_i} f'(v_i) = f(u) \sum_{i=1}^r p_{uv_i} = f(u)$.

Claim 5.3. f' is c-Lipschitz.

For any edge uv with u of depth $i-1$ and v of depth i, if u is a good vertex, then uv is a good edge, and

$$|f'(u) - f'(v)| \leq c_i.$$

If u is a bad vertex, we have $f'(u) = f'(v)$, and thus,

$$|f'(u) - f'(v)| \leq c_i.$$

Let X' be the random variable that is the restriction of f' to the leaves; X' is c-Lipschitz. We can apply Azuma's Inequality to X'. Namely,

$$\Pr(|X' - E(X')| < a) \leq 2e^{-\frac{a^2}{2\sum_{i=1}^n c_i^2}}.$$

From the definition of f, we have

$$E(X') = E(X).$$

Let B denote the set of bad leaves in the decision T of X. Clearly, we have

$$\Pr(u : X(u) \neq X'(u)) \leq \Pr(B).$$

Therefore, we have

$$\begin{aligned} \Pr(|X - E(X)| < a) &\leq \Pr(X \neq X') + \Pr(|X' - E(X')| < a) \\ &\leq 2e^{-\frac{a^2}{2\sum_{i=1}^n c_i^2}} + \Pr(B). \end{aligned}$$

The proof of the theorem is complete. $\square$

For some applications, even nearly Lipschitz condition is still not feasible. Here we consider an extension of Azuma's inequality. Our starting point is the following well-known concentration inequality (see [McDiarmid 98]).

Theorem 5.4. *Let X be the martingale associated with a filter $\mathbf{F}$ satisfying*

1. $\mathrm{Var}(X_i | \mathcal{F}_{i-1}) \leq \sigma_i^2$, *for* $1 \leq i \leq n$;
2. $|X_i - X_{i-1}| \leq M$, *for* $1 \leq i \leq n$.

Then, we have

$$\Pr(X - E(X) \geq a) \leq e^{-\frac{a^2}{2(\sum_{i=1}^n \sigma_i^2 + Ma/3)}}.$$

In this paper, we consider a strenghtened version of the above inequality where the variance $\mathrm{Var}(X_i|\mathcal{F}_{i-1})$ is instead upper bounded by a constant factor of X_{i-1}.

We first need some terminology. For a filter $\mathbf{F}$,

$$\{0, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n = \mathcal{F}.$$

A sequence of random variables $X_0, X_1, \ldots, X_n$ is called a *submartingale* if X_i is $\mathcal{F}_i$-measurable and $E(X_i \mid \mathcal{F}_{i-1}) \leq X_{i-1}$, for $1 \leq i \leq n$.

A sequence of random variables $X_0, X_1, \ldots, X_n$ is said to be a *supermartingale* if X_i is $\mathcal{F}_i$-measurable and $E(X_i \mid \mathcal{F}_{i-1}) \geq X_{i-1}$, for $1 \leq i \leq n$.

We have the following theorem.

Theorem 5.5. *Suppose that a submartingale X, associated with a filter $\mathbf{F}$, satisfies*

$$\mathrm{Var}(X_i|\mathcal{F}_{i-1}) \leq \phi_i X_{i-1}$$

and

$$X_i - E(X_i|\mathcal{F}_{i-1}) \leq M$$

for $1 \leq i \leq n$. Then, we have

$$\Pr(X_n > X_0 + a) \leq e^{-\frac{a^2}{2((X_0+a)(\sum_{i=1}^n \phi_i)+Ma/3)}}.$$

Proof. For a positive λ (to be chosen later), we consider

$$\begin{aligned} E(e^{\lambda X_i}|\mathcal{F}_{i-1}) &= e^{\lambda E(X_i|\mathcal{F}_{i-1})} E(e^{\lambda(X_i - E(X_i|\mathcal{F}_{i-1}))}|\mathcal{F}_{i-1}) \\ &= e^{\lambda E(X_i|\mathcal{F}_{i-1})} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} E((X_i - E(X_i|\mathcal{F}_{i-1})^k)|\mathcal{F}_{i-1}) \\ &\leq e^{\lambda E(X_i|\mathcal{F}_{i-1}) + \sum_{k=2}^{\infty} \frac{\lambda^k}{k!} E((X_i - E(X_i|\mathcal{F}_{i-1})^k)|\mathcal{F}_{i-1})}. \end{aligned}$$

Let $g(y) = 2\sum_{k=2}^{\infty} \frac{y^{k-2}}{k!} = \frac{2(e^y-1-y)}{y^2}$. We use the following facts:

- $g(y) \leq 1$, for $y < 0$.
- $\lim_{y \to 0} g(y) = 1$.
- $g(y)$ is monotone increasing, when $y \geq 0$.

When $b < 3$, we have

$$g(b) = 2\sum_{k=2}^{\infty}\frac{b^{k-2}}{k!} \le \sum_{k=2}^{\infty}\frac{b^{k-2}}{3^{k-2}} = \frac{1}{1-b/3}. \tag{5.1}$$

Since $X_i - E(X_i|\mathcal{F}_{i-1}) \le M$, we have

$$\sum_{k=2}^{\infty}\frac{\lambda^k}{k!}E((X_i - E(X_i|\mathcal{F}_{i-1})^k)|\mathcal{F}_{i-1}) \le \frac{g(\lambda M)}{2}\lambda^2\mathrm{Var}(X_i|\mathcal{F}_{i-1}).$$

We define $\lambda_i \ge 0$ for $0 < i \le n$, satisfying $\lambda_{i-1} = \lambda_i + \frac{g(\lambda_0 M)}{2}\phi_i\lambda_i^2$, while λ_0 will be chosen later. Then,

$$\lambda_n \le \lambda_{n-1} \le \cdots \le \lambda_0,$$

and

$$\begin{aligned} E(e^{\lambda_i X_i}|\mathcal{F}_{i-1}) &\le e^{\lambda_i E(X_i|\mathcal{F}_{i-1}) + \frac{g(\lambda_i M)}{2}\lambda_i^2\mathrm{Var}(X_i|\mathcal{F}_{i-1})} \\ &\le e^{\lambda_i X_{i-1} + \frac{g(\lambda_0 M)}{2}\lambda_i^2\phi_i X_{i-1}} \\ &= e^{\lambda_{i-1}X_{i-1}}, \end{aligned}$$

since $g(y)$ is increasing for $y > 0$.

By Markov's inequality, we have

$$\begin{aligned} \Pr(X_n > X_0 + a) &\le e^{-\lambda_n(X_0+a)}E(e^{\lambda_n X_n}) \\ &= e^{-\lambda_n(X_0+a)}E(E(e^{\lambda_n X_n}|\mathcal{F}_{n-1})) \\ &\le e^{-\lambda_n(X_0+a)}E(e^{\lambda_{n-1}X_{n-1}}) \\ &\vdots \\ &\le e^{-\lambda_n(X_0+a)}E(e^{\lambda_0 X_0}) \\ &= e^{-\lambda_n(X_0+a)+\lambda_0 X_0}. \end{aligned}$$

Note that

$$\begin{aligned} \lambda_n &= \lambda_0 - \sum_{i=1}^{n}(\lambda_{i-1} - \lambda_i) \\ &= \lambda_0 - \sum_{i=1}^{n}\frac{g(\lambda_0 M)}{2}\phi_i\lambda_i^2 \\ &\ge \lambda_0 - \frac{g(\lambda_0 M)}{2}\lambda_0^2\sum_{i=1}^{n}\phi_i. \end{aligned}$$

Hence,

$$\begin{aligned} \Pr(X_n > X_0 + a) &\le e^{-\lambda_n(X_0+a)+\lambda_0 X_0} \\ &\le e^{-(\lambda_0 - \frac{g(\lambda_0 M)}{2}\lambda_0^2\sum_{i=1}^{n}\phi_i)(X_0+a)+\lambda_0 X_0} \\ &= e^{-\lambda_0 a + \frac{g(\lambda_0 M)}{2}\lambda_0^2(X_0+a)\sum_{i=1}^{n}\phi_i}. \end{aligned}$$

Now we choose $\lambda_0 = \frac{a}{(X_0+a)(\sum_{i=1}^n \phi_i)+Ma/3}$. Using the fact that $\lambda_0 M < 3$ and Inequality (5.1), we have

$$\begin{aligned}\Pr(X_n > X_0 + a) &\leq e^{-\lambda_0 a + \lambda_0^2 (X_0+a)\sum_{i=1}^n \phi_i \frac{1}{2(1-\lambda_0 M/3)}}\\ &\leq e^{-\frac{a^2}{2((X_0+a)(\sum_{i=1}^n \phi_i)+Ma/3)}}.\end{aligned}$$

The proof of the theorem is finished. □

The condition of Theorem 5.5 can be further relaxed using the same technique as in Theorem 5.1, and we have the following theorem. The proof will be omitted.

Theorem 5.6. *For a filter* $\mathbf{F}$

$$\{0, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n = \mathcal{F},$$

suppose that a random variable X_j *is* $\mathcal{F}_i$*-measurable, for* $1 \leq i \leq n$. *Let* B_1 *be the bad set in the decision tree associated with* X*s where at least one of the following conditions is violated:*

$$\begin{aligned} E(X_i \mid \mathcal{F}_{i-1}) &\leq X_{i-1},\\ \mathrm{Var}(X_i|\mathcal{F}_{i-1}) &\leq \phi_i X_{i-1},\\ X_i - E(X_i|\mathcal{F}_{i-1}) &\leq M.\end{aligned}$$

Then, we have

$$\Pr(X_n > X_0 + a) \leq e^{-\frac{a^2}{2((X_0+a)(\sum_{i=1}^n \phi_i)+Ma/3)}} + \Pr(B_1).$$

The theorem for supermartingale is slightly different due to the asymmetry of the condition on variance.

Theorem 5.7. *Suppose that a supermartingale* X, *associated with a filter* $\mathbf{F}$, *satisfies, for* $1 \leq i \leq n$,

$$\mathrm{Var}(X_i|\mathcal{F}_{i-1}) \leq \phi_i X_{i-1}$$

and

$$E(X_i|\mathcal{F}_{i-1}) - X_i \leq M.$$

Then, we have

$$\Pr(X_n < X_0 - a) \leq e^{-\frac{a^2}{2(X_0(\sum_{i=1}^n \phi_i)+Ma/3)}},$$

for any $a \leq X_0$.

Proof. The proof is similar to that of Theorem 5.5. The following inequality still holds:

$$\begin{aligned} E(e^{-\lambda X_i}|\mathcal{F}_{i-1}) &= e^{-\lambda E(X_i|\mathcal{F}_{i-1})}E(e^{-\lambda(X_i-E(X_i|\mathcal{F}_{i-1}))}|\mathcal{F}_{i-1}) \\ &= e^{-\lambda E(X_i|\mathcal{F}_{i-1})}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}E((E(X_i|\mathcal{F}_{i-1})-X_i)^k)|\mathcal{F}_{i-1}) \\ &\le e^{-\lambda E(X_i|\mathcal{F}_{i-1})+\sum_{k=2}^{\infty}\frac{\lambda^k}{k!}E((E(X_i|\mathcal{F}_{i-1})-X_i)^k)|\mathcal{F}_{i-1})} \\ &\le e^{-\lambda_i E(X_i|\mathcal{F}_{i-1})+\frac{g(\lambda M)}{2}\lambda^2 \mathrm{Var}(X_i|\mathcal{F}_{i-1})} \\ &\le e^{-\lambda_i X_{i-1}+\frac{g(\lambda M)}{2}\lambda^2\phi_i X_{i-1}}. \end{aligned}$$

We now define $\lambda_i \ge 0$, for $0 \le i < n$ satisfying $\lambda_{i-1} = \lambda_i - \frac{g(\lambda_n)}{2}\phi_i\lambda_i^2$; λ_n will be defined later. Then, we have

$$\lambda_0 \le \lambda_1 \le \cdots \le \lambda_n,$$

and

$$\begin{aligned} E(e^{-\lambda_i X_i}|\mathcal{F}_{i-1}) &\le e^{-\lambda_i E(X_i|\mathcal{F}_{i-1})+\frac{g(\lambda_i M)}{2}\lambda_i^2 \mathrm{Var}(X_i|\mathcal{F}_{i-1})} \\ &\le e^{-\lambda_i X_{i-1}+\frac{g(\lambda_n M)}{2}\lambda_i^2\phi_i X_{i-1}} \\ &= e^{-\lambda_{i-1}X_{i-1}}. \end{aligned}$$

By Markov's inequality, we have

$$\begin{aligned} \Pr(X_n < X_0 - a) &= \Pr(-\lambda_n X_n > -\lambda_n(X_n - a)) \\ &\le e^{\lambda_n(X_0-a)}E(e^{-\lambda_n X_n}) \\ &= e^{\lambda_n(X_0-a)}E(E(e^{-\lambda_n X_n}|\mathcal{F}_{n-1})) \\ &\le e^{\lambda_n(X_0-a)}E(e^{-\lambda_{n-1}X_{n-1}}) \\ &\vdots \\ &\le e^{\lambda_n(X_0-a)}E(e^{-\lambda_0 X_0}) \\ &= e^{\lambda_n(X_0-a)-\lambda_0 X_0}. \end{aligned}$$

We note that

$$\begin{aligned} \lambda_0 &= \lambda_n + \sum_{i=1}^{n}(\lambda_{i-1}-\lambda_i) \\ &= \lambda_n - \sum_{i=1}^{n}\frac{g(\lambda_n M)}{2}\phi_i\lambda_i^2 \\ &\ge \lambda_n - \frac{g(\lambda_n M)}{2}\lambda_n^2\sum_{i=1}^{n}\phi_i. \end{aligned}$$

Thus, we have

$$\begin{aligned}\Pr(X_n < X_0 - a) &\le e^{\lambda_n(X_0-a)-\lambda_0 X_0}\\ &\le e^{\lambda_n(X_0-a)-(\lambda_n - \frac{g(\lambda_n M)}{2}\lambda_n^2 \sum_{i=1}^n \phi_i)X_0}\\ &= e^{-\lambda_n a + \frac{g(\lambda_n M)}{2}\lambda_n^2 X_0 \sum_{i=1}^n \phi_i}.\end{aligned}$$

We choose $\lambda_n = \frac{a}{X_0(\sum_{i=1}^n \phi_i)+Ma/3}$. We have $\lambda_n M < 3$ and

$$\begin{aligned}\Pr(X_n < X_0 - a) &\le e^{-\lambda_n a + \lambda_n^2 X_0 \sum_{i=1}^n \phi_i \frac{1}{2(1-\lambda_n M/3)}}\\ &\le e^{-\frac{a^2}{2(X_0(\sum_{i=1}^n \phi_i)+Ma/3)}}.\end{aligned}$$

It remains to verify that all λ_i are nonnegative. Indeed,

$$\begin{aligned}\lambda_i &\ge \lambda_0\\ &\ge \lambda_n - \frac{g(\lambda_n M)}{2}\lambda_n^2 \sum_{i=1}^n \phi_i\\ &\ge \lambda_n\left(1 - \frac{1}{2(1-\lambda_n M/3)}\lambda_n \sum_{i=1}^n \phi_i\right)\\ &= \lambda_n(1 - \frac{a}{2X_0})\\ &\ge 0.\end{aligned}$$

The proof of the theorem is complete. □

Again, the above theorem can further relaxed as follows.

Theorem 5.8. *For a filter* $\mathbf{F}$

$$\{0, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n = \mathcal{F},$$

suppose that a random variable X_j *is* $\mathcal{F}_i$*-measurable, for* $1 \le i \le n$. *Let* B_2 *be the bad set where at least one of the following conditions is violated:*

$$\begin{aligned}E(X_i \mid \mathcal{F}_{i-1}) &\ge X_{i-1},\\ \mathrm{Var}(X_i|\mathcal{F}_{i-1}) &\le \phi_i X_{i-1},\\ E(X_i|\mathcal{F}_{i-1}) - X_i &\le M.\end{aligned}$$

Then, we have

$$\Pr(X_n < X_0 - a) \le e^{-\frac{a^2}{2(X_0(\sum_{i=1}^n \phi_i)+Ma/3)}} + \Pr(B_2),$$

for any $a \le X_0$.

6. Main Theorems for the Growth-Deletion Model

We say that a random graph G is "almost surely edge-independent" if there are two edge-independent random graphs G_1 and G_2 on the same vertex set satisfying the following:

1. G dominates G_1.
2. G is dominated by G_2.
3. For any two vertices u and v, let $p_{uv}^{(i)}$ be the probability of edge uv in G_i for $i = 1, 2$. We have
$$p_{uv}^{(1)} = (1 - o(1))p_{uv}^{(2)}.$$

We will prove the following:

Theorem 6.1. *Suppose that $p_3 < p_1$, $p_4 < p_2$, and $\log n \ll m < t^{\frac{p_1}{2(p_1+p_2)}}$. Then,*

1. *Almost surely the degree sequence of the growth-deletion model*
$$G(p_1, p_2, p_3, p_4, m)$$
follows the power law distribution with the exponent
$$\beta = 2 + \frac{p_1 + p_3}{p_1 + 2p_2 - p_3 - 2p_4}.$$
2. *$G(p_1, p_2, p_3, p_4, m)$ is almost surely edge-independent. It dominates and is dominated by an edge-independent graph with probability $p_{ij}^{(t)}$ of having an edge between vertices i and j, $i < j$, at time t, satisfying:*

$$p_{ij}^{(t)} \approx \begin{cases} \frac{p_2 m}{2p_4\tau(2p_2-p_4)} \frac{l^{2\alpha-1}}{i^\alpha j^\alpha} \left(1 + \left(1 - \frac{p_4}{p_2}\right)\left(\frac{j}{t}\right)^{\frac{1}{2\tau}+2\alpha-1}\right) & \text{if } i^\alpha j^\alpha \gg \frac{p_2 m t^{2\alpha-1}}{4\tau^2 p_4} \\ 1 - (1 + o(1))\frac{2p_4\tau}{p_2 m} i^\alpha j^\alpha t^{1-2\alpha} & \text{if } i^\alpha j^\alpha \ll \frac{p_2 m t^{2\alpha-1}}{4\tau^2 p_4} \end{cases}$$

where $\alpha = \frac{p_1(p_1+2p_2-p_3-2p_4)}{2(p_1+p_2-p_4)(p_1-p_3)}$ and $\tau = \frac{(p_1+p_2-p_4)(p_1-p_3)}{p_1+p_3}$.

Without the assumption on m, we have the following general but weaker result:

Theorem 6.2. *In $G(p_1, p_2, p_3, p_4, m)$ with $p_3 < p_1$ and $p_4 < p_2$, let S be the set of vertices with index i satisfying*
$$i \gg m^{\frac{1}{\alpha}} t^{1-\frac{1}{2\alpha}}.$$
Let G_S be the induced subgraph of $G(p_1, p_2, p_3, p_4, m)$ on S. Then,

1. *G_S dominates a random power law graph G_1, in which the expected degrees are given by*

$$d_i \approx \frac{p_2 m}{2p_4\tau(2p_2 - p_4)(\frac{p_1}{p_1-p_3} - \alpha)} \frac{t^\alpha}{i^\alpha}.$$

2. *G_S is dominated by a random power law graph G_2, in which the expected degrees are given by*

$$d_i \approx \frac{m}{2p_4\tau(\frac{p_1}{p_1-p_3} - \alpha)} \frac{t^\alpha}{i^\alpha}.$$

Theorem 6.3. *In $G(p_1, p_2, p_3, p_4, m)$ with $p_3 < p_1$ and $p_4 < p_2$, let T be the set of vertices with index i satisfying*

$$i \ll m^{\frac{1}{\alpha}} t^{1-\frac{1}{2\alpha}}.$$

Then, the induced subgraph G_T of $G(p_1, p_2, p_3, p_4, m)$ is almost a complete graph. Namely, G_T dominates an edge-independent graph with $p_{ij} = 1 - o(1)$.

Let n_t (or τ_t) be the number of vertices (or edges) at time t. We assume that the initial graph has n_0 vertices and τ_0 edges. When t is large enough, the graph at time t depends on the initial graph only in a mild manner. The number of vertices n_0 and edges τ_0 in the initial graph affect only a lower order term to random variables under consideration. We first establish the following lemmas on the number of vertices and the number of edges.

Lemma 6.4. *For any t and $k > 1$, in $G(p_1, p_2, p_3, p_4, m)$ with an initial graph on n_0 vertices, the number n_t of vertices at time t satisfies*

$$(p_1 - p_3)t - \sqrt{2kt\log t} \le n_t - n_0 \le (p_1 - p_3)t + \sqrt{2kt\log t}, \tag{6.1}$$

with probability at least $1 - \frac{2}{t^k}$.

Proof. The expected number of vertices n_t satisfies the following recurrence relation:

$$E(n_{t+1}) = E(n_t) + p_1 - p_3.$$

Hence, $E(n_{t+1}) = n_0 + (p_1 - p_3)t$. Since we assume that $p_3 < p_1$, the graph grows as time t increases. By Azuma's martingale inequality, we have

$$\Pr(|n_t - E(n_t)| > a) \le 2e^{-\frac{a^2}{2t}}.$$

By choosing $a = \sqrt{2kt\log t}$, with probability at least $1 - \frac{2}{t^k}$, we have

$$(p_1 - p_3)t - \sqrt{2kt\log t} \leq n_t - n_0 \leq (p_1 - p_3)t + \sqrt{2kt\log t}. \tag{6.2}$$

□

Lemma 6.5. *The number τ_t of edges in $G(p_1, p_2, p_3, p_4, m)$, with an initial graph on n_0 vertices and τ_0 edges, satisfies at time t*

$$|E(\tau_t) - \tau_0 - \tau mt| = O(\sqrt{t\log t}),$$

where $\tau = \frac{(p_1+p_2-p_4)(p_1-p_3)}{p_1+p_3}$.

Proof. The expected number of edges satisfies

$$E(\tau_{t+1}) = E(\tau_t) + mp_1 + mp_2 - p_3E(\frac{2\tau_t}{n_t}) - mp_4. \tag{6.3}$$

Let C denote a large constant satisfying the following:

1. $C > \frac{8p_3}{(p_1-p_3)^2}$.

2. $C > 4\sqrt{\frac{s}{\log s}}$ for some large constant s.

We shall inductively prove the following inequality:

$$|E(\tau_t) - \tau_0 - m\tau t| < Cm\sqrt{t\log t} \quad \text{ for } t \geq s. \tag{6.4}$$

When $t = s$, we have

$$|E(\tau_s) - \tau_0 - m\tau s| \leq 2ms \leq Cm\sqrt{s\log s},$$

by the definition of C.

By the induction assumption, we assume that $|E(\tau_t) - \tau_0 - \tau mt| \leq C\sqrt{t\log t}$ holds. Then, we consider

$$\begin{aligned}
&|E(\tau_{t+1}) - \tau_0 - \tau m(t+1)| \\
&= \left|E(\tau_t) + mp_1 + mp_2 - p_3E\left(\frac{2\tau_t}{n_t}\right) - mp_4 - \tau_0 - \tau m(t+1)\right| \\
&= \left|E(\tau_t) - \tau_0 - \tau mt - 2p_3E\left(\frac{\tau_t}{n_t}\right) + 2p_3\frac{m\tau}{p_1 - p_3}\right|
\end{aligned}$$

$$
\begin{aligned}
&= \left| \left(1 - \frac{2p_3}{(p_1-p_3)t}\right)(E(\tau_t) - m\tau t - \tau_0) \right.\\
&\quad \left. - 2p_3\left(E\left(\frac{\tau_t}{n_t}\right) - \frac{E(\tau_t)}{(p_1-p_3)t}\right) - \frac{2p_3}{(p_1-p_3)t}\tau_0 \right|\\
&\le \left| \left(1 - \frac{2p_3}{(p_1-p_3)t}\right)(E(\tau_t) - m\tau t - \tau_0) \right|\\
&\quad + 2p_3\left| E\left(\frac{\tau_t}{n_t}\right) - \frac{E(\tau_t)}{(p_1-p_3)t}\right| + \frac{2p_3}{(p_1-p_3)t}\tau_0\\
&\le |E(\tau_t) - \tau_0 - \tau m t| + 2p_3\left| E\left(\tau_t\left(\frac{1}{n_t} - \frac{1}{(p_1-p_3)t}\right)\right)\right| + O\left(\frac{1}{t}\right).
\end{aligned}
$$

We wish to substitute n_t by $nt + n_0 + O(\sqrt{2kt\log t})$ if possible. However,

$$
E\left(\tau_t\left(\frac{1}{n_t} - \frac{1}{(p_1-p_3)t}\right)\right)
$$

can be large. We consider S, the event that $|n_t - n_0 - (p_1+p_3)t| < 4\sqrt{t\log t}$. We have $\Pr(S) > 1 - \frac{1}{t^2}$ from Lemma 6.4. Let $\mathbf{1}_S$ be the indicator random variable for the event S, and $\bar{S}$ denotes the complement event of S. We can derive an upper bound for $|E((\tau_{t+1} - \tau_0 - \tau m(t+1))\mathbf{1}_S)|$ in a similar argument as above and obtain

$$
\begin{aligned}
&|E((\tau_{t+1} - \tau_0 - \tau m(t+1))\mathbf{1}_S)|\\
&\qquad \le |E((\tau_t - \tau_0 - \tau m t)\mathbf{1}_S)|\\
&\qquad\quad + 2p_3\left| E\left(\tau_t\left(\frac{1}{n_t} - \frac{1}{(p_1-p_3)t}\right)\mathbf{1}_S\right)\right| + O\left(\frac{1}{t}\right).
\end{aligned} \tag{6.5}
$$

We consider each term in the last inequality separately.

$$
\begin{aligned}
&2p_3\left| E\left(\tau_t\left(\frac{1}{n_t} - \frac{1}{(p_1-p_3)t}\right)\mathbf{1}_S\right)\right|\\
&\qquad \le 2p_3 m t\left| \frac{1}{(p_1-p_3)t - 4\sqrt{t\log t}} - \frac{1}{(p_1-p_3)t}\right|\\
&\qquad \le \frac{\sqrt{8p_3}}{(p_1-p_3)^2}\sqrt{\frac{\log t}{t}} + O\left(\frac{\log t}{t}\right).
\end{aligned} \tag{6.6}
$$

Since $\Pr(\bar{S}) \le \frac{1}{t^2}$ and $\tau_t \le \tau_0 + mt$, we have

$$
\begin{aligned}
&|E((\tau_{t+1} - \tau_0 - \tau m(t+1)))|\\
&= |E((\tau_{t+1} - \tau_0 - \tau m(t+1))\mathbf{1}_S)| + |E((\tau_{t+1} - \tau_0 - \tau m(t+1))\mathbf{1}_{\bar{S}})|\\
&\le |E((\tau_{t+1} - \tau_0 - \tau m(t+1))\mathbf{1}_S)| + 2m(t+1)\Pr(\bar{S})\\
&\le |E((\tau_{t+1} - \tau_0 - \tau m(t+1))\mathbf{1}_S)| + 2m(t+1)\frac{1}{t^2}.
\end{aligned}
$$

By Inequalities (6.5) and (6.6), we have

$$\begin{aligned}
&|E((\tau_{t+1} - \tau_0 - \tau m(t+1)))| \\
&\quad\le\ |E((\tau_t - \tau_0 - \tau mt)\mathbf{1}_S)| + \left(\frac{\sqrt{8p_3}}{(p_1-p_3)^2}\sqrt{\frac{\log t}{t}} + O\left(\frac{\log t}{t}\right)\right) + 2mt\frac{1}{t^2} \\
&\quad\le\ |E(\tau_t - \tau_0 - \tau mt)| + \frac{\sqrt{8p_3}}{(p_1-p_3)^2}\sqrt{\frac{\log t}{t}} + O\left(\frac{\log t}{t}\right) + 4m\frac{1}{t} \\
&\quad\le\ Cm\sqrt{t\log t} + \left(\frac{\sqrt{8p_3}}{(p_1-p_3)^2}\sqrt{\frac{\log t}{t}} + O\left(\frac{\log t}{t}\right)\right) \\
&\quad\le\ Cm\sqrt{(t+1)\log(t+1)}.
\end{aligned}$$

The proof of Lemma 6.5 is complete. □

To derive the concentration result on τ_t, we will need to bound $E(\tau_t)$ as the initial number of edge τ_0 changes.

Lemma 6.6. *We consider two random graphs G_t and G'_t in $G(p_1, p_2, p_3, p_4, m)$. Suppose that G_t initially has τ_0 edges and n_0 vertices, and G'_t initially have τ'_0 edges and n'_0 vertices. Let τ_t and τ'_t denote the number of edges in G_t and G'_t, respectively. If $n_0 - n'_0 = O(1)$, then we have*

$$|E(\tau_t) - E(\tau'_t)| \le |\tau_0 - \tau'_0| + O(\log t).$$

Proof. From Equation (6.3), we have

$$\begin{aligned}
E(\tau_{t+1}) &= E(\tau_t) + mp_1 + mp_2 - 2p_3 E\left(\frac{\tau_t}{n_t}\right) - mp_4, \\
E(\tau'_{t+1}) &= E(\tau'_t) + mp_1 + mp_2 - 2p_3 E\left(\frac{\tau'_t}{n'_t}\right) - mp_4.
\end{aligned}$$

Then,

$$E(\tau_{t+1} - \tau'_{t+1}) = E(\tau_t - \tau'_t) - 2p_3 E\left(\frac{\tau_t}{n_t} - \frac{\tau'_t}{n'_t}\right). \tag{6.7}$$

Since both $n_t - n_0$ and $n'_t - n'_0$ follow the same distribution, we have

$$Pr(n_t = x) = Pr(n'_t = x + n'_0 - n_0) \qquad \text{for any } x.$$

We can rewrite $E\left(\frac{\tau_t}{n_t} - \frac{\tau'_t}{n'_t}\right)$ as follows:

$$
\begin{aligned}
& E\left(\frac{\tau_t}{n_t} - \frac{\tau_t'}{n_t'}\right) \\
&= \sum_x \frac{1}{x} E(\tau_t | n_t = x) Pr(n_t = x) - \sum_y \frac{1}{y} E(\tau_t' | n_t' = y) Pr(n_t' = y) \\
&= \sum_x \frac{1}{x} E(\tau_t | n_t = x) Pr(n_t = x) \\
&\quad - \sum_x \frac{1}{x + n_0' - n_0} E(\tau_t' | n_t' = x + n_0' - n_0) Pr(n_t' = x + n_0' - n_0) \\
&= \sum_x Pr(n_t = x) \left(\frac{1}{x} E(\tau_t | n_t = x) - \frac{1}{x + n_0' - n_0} E(\tau_t' | n_t' = x + n_0' - n_0) \right) \\
&= \sum_x Pr(n_t = x) \left(\frac{1}{x} (E(\tau_t | n_t = x) - E(\tau_t' | n_t' = x + n_0' - n_0)) \right. \\
&\quad \left. - \left(\frac{1}{x} - \frac{1}{x + n_0' - n_0} \right) E(\tau_t' | n_t' = x + n_0' - n_0) \right).
\end{aligned}
$$

From Lemma 6.4, with probability at least $1 - \frac{2}{t^2}$, we have

$$|n_t - n_0 - (p_1 - p_3)t| \leq 2\sqrt{t \log t}.$$

Let S denote the set of x satisfying $|x - n_0 - (p_1 - p_3)t| \leq 2\sqrt{t \log t}$. The probability for x not in S is at most $\frac{2}{t^2}$. If this case happens, the contribution to $E(\frac{\tau_t}{n_t} - \frac{\tau_t'}{n_t'})$ is $O(\frac{1}{t})$, which is a minor term.

In addition, τ_t' is always upper bounded by $\tau_0' + mt$. We can bound the second term as follows.

$$
\begin{aligned}
& \left| \sum_x \left(\frac{1}{x} - \frac{1}{x + n_0' - n_0} \right) E(\tau_t' | n_t' = x + n_0' - n_0) Pr(n_t = x) \right| \\
&= \left| \sum_{x \in S} \left(\frac{1}{x} - \frac{1}{x + n_0' - n_0} \right) E(\tau_t' | n_t' = x + n_0' - n_0) Pr(n_t = x) \right| + O\left(\frac{1}{t}\right) \\
&\leq \frac{|n_0' - n_0| \sum_x (\tau_0' + mt) Pr(n_t = x)}{(n_0 + (p_1 - p_3)t - 2\sqrt{t \log t})(n_0' + (p_1 - p_3)t - 2\sqrt{t \log t})} + O\left(\frac{1}{t}\right) \\
&\leq \frac{|n_0' - n_0| (\tau_0' + mt)}{(n_0 + (p_1 - p_3)t - 2\sqrt{t \log t})(n_0' + (p_1 - p_3)t - 2\sqrt{t \log t})} + O\left(\frac{1}{t}\right) \\
&= O\left(\frac{1}{t}\right).
\end{aligned}
$$

Hence, we obtain

$$\begin{aligned}
E\left(\frac{\tau_t}{n_t} - \frac{\tau'_t}{n'_t}\right) \\
&= \sum_x Pr(n_t = x)\frac{1}{x}(E(\tau_t|n_t = x) - E(\tau'_t|n'_t = x + n'_0 - n_0)) + O\left(\frac{1}{t}\right) \\
&= \sum_{x\in S} Pr(n_t = x)\frac{1}{x}(E(\tau_t|n_t = x) - E(\tau'_t|n'_t = x + n'_0 - n_0)) + O\left(\frac{1}{t}\right) \\
&= \frac{1}{n_0 + (p_1 - p_3)t + O(\sqrt{t\log t})} \\
&\quad \times \sum_{x\in S} Pr(n_t = x)(E(\tau_t|n_t = x) - E(\tau'_t|n'_t = x + n'_0 - n_0)) + O\left(\frac{1}{t}\right) \\
&= \frac{1}{n_0 + (p_1 - p_3)t + O(\sqrt{t\log t})}\left(\sum_{x\in S} E(\tau_t|n_t = x)Pr(n_t = x)\right. \\
&\quad \left. - \sum_x E(\tau'_t|n'_t = x + n'_0 - n_0)Pr(n'_t = x + n'_0 - n_0)\right) + O\left(\frac{1}{t}\right) \\
&= \frac{1}{n_0 + (p_1 - p_3)t + O(\sqrt{t\log t})}(E(\tau_t) - E(\tau'_t)) + O\left(\frac{1}{t}\right).
\end{aligned}$$

Combine this with Equation (6.7), and we have

$$E(\tau_{t+1} - \tau'_{t+1}) = \left(1 - \frac{2p_3}{n_0 + (p_1 - p_3)t + O(\sqrt{t\log t})}\right)E(\tau_t - \tau'_t) + O\left(\frac{1}{t}\right).$$

Therefore, we have

$$\begin{aligned}
&|E(\tau_{t+1} - \tau'_{t+1})| \\
&\quad \le \left|\left(1 - \frac{2p_3}{n_0 + (p_1 - p_3)t + O(\sqrt{t\log t})}\right)E(\tau_t - \tau'_t)\right| + O\left(\frac{1}{t}\right) \\
&\quad \le |E(\tau_t - \tau'_t)| + O\left(\frac{1}{t}\right) \\
&\quad\quad \vdots \\
&\quad \le |\tau_0 - \tau'_0| + \sum_{i=1}^{t} O\left(\frac{1}{i}\right) \\
&\quad = |\tau_0 - \tau'_0| + O(\log t).
\end{aligned}$$

The proof of the lemma is complete. □

In order to prove the concentration result for the number of edges for $G(p_1, p_2, p_3, p_4, m)$, we shall use the general martingale inequality. To establish the near Lipschitz coefficients, we will derive upper bounds for the degrees by considering the special case without deletion. For $p_1 = \alpha$, $p_2 = 1 - \alpha$, and $p_3 = p_4 = 0$, $G(\alpha, 1-\alpha, 0, 0, m)$ is just the preferential attachment model. The number of edge increases by m at a time. The total number of edges at time t is exactly $mt + \tau_0$, where τ_0 is the number of edge of the initial graph at $t = 0$.

We label the vertex u by i if u is generated at time i. Let $d_i(t)$ denote the degree of the vertex i at time t.

Lemma 6.7. *For the preferential attachment model $G(\gamma, 1-\gamma, 0, 0, m)$, we have, with probability at least $1 - t^{-k}$ (any $k > 1$), the degree of vertex i at time t satisfies*

$$d_i(t) \le mk \left(\frac{t}{i}\right)^{1-\gamma/2} \log t.$$

Proof. For the preferential attachment model $G(\gamma, 1-\gamma, 0, 0, m)$, the total number of edge at time t is

$$\tau_t = mt + \tau_0.$$

The recurrence for the expected value of $d_i(t)$ satisfies

$$E(d_i(t+1)|d_i(t)) = d_i(t) + m\gamma\frac{d_i(t)}{2\tau_t} + m(1-\gamma)\frac{d_i(t)}{\tau_t} = \left(1 + \frac{m(2-\gamma)}{2\tau_t}\right) d_i(t).$$

We denote $\theta_t = 1 + \frac{m(2-\gamma)}{2\tau_t}$. Let X_t be the scaled version of $d_i(t)$ defined as follows:

$$X_t = \frac{d_i(t)}{\prod_{j=i+1}^{t-1} \theta_j}.$$

We have

$$E(X_{t+1}|X_t) = \frac{E(d_i(t+1)|d_i(t))}{\prod_{j=i+1}^{t} \theta_j} = \frac{\theta_t d_i(t)}{\prod_{j=i+1}^{t} \theta_j} = X_t.$$

Thus, X_t forms a martingale with $E(X_t) = X_{i+1} = d_i(i+1) = m$. We apply Theorem 5.5. First, we compute

$$\begin{aligned}
\mathrm{Var}(X_{t+1}|X_t) &= \frac{1}{\prod_{j=i+1}^{t}\theta_j^2}\mathrm{Var}(d_i(t+1)|d_i(t)) \\
&\leq \frac{1}{\prod_{j=i+1}^{t}\theta_j^2}E((d_i(t+1)-d_i(t))^2|d_i(t)) \\
&\leq \frac{1}{\prod_{j=i+1}^{t}\theta_j^2}m\left(\gamma\frac{d_i(t)}{2\tau_t}+(1-\gamma)\frac{d_i(t)}{\tau_t}\right) \\
&= \frac{1}{\prod_{j=i+1}^{t}\theta_j^2}(\theta_t-1)d_i(t) \\
&= \frac{\theta_t-1}{\theta_t\prod_{j=i+1}^{t}\theta_j}X_t.
\end{aligned}$$

Let $\phi_t = \frac{\theta_t-1}{\theta_t\prod_{j=i+1}^{t}\theta_j}$. We have

$$\begin{aligned}
\phi_t &= \frac{\theta_t-1}{\theta_t\prod_{j=i+1}^{t}\theta_j} \\
&= \frac{\frac{m(2-\gamma)}{2(mt+\tau_0)}}{\left(1+\frac{m(2-\gamma)}{2(mt+\tau_0)}\right)\prod_{j=i+1}^{t}\left(1+\frac{m(2-\gamma)}{2(mj+\tau_0)}\right)} \\
&\approx \left(1-\frac{\gamma}{2}\right)\frac{1}{t}e^{-(1-\frac{\gamma}{2})\sum_{j=i+1}^{t}\frac{1}{j+\frac{\tau_0}{m}}} \\
&\approx \left(1-\frac{\gamma}{2}\right)\frac{1}{t}\left(\frac{i}{t}\right)^{1-\gamma/2} \\
&\approx \left(1-\frac{\gamma}{2}\right)\frac{i^{1-\gamma/2}}{t^{2-\gamma/2}}.
\end{aligned}$$

In particular, we have

$$\begin{aligned}
\sum_{j=i+1}^{t-1}\phi_j &\approx \left(1-\frac{\gamma}{2}\right)\sum_{j=i+1}^{t-1}\frac{i^{1-\gamma/2}}{j^{2-\gamma/2}} \\
&\approx 1.
\end{aligned}$$

Concerning the last condition in Theorem 5.5, we have

$$\begin{aligned}
X_{t+1}-E(X_{t+1}|X_t) &= \frac{1}{\prod_{j=i+1}^{t}\theta_j}(d_i(t+1)-E(d_i(t+1)|d_i(t))) \\
&\leq \frac{1}{\prod_{j=i+1}^{t}\theta_j}(d_i(t+1)-d_i(t)) \\
&\leq \frac{1}{\prod_{j=i+1}^{t}\theta_j}m \\
&\leq m.
\end{aligned}$$

With $M = m$ and $\sum_{j=i+1}^{t-1} \phi_j \approx 1$, Theorem 5.5 gives

$$\Pr(X_t < m + a) \leq e^{-\frac{a^2}{2(m+a+ma/3)}}.$$

By choosing $a = m(k \log t - 1)$, with probability at least $1 - O(t^{-k})$, we have

$$X_t < m + m(k \log t - 1) = mk \log t.$$

Hence, with probability at least $1 - O(t^{-k})$, $d_i(t) \leq mk \log t (\frac{t}{i})^{1-\gamma/2}$. □

Remark 6.8. In the above proof, $d_i(t + 1) - d_i(t)$ roughly follows the Poisson distribution with mean

$$\frac{m(2-\gamma)d_i(t)}{2\tau_t} = O\left(i^{-(1-\gamma/2)} t^{-\gamma/2}\right) = O\left(t^{-\gamma/2}\right).$$

It follows with probability at least $1 - O(t^{-k})$ that $d_i(t + 1) - d_i(t)$ is bounded by $\frac{2k}{\gamma}$. Applying Theorem 5.6 with $M = \frac{2k}{\gamma}$ and $\sum_{j=i+1}^{t-1} \phi_j \approx 1$, we get

$$\Pr(X_t < m + a) \leq e^{-\frac{a^2}{2\left(m+a+\frac{2ka}{3\gamma}\right)}} + O(t^{-k}).$$

When $m \gg \log t$, we can choose $a = \sqrt{3mk \log t}$ so that m dominates $\frac{2ka}{3\gamma}$. In this case, we have

$$\Pr(X_t < m + a) \leq e^{-\frac{a^2}{2\left(m+a+\frac{2ka}{3\gamma}\right)}} + O(t^{-k}) = O(t^{-k}).$$

With probability at least $1 - O(t^{-k})$, we have $d_i(t) \leq (m + \sqrt{3mk \log t})(\frac{t}{i})^{1-\gamma/2}$.

Similarly arguments using Theorem 5.8 give the lower bound of the same order. If i survives at time t in the preferential attachment model $G(\gamma, 1 - \gamma, 0, 0, m)$, then, with probability at least $1 - O(t^{-k})$, we have

$$d_i(t) \geq \left(m - \sqrt{3mk \log t}\right) \left(\frac{t}{i}\right)^{1-\gamma/2}.$$

The above bounds will be further generalized for model $G(p_1, p_2, p_3, p_4, m)$ later in Lemma 6.11 with similar ideas.

Lemma 6.9. *For any k, i, and t in graph $G(p_1, p_2, p_3, p_4, m)$, the degree of i at time t satisfies*

$$d_i(t) \leq Ckm \log \left(\frac{t}{i}\right)^{1-\frac{p_1}{2(p_1+p_2)}} \tag{6.8}$$

with probability at least $1 - O(\frac{1}{t^k})$, for some absolute constant C.

Proof. We compare $G(p_1, p_2, p_3, p_4, m)$ with the following preferential attachment model $G(p_1, p_2, 0, 0, m)$ without deletion:

At each step,
- with probability p_1, take a vertex-growth step and add m edges from the new vertex to the current graph;
- with probability p_2, take an edge-growth step and m edges are added into the current graph;
- with probability $1 - p_1 - p_2$, do nothing.

We wish to show that the degree $d_u(t)$ in the model $G(p_1, p_2, p_3, p_4, m)$ (with deletion) is dominated by the degree sequence $d_u(t)$ in the model $G(p_1, p_2, 0, 0, m)$. Basically, it is a balls-and-bins argument, similar to the one given in [Cooper et al. 04]. The number of balls in the first bin (denoted by a_1) represents the degree of u while the number of balls in the other bin (denoted by a_2) represents the sum of degrees of the vertices other than u. When an edge incident to u is added to the graph $G(p_1, p_2, p_3, p_4, m)$, it increases both a_1 and a_2 by 1. When an edge not incident to u is added into the graph, a_2 increases by 2 while a_1 remains the same. Without loss of generality, we can assume that a_1 is less than a_2 in the initial graph. If an edge uv, which is incident to u, is deleted later, we delay adding this edge until the very moment that the edge is to be deleted. At the moment of adding the edge uv, the two bins have a_1 and a_2 balls, respectively. When we delay adding the edge uv, the number of balls in each bin is still a_1 and a_2, respectively, compared with $a_1 + 1$ and $a_2 + 1$ in the original random process. Since $a_1 < a_2$, the random process with delay dominates the original random process. If an edge vw, which is not incident to u, is deleted, we also delay adding this edge until the very moment that the edge is to be deleted. Equivalently, we compare the process with a_1 and a_2 balls in the bins to the process with a_1 and $a_2 + 2$ balls. The random process without delay dominates the one with delay. Therefore, for any u, the degrees of u in the model without deletion dominates the degrees in the model with deletion.

It remains to derive an appropriate upper bound of $d_u(t)$ for model

$$G(p_1, p_2, 0, 0, m).$$

If a vertex u is added at time i, we label it by i. Let us remove the idle steps and re-parameterize the time. For $\gamma = \frac{p_1}{p_1+p_2}$, we have $G(p_1, p_2, 0, 0, m) = G(\gamma, 1-\gamma, 0, 0, m)$. We can use the upper bound for the degrees of $G(\gamma, 1-\gamma, 0, 0, m)$ as in Lemma 6.7. This completes the proof for Lemma 6.9. $\square$

Lemma 6.5 can be further strengthened as follows:

Lemma 6.10. *In $G(p_1, p_2, p_3, p_4, m)$ with initial graph on n_0 vertices and τ_0 edges, the total number of edges at time t is*

$$\tau_t = \tau_0 + \tau m t + O\left(kmt^{1-\frac{p_1}{4(p_1+p_2)}} \log^{3/2} t\right)$$

with probability at least $1 - O(t^{-k})$ where $\tau = \frac{(p_1+p_2-p_4)(p_1-p_3)}{p_1+p_3}$.

Proof. For a fixed s with $s \le t$, we define $\tau_s(t) = \#\{ij \in E(G_t) | s \le i, j \le t\}$. We use Lemma 6.9 with the initial graph to be taken as the graph G_s at time s. Then, Lemma 6.9 implies that, with probability at least $1 - O(\frac{1}{t^{k-1}})$, we have

$$\begin{aligned} \tau_t - \tau_s(t) &\le \sum_{i \le s} d_i(t) \\ &\le \sum_{i \le s} C \left(\frac{t}{i}\right)^{1-\frac{p_1}{2(p_1+p_2)}} mk \log t \\ &\le \frac{C}{\frac{p_1}{2(p_1+p_2)}} mkt \log t \left(\frac{t}{s}\right)^{-\frac{p_1}{2(p_1+p_2)}}. \end{aligned}$$

By choosing $s = \sqrt{t}$, we have

$$\tau_t = \tau_s(t) + O\left(t^{1-\frac{p_1}{4(p_1+p_2)}} mk \log t\right). \tag{6.9}$$

We want to show that with probability at least $1 - O(t^{-k/2+1})$, we have

$$\begin{aligned} |\tau_t - E(\tau_t)| &\le |\tau_t - \tau_s(t)| + |E(\tau_t) - E(\tau_s(t))| + |\tau_s(t) - E(\tau_s(t))| \\ &\le O\left(mkt^{1-\frac{p_1}{4(p_1+p_4)}} \log^{3/2} t\right). \end{aligned}$$

It suffices to show that $\tau_s(t) - E(\tau_s(t)) = O(mkt^{1-\frac{p_1}{4(p_1+p_4)}} \log^{3/2} t)$.

We use the general martingale inequality as in Theorem 5.1 as follows: let $c_i = Ckm(\frac{i}{s})^{1-\frac{p_1}{2(p_1+p_2)}} \log t$ where C is the constant in Equation (6.8). The nodes of the decision tree T are just graphs generated by graph model $G(p_1, p_2, p_3, p_4, m)$. A path from the root to a leaf in the decision tree T is associated with a chain of the graph evolution. The value $f(i)$ at each node G_i (as defined in the proof of Theorem 5.1) is the expected number of edges at time t with initial graph G_i at time i. We note that X_i might be different from the number of edges of G_i, which is denoted by τ_i.

Let G_{i+1} be any child node of G_i in the decision T. We define $f(i+1)$ and τ_{i+1} in a similar way. By Lemma 6.6, we have

$$|f(i+1) - f(i)| \le |\tau_{i+1} - \tau_i| + O(\log t) \le (1 + o(1))c_i.$$

We say that an edge of the decision tree T is *bad* if and only if it deletes a vertex of degree greater than $(1+o(1))c_i$ at time i. A leaf of T is *good* if none of the graphs in the chain contains a vertex with degree larger than $(1+o(1))c_i$ at time i. Therefore, the probability for the set B consisting of bad leaves is at most

$$\Pr(B) \leq \sum_{l=s}^{t} O\left(l^{-k}\right) = O(s^{-k+1}).$$

By Theorem 5.1, we have

$$\begin{aligned}
\Pr(|\tau_s(t) - E(\tau_s(t))| > a) &\leq 2e^{-\frac{a^2}{\sum_{l=s}^t c_l^2}} + \Pr(B) \\
&\leq 2e^{-\frac{a^2}{\sum_{l=s}^t \left(\frac{l}{s}\right)^{2-\frac{p_1}{(p_1+p_2)}} (Cmk\log t)^2}} + O(s^{-k+1}) \\
&\leq 2e^{-\frac{a^2}{C'C^2 t^{3-\frac{p_1}{p_1+p_2}} s^{-2+\frac{p_1}{p_1+p_2}} m^2k^2\log^2 t}} + O(s^{-k+1}).
\end{aligned}$$

We choose $s = \sqrt{t}$ and $a = \sqrt{C'}Ct^{1-\frac{p_1}{4(p_1+p_2)}} mk\log^{3/2} t$. With probability at least $1 - O(t^{-k/2+1})$, we have

$$|\tau_s(t) - E(\tau_s(t))| = O\left(t^{1-\frac{p_1}{4(p_1+p_2)}} mk\log^{3/2} t\right),$$

as desired. □

Lemma 6.11. *For the model $G(p_1, p_2, p_3, p_4, m)$, let $\alpha = \frac{p_1(p_1+2p_2-p_3-2p_4)}{2(p_1+p_2-p_4)(p_1-p_3)}$ and $\gamma = \frac{p_1}{p_1+p_2}$. If $\log t \ll m \leq t^{\gamma/2}$, we have the following:*

1. *For $p_3 > 0$ and $\epsilon > 0$, with probability at least $1-\epsilon$, no vertex born before $\epsilon t^{\frac{p_3}{p_1}}$ survives at time t.*

2. *If the vertex i survives at time t, then, with probability at least $1 - O(t^{-k})$, the degree $d_i(t)$ in a graph G of the model $G(p_1, p_2, p_3, p_4)$ satisfies*

$$\begin{aligned}
d_i(t) &\geq (m - C\sqrt{mk\log t})(1 - Ci^{-\gamma/4}\log^{3/2} i)\left(\frac{t}{i}\right)^{\alpha}, \\
d_i(t) &\leq (m + C\sqrt{mk\log t})(1 + Ci^{-\gamma/4}\log^{3/2} i)\left(\frac{t}{i}\right)^{\alpha},
\end{aligned}$$

   for some constant C depending on p_1, p_2, p_3, and p_4.

Proof. For a fixed t and $i \leq t$, let Z_i denote the number of vertices left at time i with indices less than $t_0 = \epsilon t^{\frac{p_3}{p_1}}$ (i.e., born before t_0). Clearly, $Z_{t_0} \leq t_0$. For

$t_0 \le i \le t$, we have

$$Z_{i+1} = \begin{cases} Z_i - 1 & \text{with probability } p_3 \frac{Z_i}{n_i} \\ Z_i & \text{otherwise.} \end{cases} \tag{6.10}$$

We wish to upper bound the expected value of Z_{i+1}. From Inequality (6.2) we have

$$\Pr(n_i > (p_1 - p_3)i + O(\sqrt{2ki \log i})) \le i^{-k}.$$

We write

$$\begin{aligned}
&E(Z_{i+1}) \\
&= E\left(Z_{i+1}\mathbf{1}_{n_i \le (p_1-p_3)i + O\left(\sqrt{2ki\log i}\right)}\right) + E\left(Z_{i+1}\mathbf{1}_{n_i > (p_1-p_3)i + O\left(\sqrt{2ki\log i}\right)}\right) \\
&\le E\left(Z_{i+1}\mathbf{1}_{n_i \le (p_1-p_3)i + O\left(\sqrt{2ki\log i}\right)}\right) \\
&\quad + t_0 \Pr\left(n_i > (p_1 - p_3)i + O\left(\sqrt{2ki\log i}\right)\right) \\
&\le E(Z_i) - p_3 \frac{E(Z_i)}{(p_1 - p_3)i + O\left(\sqrt{2ki\log i}\right)} \\
&\quad + t_0 \Pr\left(n_i > (p_1 - p_3)i + O\left(\sqrt{2ki\log i}\right)\right) \\
&\le E(Z_i)\left(1 - \frac{p_3}{(p_1 - p_3)i + O(\sqrt{2ki\log i})}\right) + t_0 i^{-k}.
\end{aligned}$$

The above recursive formula of $E(Z_i)$ can be solved as follows. Let $a_i = E(Z_i) - t^{-1}$. If $k > 3$, we have

$$\begin{aligned}
a_{i+1} - \left(1 - \frac{p_3}{(p_1 - p_3)i + O(\sqrt{2ki\log i})}\right) a_i \\
\le t_0 i^{-k} - \tfrac{p_3}{(p_1-p_3)i + O(\sqrt{2ki\log i})} t^{-1} \le 0.
\end{aligned}$$

Since $a_{t_0} \le E(Z_{t_0}) \le t_0$, we get

$$\begin{aligned}
a_t &\le a_{t_0} \prod_{i=t_0}^{t-1} \left(1 - \frac{p_3}{(p_1 - p_3)i + O(\sqrt{2ki\log i})}\right) \\
&\le t_0 e^{-\sum_{i=t_0}^{t-1} \frac{p_3}{(p_1-p_3)i + O(\sqrt{2ki\log i})}} \\
&= (1 + o(1)) t_0 e^{-\frac{p_3}{p_1 - p_3}\ln(t/t_0)}.
\end{aligned}$$

We note that

$$\ln \frac{t}{t_0} = \left(1 - \frac{p_3}{p_1}\right) \ln t - \ln \epsilon.$$

Hence,

$$\begin{aligned} a_t &\le (1+o(1))t_0 e^{-\frac{p_3}{p_1-p_3}\ln(t/t_0)} \\ &= (1+o(1))\epsilon t^{\frac{p_3}{p_1}} e^{-\frac{p_3}{p_1}\ln t + \frac{p_3}{p_1-p_3}\ln \epsilon} \\ &= (1+o(1))\epsilon^{\frac{p_1}{p_1-p_3}}. \end{aligned}$$

Therefore, we have

$$\begin{aligned} \Pr(Z_t > 0) &\le a_t + \frac{1}{t_0} \\ &\le (1+o(1))\epsilon^{\frac{p_1}{p_1-p_3}} + \frac{1}{t_0} \\ &\le \epsilon. \end{aligned}$$

This implies that, with probability at least $1-\epsilon$, the number of vertices, which are born before t_0 and survive at time t, is zero. Item 1 is proved.

Let $\mathcal{F}_t$ be the σ-algebras generated by all subsets of the probability space at time t. Under the condition that vertex i survives at time t, we have

$$\begin{aligned} E(d_i(t+1)|\mathcal{F}_t) &\approx d_i(t) + p_1 m \frac{d_i(t)}{2\tau_t} + p_2 m \frac{d_i(t)}{\tau_t} - p_3 \frac{d_i(t)}{n_t} - p_4 m \frac{d_i(t)}{\tau_t} \\ &= \left(1 + m\frac{(p_1+2p_2-2p_4)}{2\tau_t} - \frac{p_3}{n_t}\right) d_i(t). \end{aligned} \tag{6.11}$$

To see this, with probability p_1, m edges from a new vertex will be added to the graph. For this case, the probability that the vertex i is selected as an endpoint of these m edges is $\frac{mX_t}{2\tau_t}$. The terms containing p_2 and p_4 are the contributions from the edge-addition step and the edge-deletion step, respectively. The term containing p_3 is the contribution from the vertex-deletion. We note that repetition in the edge-deletion step only causes an error of minor term in the above computation.

By Lemma 6.10, with probability at least $1 - O(t^{-k})$, the total number of edges is

$$\tau_t = \tau m t + O(kmt^{1-\gamma/4}\log^{3/2} t).$$

Recall that $\tau = \frac{(p_1+p_2-p_4)(p_1-p_3)}{p_1+p_3}$ and $\gamma = \frac{p_1}{p_1+p_2}$. By Lemma 6.4, the number n_t of vertices at time t satisfies

$$n_t = (p_1 - p_3)t + O(\sqrt{2kt\log t}),$$

with probability at least $1 - \frac{2}{t^k}$.

Substitute τ_t and n_t into the Recurrence Forumla (6.11) and simplify. Thus, with probability at least $1 - O(t^{-k})$, we have

$$\begin{aligned} E(d_i(t+1)|\mathcal{F}_t) &= \left(1 + m\frac{(p_1+2p_2-2p_4)}{2\tau_t} - \frac{p_3}{n_t}\right) d_i(t) \\ &= \left(1 + m\frac{(p_1+2p_2-2p_4)}{2(\tau mt + O(kmt^{1-\gamma/4}\log t))}\right. \\ &\quad \left. - \frac{p_3}{(p_1-p_3)t + O(\sqrt{2kt\log t})}\right) d_i(t) \\ &= \left(1 + \frac{\alpha}{t} + O(t^{-1-\gamma/4}\log t)\right) d_i(t). \end{aligned}$$

Let $\theta_t = 1 + \frac{\alpha}{t} + Ct^{-1-\gamma/4}$ for some large constant C. With probability at least $1 - O(t^{-k})$, we have

$$E(d_i(t+1)|\mathcal{F}_t) \le \theta_t d_i(t).$$

Now we apply Theorem 5.6 to random variables

$$X_t = \frac{1}{\prod_{j=i+1}^{t-1}\theta_j} d_i(t).$$

With probability at least $1 - O(t^{-k})$, we have

$$\begin{aligned} E(X_{t+1}|\mathcal{F}_t) &= E\left(\frac{1}{\prod_{j=i}^{t}\theta_j} d_i(t+1)|\mathcal{F}_t\right) \\ &\le \frac{1}{\prod_{j=i+1}^{t}\theta_j}\theta_t d_i(t) \\ &= X_t. \end{aligned}$$

In other words, X_t is nearly a submartingale. We compute

$$\begin{aligned} \prod_{j=i+1}^{t-1}\theta_j &= \prod_{j=i+1}^{t-1}\left(1 + \frac{\alpha}{j} + Cj^{-1-\gamma/4}\log^{3/2} j\right) \\ &\le e^{\sum_{j=i+1}^{t-1}\left(\frac{\alpha}{j} + Cj^{-1-\gamma/4}\log^{3/2} j\right)} \\ &= e^{\alpha(\log t - \log i) + O(i^{-\gamma/4}\log^{3/2} i)} \\ &= \left(1 + O\left(i^{-\gamma/4}\log^{3/2} i\right)\right)\left(\frac{t}{i}\right)^{\alpha}. \end{aligned}$$

Next, we consider the variance $\mathrm{Var}(X_{t+1}|\mathcal{F}_t)$:

$$\begin{aligned} \mathrm{Var}(X_{t+1}|\mathcal{F}_t) &= \frac{1}{\prod_{j=i+1}^{t}\theta_j^2}\mathrm{Var}(d_i(t+1)|\mathcal{F}_t) \\ &\le \frac{1}{\prod_{j=i+1}^{t}\theta_j^2} E((d_i(t+1) - d_i(t))^2|\mathcal{F}_t). \end{aligned}$$

The second moment $E((d_i(t+1)-d_i(t))^2|\mathcal{F}_t)$ consists of four items, which correspond to four steps: vertex-growth step, edge-growth step, vertex-deletion step, and edge-deletion step. Recall that the graphs are always simple. We have

$$\begin{aligned} E((d_i(t+1)-d_i(t))^2|\mathcal{F}_t) &\leq p_1 m\frac{d_i(t)}{2\tau_t} + p_2 m\frac{d_i(t)}{\tau_t} + p_3\frac{d_i}{n_t} + p_4 m\frac{d_i(t)}{\tau_t} \\ &= \left(\frac{p_1+2p_2+2p_4}{2\tau} + \frac{p_3}{p_1-p_3} + o(1)\right)\frac{1}{t}d_i(t). \end{aligned}$$

Let $\phi_t = \left(\frac{p_1+2p_2+2p_4}{2\tau} + \frac{p_3}{p_1-p_3} + o(1)\right)\frac{1}{t\theta_t\prod_{j=i+1}^{t}\theta_j}$. Then $\mathrm{Var}(X_{t+1}|\mathcal{F}_t) \leq \phi_t X_t$. We have

$$\begin{aligned} \sum_{j=i+1}^{t-1}\phi_j &= O\left(\sum_{j=i+1}^{t-1}\frac{1}{j\theta_j\prod_{l=i+1}^{t}\theta_l}\right) \\ &= O\left(\sum_{j=i+1}^{t-1}\frac{i^\alpha}{j^{1+\alpha}}\right) \\ &= O(1). \end{aligned}$$

Let us estimate $|d_i(t+1)-d_i(t)|$. It is upper bounded by 1 if it takes a vertex-growth step or a vertex-deletion step (with i surviving). It is at most m if it takes an edge-growth step or an edge-deletion step. We can further lower the upper bound by considering trade-off with probability.

For an edge-growth step, it follows the Poisson distribution with mean

$$\begin{aligned} \mu &= m\frac{d_i(t)}{2\tau_t} \\ &\leq m\frac{Cmk\log t(\frac{t}{i})^{1-\gamma}}{2m\tau t + o(t)} \\ &= O(mkt^{-\gamma}\log t) \\ &\leq o(t^{-\gamma/3}). \end{aligned}$$

By using Lemma 6.9 and $m < t^{\gamma/2}$, with probability at least $1-O(t^{-k})$, $d_i(t)-d_i(t+1)$ is bounded by $2k < \frac{6k}{\gamma}$.

For an edge-deletion step, it follows the Poisson distribution with mean

$$\mu = \frac{m}{\tau_t} \approx \frac{1}{\tau t}.$$

With probability at least $1-O(t^{-k})$, $d_i(t)-d_i(t+1)$ is bounded by $2k < \frac{6k}{\gamma}$.

Therefore,

$$\begin{aligned} |E(X_{t+1}|\mathcal{F}_t) - X_{t+1}| &\leq \frac{|E(d_i(t+1)|\mathcal{F}_t) - d_i(t+1)|}{\prod_{j=i+1}^{t}\theta_j} \\ &\leq 2|d_i(t+1) - d_i(t)| \\ &\leq \frac{12k}{\gamma}. \end{aligned}$$

By applying Theorem 5.6 with $M = \frac{12k}{\gamma}$ and $\sum_{j=i+1}^{t-1}\phi_j = O(1)$, we have

$$\Pr(X_t < m + a) \leq e^{-\frac{a^2}{2\left(C(m+a)+\frac{12ka}{3\gamma}\right)}} + O(t^{-k}).$$

When $m \gg \log t$, we can choose $a = \sqrt{3Cmk\log t}$ so that m dominates $\frac{12ka}{3\gamma}$. In this case, we have

$$\Pr(X_t < m + a) \leq e^{-\frac{a^2}{2\left(C(m+a)+\frac{2ka}{3\gamma}\right)}} + O(t^{-k}) = O(t^{-k}).$$

With probability at least $1 - O(t^{-k})$, we have

$$d_i(t) \leq (m + \sqrt{3Cmk\log t})\left(1 + O\left(i^{-\gamma/4}\log^{3/2} i\right)\left(\frac{t}{i}\right)^{\alpha}\right).$$

The proof of the lower bound is similar by using Theorem 5.8 instead. Let $\theta'_t = 1 + \frac{\alpha}{t} - Ct^{-1-\delta}$ for some large constant C. With probability at least $1 - O(\frac{1}{t^k})$, we have

$$E(d_i(t+1)|\mathcal{F}_t) \geq \theta'_t d_i(t).$$

Now, we apply Theorem 5.8 to random variables

$$Y_t = \frac{1}{\prod_{j=i+1}^{t-1}\theta'_j} d_i(t).$$

With probability at least $1 - O(t^{-k})$, we have

$$\begin{aligned} E(Y_{t+1}|\mathcal{F}_t) &= E\left(\frac{1}{\prod_{j=i}^{t}\theta'_j} d_i(t+1)|\mathcal{F}_t\right) \\ &= \frac{1}{\prod_{j=i+1}^{t}\theta'_j} E(d_i(t+1)|\mathcal{F}_t) \\ &\geq \frac{1}{\prod_{j=i+1}^{t}\theta'_j}\theta'_t d_i(t) \\ &= Y_t. \end{aligned}$$

Hence, X_t is nearly a supermartingale. We have

$$\begin{aligned}\prod_{j=i+1}^{t-1} \theta'_j &= \prod_{j=i+1}^{t-1} \left(1 + \frac{\alpha}{j} - Cj^{-1-\gamma/4} \log^{3/2} j\right) \\ &= e^{\sum_{j=i+1}^{t-1} \left(\frac{\alpha}{j} - O\left(j^{-1-\gamma/4} \log^{3/2} j\right)\right)} \\ &= e^{\alpha(\log t - \log i) - O(i^{-\gamma/4} \log^{3/2} i)} \\ &= \left(1 - O\left(i^{-\gamma/4} \log^{3/2} i\right)\right) \left(\frac{t}{i}\right)^{\alpha}.\end{aligned}$$

Similarly, let $\phi'_t = \frac{\left(\frac{p_1+2p_2+2p_4}{2\tau} + \frac{p_3}{p_1-p_3} + o(1)\right)}{t\theta'_t \prod_{j=i+1}^{t} \theta'_j}$. Then, $\mathrm{Var}(Y_{t+1}|\mathcal{F}_t) \leq \phi'_t Y_t$,

$$\begin{aligned}\sum_{j=i+1}^{t-1} \phi'_j &= O\left(\sum_{j=i+1}^{t-1} \frac{1}{j\theta'_j \prod_{l=i+1}^{t} \theta'_l}\right) \\ &= O\left(\sum_{j=i+1}^{t-1} \frac{i^{\alpha}}{j^{1+\alpha}}\right) \\ &= O(1),\end{aligned}$$

and

$$\begin{aligned}|E(Y_{t+1}|\mathcal{F}_t) - Y_{t+1}| &= \frac{|E(d_i(t+1)|\mathcal{F}_t) - d_i(t+1)|}{\prod_{j=i+1}^{t} \theta_j} \\ &\leq 2|d_i(t+1) - d_i(t)| \\ &\leq \frac{12k}{\gamma}.\end{aligned}$$

Using Theorem 5.8 with $M = \frac{12k}{\gamma}$ and $\sum_{j=i+1}^{t-1} \phi'_j = O(1)$, we have

$$\Pr(X_t < m - a) \leq e^{-\frac{a^2}{2\left(Cm + \frac{12ka}{3\gamma}\right)}} + O(t^{-k}).$$

When $m \gg \log t$, we can choose $a = \sqrt{3Cmk \log t}$ so that m dominates $\frac{12ka}{3\gamma}$. In this case, we have

$$\Pr(X_t < m - a) \leq e^{-\frac{a^2}{2\left(Cm + \frac{2ka}{3\gamma}\right)}} + O(t^{-k}) = O(t^{-k}).$$

With probability at least $1 - O(t^{-k})$, we have

$$d_i(t) \geq (m - \sqrt{3Cmk \log t}) \left(1 - O\left(i^{-\gamma/4} \log^{3/2} i\right) \left(\frac{t}{i}\right)^{\alpha}\right).$$

The proof of Lemma 6.11 is complete. $\square$

7. The Proofs for the Main Theorems

Now we are ready to prove Theorems 6.1 to 6.3.

Proof of Theorem 6.1. The probability that a vertex i survives at time t is

$$\prod_{l=i+1}^{t}\left(1-\frac{p_3}{n_l}\right) \approx e^{\sum_{l=i+1}^{t} -\frac{p_3}{(p_1-p_3)t}} \approx \left(\frac{i}{t}\right)^{\frac{p_3}{p_1-p_3}}.$$

Suppose that i survives at time t. By Lemma 6.11, with high probability, we have

$$d_i(t) = (1+o(1))m\left(\frac{t}{i}\right)^{\alpha}.$$

Recall that $\alpha = \frac{p_1(p_1+2p_2-p_3-2p_4)}{2(p_1+p_2-p_4)(p_1-p_3)}$. The number of vertices with degree between x_1 and x_2 can be written by

$$\begin{aligned}
&\sum_{(1+o(1))(\frac{x_2}{m})^{-1/\alpha}t\le i\le(1+o(1))(\frac{x_1}{m})^{-1/\alpha}t} \left(\frac{i}{t}\right)^{\frac{p_3}{p_1-p_3}} \\
&\approx \left(\left(\frac{x_1}{m}\right)^{\frac{-p_1}{\alpha(p_1-p_3)}} - \left(\frac{x_2}{m}\right)^{\frac{-p_1}{\alpha(p_1-p_3)}}\right)\frac{p_1-p_3}{p_1} \\
&\approx \frac{p_1-p_3}{p_1}\left(\left(\frac{x_1}{m}\right)^{-\beta+1} - \left(\frac{x_2}{m}\right)^{-\beta+1}\right).
\end{aligned}$$

We note that

$$\frac{-p_1}{\alpha(p_1-p_3)} = -\frac{2(p_1+p_2-p_4)}{p_1+2p_2-p_3-2p_4} = -\beta+1.$$

The number of vertices with degree between x and $x+\Delta x$ is

$$\left(1-\frac{p_3}{p_1}+o(1)\right)\left(\left(\frac{x}{m}\right)^{-\beta+1} - \left(\frac{x+\Delta x}{m}\right)^{-\beta+1}\right) \approx \left(1-\frac{p_3}{p_1}\right)\frac{\beta m^{\beta-1}}{x^{\beta}}\Delta x.$$

Hence, $G(p_1,p_2,p_3,p_4,m)$ is a power law graph with exponent

$$\beta = 2 + \frac{p_1+p_3}{p_1+2p_2-p_3-2p_4}.$$

This completes the proof for Item 1.

For $t_0 = \lfloor \frac{1}{g(t)} t^{\frac{p_3}{p_1}} \rfloor$, where $g(t)$ is an arbitrarily slow growing function, Lemma 6.11 implies that almost surely any surviving vertices are born after t_0. To prove Item 2, for some fixed $l \leq t$, we consider $w_i^{(l)} = m(\frac{l}{i})^\alpha$ and $\tau_l = m\tau l$, for $t \geq l \geq i \geq t_0$ where $\tau = \frac{(p_1+p_2-p_4)(p_1-p_3)}{p_1+p_3} l$.

For $l = t_0, \ldots, t$, let $G^l(p_1, p_2, p_3, p_4, m)$ denote the graph at time l generated by the evolution model $G(p_1, p_2, p_3, p_4, m)$. Now, we construct an edge-independent random graph H^l as follows. At $l = t_0$, H^l is an empty graph. By the induction assumption, we assume that the edge-independent random graph H^j has been constructed, for $j \leq l$.

If at time $l+1$ we have a vertex-growth step in $G^{l+1}(p_1, p_2, p_3, p_4, m)$, we add a new vertex labeled by $l+1$ to H_l. Let F_v^l be the edge-independent random graph with

$$p_{i,l+1} = (1 - o(1))m\frac{w_i^{(l)}}{2\tau_l}.$$

We define $H^{l+1} = H^l \cup F_v^l$.

If at time $l+1$ we have an edge-growth step in $G^{l+1}(p_1, p_2, p_3, p_4, m)$, let F_e^l be the edge-independent random graph with

$$p_{i,j} = m(1 - o(1))\frac{w_i^{(l)} w_j^{(l)}}{4\tau_l^2},$$

for all pairs of vertices (i, j) in H^l. We define $H^{l+1} = H^l \cup F_e^l$.

If at time $l+1$ we have a vertex-deletion step in $G^{l+1}(p_1, p_2, p_3, p_4, m)$, we delete the same vertex from H^l and call the resulted graph H^{l+1}.

If at time $l+1$ we have an edge-deletion step in $G^{l+1}(p_1, p_2, p_3, p_4, m)$, let F_d^l be the random graph with uniform probability $p = \frac{m}{\tau_l}$. We define $H^{l+1} = H^l \setminus F_d^l$.

Clearly, H^{l+1} is edge-independent if H^l is edge-independent.

From the above construction, for any two vertices i and j $(i < j)$ in H^l, the edge probability $p_{ij}^{(l)}$ satisfies the following recurrence formula:

$$p_{ij}^{(l+1)} = \begin{cases} m(1-o(1))\frac{w_i^{(j)}}{2\tau_j} & \text{if } l = j-1 \\ p_{ij}^l & \text{with probability } p_1 + p_3 \\ p_{ij}^l + \left(1 - p_{ij}^{(l)}\right) m(1-o(1))\frac{w_i^{(l)} w_j^{(l)}}{4\tau_l^2} & \text{with probability } p_2 \\ p_{ij}^l \left(1 - \frac{m}{2\tau_l}\right) & \text{with probability } p_4 \\ 0 & \text{if } i \text{ and } j \text{ are deleted or } l < j \end{cases}$$

Let $a_l = (1 - o(1))p_2 m\frac{w_i^{(l)} w^{(l)} y_j}{4\tau_l^2} = (1 - o(1))p_2 m\frac{1}{4\tau^2 i^\alpha j^\alpha} l^{2\alpha-2}$ and $b_l = (1 + o(1))p_4\frac{m}{2\tau_l} = (1 + o(1))\frac{p_4}{2\tau}\frac{1}{l}$. The expected value $E(p_{ij}^{(l)})$ satisfies the following recurrence formula:

$$E\left(p_{ij}^{(l+1)}\right) = (1 - a_l - b_l)E\left(p_{ij}^{(l)}\right) + a_l. \tag{7.1}$$

This implies that

$$\begin{aligned}
E\left(p_{ij}^{(t)}\right) &= (1-a_{t-1}-b_{t-1})E\left(p_{ij}^{(t-1)}\right)+a_{t-1}\\
&= (1-a_{t-1}-b_{t-1})\left((1-a_{t-2}-b_{t-2})E\left(p_{ij}^{(t-2)}\right)+a_{t-2}\right)+a_{t-1}\\
&\vdots\\
&= \sum_{s=j+1}^{t-1} a_s \prod_{l=s}^{t-1}(1-a_l-b_l)+E\left(p_{ij}^{(j)}\right)\prod_{l=j}^{t-1}(1-a_l-b_l).
\end{aligned}$$

Before we proceed to prove that $p_{ij}^{(t)}$ concentrates on $E(p_{ij}^{(t)})$, we simplify the expression for $E(p_{ij}^{(t)})$ by solving the recurrence in Equation (7.1). We consider the following two cases.

Case 1. $a_t = o(b_t)$.

For any $l \le t$, we have

$$\frac{a_l}{a_t} \approx \left(\frac{l}{t}\right)^{2\alpha-2} \le \left(\frac{l}{t}\right)^{-1} \approx \frac{b_l}{b_t}.$$

Hence, $a_l \le \frac{a_t}{b_t} b_l = o(b_l)$.

Suppose that the Recurrence Formula (7.1) has a solution in the following form: $p_{ij}^{(l)} \approx Cl^x$ for all $l \le t$. By substituting into the formula, we have

$$C(l+1)^x \approx Cl^x(1-a_l-b_l)+a_l.$$

Here we apply the estimation

$$(l+1)^x \approx l^x\left(1+\frac{x}{l}\right).$$

We have

$$Cl^x\left(\frac{x}{l}+a_l+b_l\right) \approx a_l$$

and

$$Cl^x \approx \frac{a_l}{\frac{x}{l}+a_l+b_l} \approx \frac{\frac{p_2 m}{4\tau^2 i^\alpha j^\alpha}}{x+\frac{p_4}{2\tau}} l^{2\alpha-1}.$$

By choosing $x = 2\alpha-1$, we have

$$C = \frac{\frac{p_2 m}{4\tau^2 i^\alpha j^\alpha}}{x+\frac{p_4}{2\tau}} = \frac{p_2 m}{2\tau(2p_2-p_4)i^\alpha j^\alpha}.$$

Let $f(l)$ be the difference $E(p_{ij}^{(l)}) - \frac{p_2 m}{2\tau(2p_2-p_4)} \frac{l^{2\alpha-1}}{i^\alpha j^\alpha}$. It is enough to establish an appropriate upper bound for $f(l)$. Since both $E(p_{ij}^{(l)})$ and $\frac{p_2 m}{2\tau(2p_2-p_4)}$ are (asymptotic) solutions of Equation (7.1), we have $f(l+1) \approx (1 - a_l - b_l) f(l)$. Hence,

$$\begin{aligned} f(t) &\approx f(j) \prod_{l=j}^{t-1} (1 - a_l - b_l) \\ &\approx \left(m \frac{w_i^{(j)}}{2\tau_j} - \frac{p_2 m}{2\tau(2p_2 - p_4)} \frac{j^{\alpha-1}}{i^\alpha} \right) e^{-\sum_{l=j}^{t}(a_l + b_l)} \\ &\approx \left(1 - \frac{p_2}{(2p_2 - p_4)} \right) \frac{m}{2\tau} \frac{j^{\alpha-1}}{i^\alpha} e^{-\sum_{l=j}^{t}(a_l + b_l)} \\ &\approx \left(1 - \frac{p_2}{(2p_2 - p_4)} \right) \frac{m}{2\tau} \frac{j^{\alpha-1}}{i^\alpha} \left(\frac{j}{t} \right)^{\frac{p_4}{2\tau}} \\ &= \frac{p_2 m}{2\tau(2p_2 - p_4) i^\alpha j^\alpha} \frac{t^{2\alpha-1}}{i^\alpha j^\alpha} \left(1 - \frac{p_4}{p_2} \right) \left(\frac{j}{t} \right)^{\frac{p_4}{2\tau} + 2\alpha - 1}. \end{aligned}$$

The solution of the Recurrence Formula (7.1) is

$$p_{ij}^{(l)} \approx \frac{p_2 m}{2p_4 \tau (2p_2 - p_4)} \frac{l^{2\alpha-1}}{i^\alpha j^\alpha} \left(1 + \left(1 - \frac{p_4}{p_2} \right) \left(\frac{j}{t} \right)^{\frac{p_4}{2\tau} + 2\alpha - 1} \right).$$

If $t \gg j$, the above solution can be expressed as

$$p_{ij}^{(t)} = (1 + o(1)) \frac{p_2 m}{2p_4 \tau (2p_2 - p_4)} \frac{t^{2\alpha-1}}{i^\alpha j^\alpha}.$$

Case 2. $b_t = o(a_t)$.

From the definition of a_l and b_l, there is a $t_1 \le \frac{t}{2}$ satisfying $b_l = o(a_l)$ for all $t_1 \le l \le t$. We can rewrite the Recurrence Formula (7.1) as

$$1 - E\left(p_{ij}^{(l+1)}\right) = (1 - a_l - b_l)\left(1 - E\left(p_{ij}^{(l)}\right)\right) + b_l. \tag{7.2}$$

Suppose that the Recurrence Formula (7.1) has a solution with the following form: $p_{ij}^{(l)} \approx 1 - C' l^y$ for all $l \le t$. We have

$$C'(l+1)^y \approx C' l^y (1 - a_l - b_l) + b_l.$$

In a similar way as in Case 1, we have

$$C' l^y \approx \frac{b_l}{\frac{x}{l} + a_l + b_l} \approx \frac{b_l}{a_l} \approx \frac{2p_4 \tau}{p_2 m} i^\alpha j^\alpha l^{1-2\alpha}.$$

We choose $y = 1 - 2\alpha$ and $C' = 2\frac{p_4}{p_2}\tau m i^\alpha j^\alpha$. Consider

$$f(l) = E\left(p_{ij}^{(l)}\right) - \left(1 - 2\frac{p_4}{p_2}\tau m i^\alpha j^\alpha l^{1-2\alpha}\right).$$

From Equation (7.1), we have $f(l+1) \approx (1 - a_l - b_l) f(l)$. Hence,

$$\begin{aligned} |f(t)| &\approx |f(t_1)| \prod_{l=t_1}^{t-1} (1 - a_l - b_l) \\ &\leq e^{-\sum_{l=t_1}^{t}(a_l + b_l)} \\ &= o(\tau m i^\alpha j^\alpha l^{1-2\alpha}). \end{aligned}$$

Hence, the solution of the Recurrence Formula (7.1) is

$$E(p_{ij}^{(l)}) = 1 - (1 + o(1))\frac{2p_4\tau}{p_2 m} i^\alpha j^\alpha l^{1-2\alpha}.$$

It is sufficient to prove that $p_{ij}^{(t)}$ concentrates on its expected value. Consider a martingale $X_l = E(p_{ij}^{(t)} | p_{ij}^{(j)}, \ldots, p_{ij}^{(l)})$, for $l = j, \ldots, t$. Since $p_{ij}^{(l+1)}$ only depends on $p_{ij}^{(l)}$ but not on the history $p_{ij}^{(s)}$ for $s < l$, this implies that $X_l = E(p_{ij}^{(t)} | p_{ij}^{(l)})$ is the expected value at time t with initial value $p_{ij}^{(l)}$ at time l. The solution of the following recurrence formula is $X_l = f(t)$:

$$f(s+1) = (1 - a_s - b_s) f(s) + a_s, \qquad \text{with } f(l) = p_{ij}^{(l)}.$$

We have

$$\begin{aligned} |X_{l+1} - X_l| &= |p_{ij}^{(l+1)} - p_{ij}^{(l)}| \prod_{s=l+1} (1 - a_s - b_s) \\ &\leq |p_{ij}^{(l+1)} - p_{ij}^{(l)}| \\ &\leq \max\{a_l(1 - p_{ij}^{(l)}), b_l p_{ij}^{(l)}\}. \end{aligned}$$

Let $c_i = \max\{a_l(1 - p_{ij}^{(l)}), b_l p_{ij}^{(l)}\}$ denote the sequence c for the c-Lipschitz condition.

For the case $a_t = o(b_t)$, we first get a crude upper bound for $p_{ij}^{(l)}$ (by setting $p_2 = 1$):

$$\begin{aligned}
p_{ij}^{(l)} &\leq 1-(1-p_{ij}^{(j)})\left(1-\prod_{s=j}^{l}\left(1-m\frac{w_i^{(s)}w_j^{(s)}}{4\tau_s^2}\right)\right)\\
&\leq \frac{j^{\alpha-1}}{2\tau i^{\alpha}}+\sum_{s=j}^{l} m\frac{w_i^{(s)}w_j^{(s)}}{4\tau_s^2}\\
&\leq (1+O(1))\frac{m(2\alpha-1)}{4\tau^2}\frac{l^{2\alpha-1}}{i^{\alpha}j^{\alpha}}.
\end{aligned}$$

Also, we have

$$b_l p_{ij}^{(l)} \leq b_l(1+O(1))\frac{m(2\alpha-1)}{4\tau^2}\frac{l^{2\alpha-1}}{i^{\alpha}j^{\alpha}} = \Theta(a_l).$$

Hence,

$$\begin{aligned}
\sum_{l=j}^{t} c_l^2 &\leq \Theta\left(\sum_{l=j}^{t} a_l^2\right)\\
&\approx \begin{cases} O\left(\frac{m^2 t^{4\alpha-3}}{(ij)^{2\alpha}}\right) & \text{if } \alpha>\frac{3}{4}\\ O\left(\frac{m^2 j^{4\alpha-3}}{(ij)^{2\alpha}}\right) & \text{if } \alpha<\frac{3}{4}\end{cases}\\
&= \begin{cases} \Theta(t^{-1}(E(p_{ij}^{(t)})^2) & \text{if } \alpha>\frac{3}{4}\\ \Theta(t^{-(2\alpha-2)}E(p_{ij}^{(t)})^2) & \text{if } \alpha<\frac{3}{4}\end{cases}\\
&= o(E(p_{ij}^{(t)})^2).
\end{aligned}$$

By Azuma's martingale inequality, almost surely, we have

$$p_{ij}^{(t)} = E(p_{ij}^{(t)}) + O\left(\sqrt{\sum_{l=j}^{t} c_l^2}\right) = (1+o(1))E(p_{ij}^{(t)}).$$

For the case $b_t = \Theta(a_t)$ and $E(p_{ij}^{(t)}) = \Theta(1)$, we have

$$\sum_{l=j}^{t} c_l^2 = O\left(\sum_{l=j}^{t} b_l^2\right) = \Theta\left(\frac{1}{j}\right) = o(1)$$

and, therefore,

$$p_{ij}^{(t)} = E(p_{ij}^{(t)}) + O\left(\sqrt{\sum_{l=j}^{t} c_l^2}\right) = (1+o(1))E\left(p_{ij}^{(t)}\right).$$

Now we prove inductively that $G^l(p_1, p_2, p_3, p_4, m)$ dominates G^l within an error estimate $o(t^{-K})$ (for some constant $K > 2$).

For $l = t_0$, the statement is trivial since H^l is an empty graph. We now assume that $G^l(p_1, p_2, p_3, p_4, m)$ dominates H^l within error estimate $o(t^{-K})$.

If at time $l + 1$ we have a vertex-growth step, we define the random graph $\phi(H)$ to be the graph resulting from adding to H^l m random edges from the new vertex. The other endpoints of these edges are chosen with probability proportional to their degrees in H. We note that $G^{l+1}(p_1, p_2, p_3, p_4) = \phi(G^l(p_1, p_2, p_3, p_4))$. Since $G^l(p_1, p_2, p_3, p_4, m)$ dominates H^l within an error estimate $o(t^{-K})$, $G^{l+1}(p_1, p_2, p_3, p_4)$ dominates $\phi(H^l)$ within an appropriate error term.

If at time $l + 1$ we have an edge-growth step, we define the random graph $\phi(H)$ to be the graph resulting from adding m random edges on the vertices of H. The endpoints of those edges are chosen with probability proportional to their degrees in H. We note that $G^{l+1}(p_1, p_2, p_3, p_4) = \phi(G^l(p_1, p_2, p_3, p_4))$. Since $G^l(p_1, p_2, p_3, p_4, m)$ dominates H^l within error estimate $o(t^{-K})$, $G^{l+1}(p_1, p_2, p_3, p_4)$ dominates $\phi(G^l)$, with a suitable error term.

If at time $l+1$ we have a vertex-deletion step, it is clear that $G^{l+1}(p_1, p_2, p_3, p_4)$ dominates H^{l+1} within the same error estimate as at time l.

If at time $l+1$ we have an edge-deletion step, we note that $G^{l+1}(p_1, p_2, p_3, p_4) = G^l(p_1, p_2, p_3, p_4) \setminus H_d^l$. Since $G^l(p_1, p_2, p_3, p_4, m)$ dominates H^l with an error estimate $o(t^{-K})$, $G^{l+1}(p_1, p_2, p_3, p_4)$ dominates $H^l \setminus F_e^l = G^{l+1}$.

The total error bound is less that t times the maximum error within each step. Hence, the error is $o(t^{-K})$ for any constant K. The proof of one direction for the domination is completed. The proof of the other direction can be treated similarly except that the opposite direction of the domination is involved and we omit that proof here. $\square$

Proof of Theorem 6.2. When $j > i \gg m^{\frac{1}{\alpha}} t^{1-\frac{1}{2\alpha}}$, we have

$$p_{ij} \approx \frac{p_2 m}{2p_4 \tau (2p_2 - p_4)} \frac{t^{2\alpha-1}}{i^\alpha j^\alpha}.$$

Let H_1 be the edge-independent random graph with $p'_{ij} = \frac{p_2 m}{2p_4 \tau (2p_2 - p_4)} \frac{t^{2\alpha-1}}{i^\alpha j^\alpha}$. Since p_{ij} can be written as a product of a function of i and a function of j, H_1 is a random graph with given expected degrees. To calculate the expected degrees of H_1, we will use the fact that the probability that i survives at t is

$(1+o(1))\left(\frac{i}{t}\right)^{\frac{p_3}{p_1-p_3}}$. Hence, the expected degree of H_i is

$$\begin{aligned} d_i &= \sum_{j\in S} p'_{ij}\left(\frac{j}{t}\right)^{\frac{p_3}{p_1-p_3}} \\ &\approx \sum j = m^{\frac{1}{\alpha}}t^{1-\frac{1}{2\alpha}}\,{}^{t} \frac{p_2 m}{2p_4\tau(2p_2-p_4)}\frac{t^{2\alpha-1}}{i^\alpha j^\alpha}\left(\frac{j}{t}\right)^{\frac{p_3}{p_1-p_3}} \\ &\approx \frac{p_2 m}{2p_4\tau(2p_2-p_4)(\frac{p_1}{p_1-p_3}-\alpha)}\frac{t^\alpha}{i^\alpha}, \end{aligned}$$

as claimed.

For the other direction, we note that

$$\begin{aligned} p_{ij} &\approx \frac{p_2 m}{2p_4\tau(2p_2-p_4)}\frac{t^{2\alpha-1}}{i^\alpha j^\alpha}\left(1+\left(1-\frac{p_4}{p_2}\right)\left(\frac{j}{t}\right)^{\frac{p_4}{2\tau}+2\alpha-1}\right) \\ &\leq \frac{p_2 m}{2p_4\tau(2p_2-p_4)}\frac{t^{2\alpha-1}}{i^\alpha j^\alpha}\left(1+\left(1-\frac{p_4}{p_2}\right)\right) \\ &= \frac{m}{2p_4\tau}\frac{t^{2\alpha-1}}{i^\alpha j^\alpha}. \end{aligned}$$

Let H_2 be the edge-independent random graph with $p''_{ij} = \frac{m}{2p_4\tau}\frac{t^{2\alpha-1}}{i^\alpha j^\alpha}$. Clearly, H_2 is a random graph with a given expected degree sequence. The proof is similar and will be omitted. □

Proof of Theorem 6.3. When $i < j \ll m^{\frac{1}{\alpha}}t^{1-\frac{1}{2\alpha}}$, we have

$$p_{ij} = 1-(1+o(1))\frac{2p_4\tau}{p_2 m}i^\alpha j^\alpha t^{1-2\alpha} = 1-o(1).$$

Therefore, it is dominating and dominated by the complete subgraph. □

Acknowledgments. We are grateful to the anonymous referees whose valuable comments have led to numerous improvements of the paper. Our research was supported in part by NSF Grants DMS 0100472 and ITR 0205061.

References

[Aiello et al. 00] W. Aiello, F. Chung, and L. Lu. “A Random Graph Model for Massive Graphs.” In *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing*, pp. 171–180. New York: ACM Press, 2000.

[Aiello et al. 01] W. Aiello, F. Chung, and L. Lu. "A Random Graph Model for Power Law Graphs." *Experimental Math.* 10:1 (2001), 53–66.

[Aiello et al. 02] W. Aiello, F. Chung, and L. Lu. "Random Evolution in Massive Graphs." In *Handbook of Massive Data Sets*, Vol. 2, edited by J. Abello et al., pp. 97–122. Dordrecht: Kluwer Academic Publishers, 2002.

[Barabási and Albert 99] Albert-László Barabási and Réka Albert. "Emergence of Scaling in Random Networks." *Science* 286 (1999), 509–512.

[Barabási et al. 00] A. Barabási, R. Albert, and H. Jeong. "Scale-Free Characteristics of Random Networks: The Topology of the World Wide Web." *Physica A* 281 (2000), 69–77.

[Broder et al. 00] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tompkins, and J. Wiener. "Graph Structure in the Web." In *Proceedings of the 9th International World Wide Web Conference on Computer Networks*, pp. 309–320. Amsterdam: North-Holland Publishing Co., 2000.

[Bollabás and Riordan 03] B. Bollabás and O. Riordan. "Robustness and Vulnerability of Scale-Free Random Graphs." *Internet Mathematics* 1:1 (2003), 1–35.

[Chung 97] F. Chung. *Spectral Graph Theory.* Providence, RI: American Mathematical Society, 1997.

[Chung and Lu 01] Fan Chung and Linyuan Lu. "The Diameter of Random Sparse Graphs." *Advances in Applied Math.* 26 (2001), 257–279.

[Chung and Lu 02a] Fan Chung and Linyuan Lu. "Connected Components in a Random Graph with Given Degree Sequences." *Annals of Combinatorics* 6 (2002), 125–145.

[Chung and Lu 02b] Fan Chung and Linyuan Lu. "The Average Distance in Random Graphs with Given Expected Degrees." *Proceeding of National Academy of Science* 99 (2002), 15879–15882.

[Chung et al. 03] Fan Chung, Linyuan Lu, and Van Vu. "The Spectra of Random Graphs with Given Expected Degrees." *Proceedings of National Academy of Sciences* 100:11 (2003), 6313–6318.

[Cooper and Frieze 03] Colin Cooper and Alan Frieze. "A General Model of Undirected Web Graphs." *Random Structures and Algorithms* 22 (2003), 311–335.

[Cooper et al. 04] Colin Cooper, Alan Frieze, and Juan Vera. "Random Vertex Deletion in a Scale Free Random Graph." *Internet Mathematics* 1:4 (2004), 463–483.

[Erdős and Gallai 61] P. Erdős and T. Gallai. "Gráfok előírt fokú pontokkal (Graphs with Points of Prescribed Degrees)" (in Hungarian). *Mat. Lapok* 11 (1961), 264–274.

[Erdős and Rényi 59] P. Erdős and A. Rényi. "On Random Graphs. I." *Publ. Math. Debrecen* 6 (1959), 290–291.

[Janson et al. 00] S. Janson, T. Łuczak, and A. Rucinski. *Random Graphs.* New York: Wiley-Interscience, 2000.

[Jeong et al. 00] H. Jeong, B. Tomber, R. Albert, Z. Oltvai, and A. L. Babárasi. "The Large-Scale Organization of Metabolic Networks." *Nature* 407 (2000), 378–382.

[Kleinberg et al. 99] J. Kleinberg, S. R. Kumar, P. Raphavan, S. Rajagopalan, and A. Tomkins. "The Web as a Graph: Measurements, Models and Methods." *Computing and Combinatorics: 5th Annual International Conference, COCOON '99, Tokyo, Japan, July 26–28, 1999, Proceedings*, edited by Takao Asano et al., Lecture Notes in Computer Science 1627, pp. 1–17. New York: Springer, 1999.

[Kim and Vu 00] J.H. Kim and V. Vu. "Concentration of Multi-Variate Polynomials and its Applications." *Combinatorica* 20:3 (2000), 417–434.

[Lotka 26] A. J. Lotka. "The Frequency Distribution of Scientific Productivity." *The Journal of the Washington Academy of the Sciences* 16 (1926), 317.

[Lu 01] Linyuan Lu. "The Diameter of Random Massive Graphs." In *Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 912–921. Philadelphia, PA: SIAM, 2001.

[McDiarmid 98] Colin McDiarmid. "Concentration." In *Probabilistic Methods for Algorithmic Discrete Mathematics*, edited by M. Habib et al., Algorithms and Combinatorics 16, pp. 195–248, Berlin: Springer, 1998.

[Milgram 67] S. Milgram. "The Small World Problem." *Psychology Today* 2 (1967), 60–67.

[Mitzenmacher 05] Michael Mitzenmacher. "A Brief History of Generative Models for Power Law and Lognormal Distributions." To appear in *Internet Mathematics* (2005).

Fan Chung, Department of Mathematics, University of California, San Diego, 9500 Gilman Drive, 0112, La Jolla, CA 92093-0112 (fan@ucsd.edu)

Linyuan Lu, Department of Mathematics, University of South Carolina, Columbia, SC 29208 (llu@math.sc.edu)

Received November 3, 2003; accepted September 3, 2004.