Limiting properties on the number of leaves in Barabasi-Albert Model

Linjun Zhang, Jinshuo Dong

Abstract

In this essay, we gave a short proof for SLLN result on the number of leaves in Barabasi-Albert random graph model, which was firstly established in [3]. Our proof is based on martingale model and Hoeffding inequality, and therefore avoids the lengthy proof in [3] by calculating moments. Moreover, we use martingale CLT to give heuristics to get CLT result in this model.

1 Introduction

The introduction of the Barabasi-Albert model (BA model) for generating random graphs was a significant step in understanding the properties of scale-free networks[1, 2]. In 1999, Barabasi and Albert proposed a method to generate random graph: start at time $n = 1$ with a graph consisting of two vertices that are connected by a single edge. At each step $n = 2, 3, \ldots$, we add a new vertex to the graph and connect it to an existing vertex, and the existing vertex is chosen randomly with probability proportional to its own degree.

In this essay, we analyze the limiting properties of the number of leaves (the nodes with degree 1) of BA model at time $n$, denoted as $X_n$. These results were firstly established in [3], but the authors gave a lengthy and algebra-heavy proof. In this essay, we use Azuma-Hoeffding inequality to present a simple proof for SLLN, and use Martingale CLT to give heuristics to obtain similar CLT result in BA model.

2 Model setup

In this section we give two ideas to interpret the BA model as a (sub)martingale.

2.1 Doob’s decomposition

Let $\mathcal{F}_n$ denote all the information about the graph $G_n$ at time $n$, with $n + 1$ vertices. If we define $Y_k$ as 0 when the $k^{th}$ incoming vertex connect itself with an existing leaf (whose node is of degree 1) and 1 else, then $Y_k|\mathcal{F}_{k-1} \sim Ber\left(1 - \frac{X_{k-1}}{2k}\right)$, and the number of leaves of $G_n$ is $X_n = 2 + \sum_{k=1}^{n} Y_k$.

From the above formulation, $X_n$ is a submartingale, therefore according to Doob’s decomposition,

$$X_n = 2 + \sum_{k=1}^{n} Y_k = \sum_{k=1}^{n} (Y_k - E[Y_k|\mathcal{F}_{k-1}]) + 2 + \sum_{k=1}^{n} E[Y_k|\mathcal{F}_{k-1}].$$

$^1$Ber denotes the Bernoulli distribution.
Let $M_n = \sum_{k=1}^n (Y_k - E[Y_k | \mathcal{F}_{k-1}])$, $A_n = 2 + \sum_{k=1}^n E[Y_k | \mathcal{F}_{k-1}]$. $M_n$ is a martingale with filtration $\mathcal{F}_n$, and $A_n$ is an a.s. increasing predictable process.

2.2 Vertex exposure martingale

Another approach to interpret $X_n$ is analog to Doob martingale. For a given time $n$, we define a sequence of random variables $V_1, \ldots, V_n$, where $V_i$ encodes the edges between vertex $i$ and vertices $1, \ldots, i - 1$. For any graph $G_n$, the sequence $V_1(G_n), \ldots, V_n(G_n)$ uniquely determines $G_n$, so there is a function $f$ such that the number of leaves in $G_n$ is $X_n = f(V_1, \ldots, V_n)$.

Let $X_n^{(k)} = E[X_n | V_1, \ldots, V_k]$ for $k = 1, \ldots, n$. Since $E[X_n^{(k)} | V_1, \ldots, V_{k-1}] = E[E[X_n | V_1, \ldots, V_k] | V_1, \ldots, V_{k-1}] = E[X_n | V_1, \ldots, V_{k-1}] = X_n^{(k-1)}$, we have $\{X_n^{(k)}\}_{k=1}^n$ is a martingale.

3 Main results

**Theorem 3.1.** If $X_n$ is the number of leaves of BA model at time $n$, then as $n \to \infty$,

$$\frac{X_n}{n} \xrightarrow{a.s.} \frac{2}{3}.$$

**Theorem 3.2.** If $X_n$ is the number of leaves of BA model at time $n$, then as $n \to \infty$,

$$\frac{X_n + \sum_{k=1}^n X_{k-1} - n}{\sqrt{2n}} \overset{L}{\to} N(0, 1).$$

Note: Theorem 2.2 is not a beautiful result, so we call it heuristics to obtain the beautiful result

$$\frac{X_n - \frac{2n}{3}}{\sqrt{\frac{2n}{9}}} \overset{L}{\to} N(0, 1),$$

which was proved in [3], but with a lengthy proof.

4 Technical proofs

Both proofs are much shorter than [3], and at first we introduce a commonly used inequality.

**Lemma 4.1.** (Azuma-Hoeffding inequality) Suppose $\{M_k : k = 0, 1, 2, 3, \ldots\}$ is a martingale and $|M_k - M_{k-1}| < c_k$, almost surely. Then for all positive integers $n$ and all positive reals $t$,

$$P(|X_n - X_0| \geq t) \leq \exp \left( -\frac{t^2}{2 \sum_{k=1}^n c_k^2} \right).$$

**Corollary 4.1.** Suppose $\{X_k = E[f(Z_1, \ldots, Z_n) | Z_1, \ldots, Z_k] : k = 0, 1, 2, 3, \ldots, n\}$ is a martingale with $f$ is $c$-Lipschitz, and $Z_i$ is independent of $Z_{i+1}, \ldots, Z_n$ conditioned on $Z_1, \ldots, Z_{i-1}$. Then for all positive integers $n$ and all positive reals $t$,

$$P(|X_n - X_0| \geq t) \leq \exp \left( -\frac{t^2}{2nc^2} \right).$$
Note: We say $f(Z_1, ..., Z_n)$ is $c$-Lipschitz if changing the value of any one coordinate of $f$ causes $f$ to change by at most $\pm c$.

**Proof of Theorem 2.1.**

**Step 1.** $\lim_{n \to \infty} \frac{E[X_n]}{n} = \frac{2}{3}$

Use our notation in section 1.1,

$$E[X_n] = E[E[X_n|F_{n-1}]] = E[X_{n-1}] + E[|Y_n|F_{n-1}] = E[X_{n-1}] + 1 - \frac{X_{n-1}}{2n}$$

$$\Rightarrow E[X_n] = \frac{2n-1}{2n} E[X_{n-1}] + 1$$

$$\Rightarrow E[X_n] - \frac{2}{3} (n+1) = \frac{2n-1}{2n} (E[X_{n-1}] - \frac{2}{3})$$

$$\Rightarrow \frac{E[X_n]}{n+1} - \frac{2}{3} = \left| \left( 1 - \frac{3}{2n+2} \right) \left( \frac{2n-1}{n} - \frac{2}{3} \right) \right| \leq |e^{-\frac{1}{n+2}}| \cdot \left| \frac{E[X_{n-1}]}{n} - \frac{2}{3} \right|$$

$$\Rightarrow \lim_{n \to \infty} \frac{E[X_n]}{n} = \frac{2}{3}$$

**Step 2.** $X_n \overset{a.s.}{\to} \frac{2}{3}$

Use our notation in section 1.2, and let $X_{n(i)} = E[X_n|V_1, ..., V_k] = E[f(V_1, ..., V_n)|V_1, ..., V_k]$. We observe that the function $f$ is $1$-Lipschitz: if we modify $V_i$ by changing the edge that vertex $i$ added, this change the number of leaves of $G_n$ by at most one. In addition, we have $X_n^{(0)} = E[X_n]$, and $X_n^{(n)} = X_n$. Applying Azuma-Hoeffding’s inequality to it,

$$P[|X_n - E[X_n]| \geq \lambda] \leq 2e^{-\lambda^2 / \lambda^2} \Rightarrow P[|\frac{X_n}{n} - \frac{E[X_n]}{n}| \geq \lambda] \leq 2e^{-\lambda^2 / \lambda^2}$$

Let $\lambda = \frac{1}{\sqrt{\lambda}}$, we have $P[|\frac{X_n}{n} - \frac{E[X_n]}{n}| \geq n^{1/4}] \leq 2e^{-\lambda^2 / \lambda^2}$. According to Borel Cantelli lemma and equation (1).

$$\lim_{n \to \infty} \frac{X_n}{n} \overset{a.s.}{=} \frac{2}{3}$$

**Proof of Theorem 2.2.**

We use the CLT for martingales derived in class to handle this. Let’s first review this.

**Lemma 4.2.** (Martingale CLT) Suppose $\{w_{n,j} : 1 \leq j \leq n\}$ is martingale difference sequence, such that (a).

$$\sum_{j=1}^n w_{n,j}^2 \overset{P}{\to} 1; \ (b). \max_j |w_{n,j}| \overset{P}{\to} 0; \ (c). E[\max_j w_{n,j}^2] \text{ is uniformly bounded, then}$$

$$\sum_{j=1}^n w_{n,j} \overset{L}{\to} N(0,1).$$
For our own martingale defined in section 1.1, we notice that
\[
\frac{X_n + \sum_{k=1}^{n} \frac{X_{k+1} - n}{2k} - n}{\sqrt{\frac{2n}{9}}} = \frac{\sum_{k=1}^{n} Y_k - E[Y_k | \mathcal{F}_{k-1}]}{\sqrt{\frac{2n}{9}}}
\]
with \(Y_k - E[Y_k | \mathcal{F}_{k-1}]\) being martingale difference sequence, and we are going to apply martingale CLT to it.

Since \(\left| \frac{Y_k - E[Y_k | \mathcal{F}_{k-1}]}{\sqrt{\frac{2n}{9}}} \right| \leq \frac{2}{\sqrt{\frac{2n}{9}}}\), condition (b) and (c) of Lemma 3.2 are satisfied in this case.

As for condition (a), we have
\[
\sum_{k=1}^{n} \left( \frac{Y_k - E[Y_k | \mathcal{F}_{k-1}]}{\sqrt{\frac{2n}{9}}} \right)^2 = \frac{9}{2n} \sum_{k=1}^{n} \left( Y_k - E[Y_k | \mathcal{F}_{k-1}] \right)^2 = \frac{9}{2n} \sum_{k=1}^{n} \left( Y_k^2 + E[Y_k | \mathcal{F}_{k-1}]^2 - 2Y_k E[Y_k | \mathcal{F}_{k-1}] \right)
\]
\[
= \frac{9}{2n} \sum_{k=1}^{n} Y_k + \frac{9}{2n} \sum_{k=1}^{n} E[Y_k | \mathcal{F}_{k-1}]^2 - 2 \sum_{k=1}^{n} Y_k E[Y_k | \mathcal{F}_{k-1}]
\]
\[
= \frac{9}{2n} X_n + \frac{9}{2n} \sum_{k=1}^{n} E[Y_k | \mathcal{F}_{k-1}]^2 - 2 \sum_{k=1}^{n} Y_k E[Y_k | \mathcal{F}_{k-1}].
\]

Since \(X_n \nabla \frac{a+1}{2}, E[Y_k | \mathcal{F}_{k-1}] = 1 - \frac{X_{k+1}}{2k} \nabla \frac{a}{2}, \) and using the fact \(\frac{\sum_{k=1}^{n} a_k}{n} \to a, \) \(b_n \to b\) implies \(\frac{\sum_{k=1}^{n} a_k b_n}{n} \to ab,\)
we have \(\frac{\sum_{k=1}^{n} E[Y_k | \mathcal{F}_{k-1}]^2}{n} \to \frac{4}{9}, \) \(\frac{\sum_{k=1}^{n} Y_k E[Y_k | \mathcal{F}_{k-1}]}{n} \to \frac{4}{9}\) a.s. Therefore \(\sum_{k=1}^{n} \left( \frac{Y_k - E[Y_k | \mathcal{F}_{k-1}]}{\sqrt{\frac{2n}{9}}} \right)^2 \to 1.\)

By Lemma 3.2, we complete our proof:
\[
\frac{X_n + \sum_{k=1}^{n} \frac{X_{k+1} - n}{2k} - n}{\sqrt{\frac{2n}{9}}} \to N(0,1).
\]

\[\square\]

References

