UDC 519.1,519.2

The number of vertices of fixed degree in the preferential attachment model with choice

Y. Malyshkin*

* *Tver State University,*
Sadovij per. 35, Tver, 170002, Russia

Abstract. The preferential attachment models are widely used to describe different web and social network (random) graphs. We concentrate on some generalizations of these models. Namely, a random tree under consideration is constructed in the following way. Let $d(1)$, $d(2)$,... be i.i.d. integer-valued random variables. At each step n a new vertex is introduced. Then we select $d(n)$ vertices, chosen from the old vertices with probabilities proportional to their degrees and conditionally independent given $d(n)$, and connect new vertex to the vertex from the sample with largest degree. In preferential attachment without choice $d(n) = 1$ for all n. We establish the upper class law of the iterated logarithm for the number of vertices of degree k at step n. The proof employs results of stochastic approximation theory along with analysis of specified martingales and system of the equations describing an evolution of the model.

Keywords: random trees, stochastic approximation, preferential attachment.

1. Introduction

In the present work, we study how the addition of choice affects the Mori's preferential attachment model. Let describe the max-choice Mori's preferential attachment tree model. This model is a time-indexed inductively constructed sequence of trees, built in the following way. First, fix number $\beta > -1$ and distribution of a random variable d with values in $\mathbb{N}$. These are the parameters of our model. Consider set of vertices $V = \{v_i\}_{i=1}^{\infty}$. Define a sequence of random trees $\{T_n\}$, $n \in \mathbb{N}$, by the following inductive rule. Let T_1 be the one-edge tree which consists of vertices v_1 and v_2 and an edge between them. Given T_n, we construct T_{n+1} by adding one vertex and drawing one edge in the following way. First, we add a vertex v_{n+2} to T_n. So, for the set $V(T_{n+1})$ of vertices of T_{n+1} we get $V(T_{n+1}) = \{v_i,\ i = 1, ..., n+2\}$. Note that $\sum_{v_i \in V(T_n)} \deg_{T_n} v_i = 2n$, where $\deg_{T_n} v_i$ is the degree of v_i in T_n. Second, we draw an edge between v_{n+2} and $Y_n \in V(T_n)$, which we choose by the rule describe below. So, for the set $E(T_{n+1})$ of edges of T_{n+1} we get $E(T_{n+1}) = E(T_n) \bigcup \{v_{n+2}, Y_n\}$. The randomness of T_{n+1} given T_n is due to randomness of Y_n. Order the set $V(T_n)$ by the vertices degrees in T_n. In other words, $V(T_n) = \{v^n_{(1)}, ..., v^n_{(n+1)}\}$, $\deg_{T_n} v^n_{(i)} \le \deg_{T_n} v^n_{(i+1)}$. Let

The presented work was performed under the State Assignment N3.8032.2017/BCh of the Ministry of Science and Education of Russia.

$\vartheta_j(n)$ denote the position of vertex v_j in ordered set $\{v^n_{(1)},...,v^n_{(n+1)}\}$, i.e. $v_j = v^n_{(\vartheta_j(n))}$ and if $\deg_{T_n} v_i = \deg_{T_n} v_j$ and $i < j$, then $\vartheta_i(n) < \vartheta_j(n)$. Consider i.i.d. random variables $\{U^i_n\}_{n\in\mathbb{N},i\in\mathbb{N}}$, distributed uniformly on $[0,1]$. Define random variables $X^i_n, i \in \mathbb{N}$ with values in $V(T_n)$ as follow. Let $X^i_n = v^n_{(1)}$ if $U^i_n \le \frac{\deg_{T_n} v^n_{(1)}+\beta}{(2+\beta)n+\beta}$ and $X^i_n = v_{(j)}$, $1 < j \le n+1$, if $\sum_{l=1}^{j-1} \frac{\deg_{T_n} v^n_{(l)}+\beta}{(2+\beta)n+\beta} < U^i_n \le \sum_{l=1}^{j} \frac{\deg_{T_n} v^n_{(l)}+\beta}{(2+\beta)n+\beta}$. Let $d_1, d_2, ...$ be i.i.d. random variable distributed as d. Finally, we take Y_n as the vertex among $X^1_n,...,X^{d_n}_n$ with the largest degree. In the case of a tie, choose the vertex with the largest index.

Remark 1 *Since T_{n+1} is well defined by $Y_1, ..., Y_n$, all its parameters are $\mathcal{F}_{n+1}$ measurable.*

Remark 2 *Prove of the main result would require analysis of event $A_n = \{(U^n_1, ..., U^n_{d_n}) \in D_{d_n}\}$ for random sets $D_i \in \mathbb{R}^i$, $i \in \mathbb{N}$. Note that if $|D_i|$ (here $|B|$ stands for the Lebesgue measure of B) does not depend on $\mathcal{F}_n$, then A_n does not depend on $\mathcal{F}_n$.*

Let formulate our theorem. Let $N_k(n)$ be the number of vertices of degree k in tree T_n, $Z_k(n) = \frac{N_k(n)}{n}$ and $W_k(n) = (Z_1(n), ..., Z_k(n))$. Theorem 5.1 of [1] states that there is a point $\rho^*_k = (x^*_1, ..., x^*_k)$ and a positively defined symmetric matrix $B_k = (b_{i,j})_{1\le i,j\le k}$ (both depends on β and distribution of d) such that $n^{1/2}(W_k(n) - \rho^*_k)$ converge in distribution to normal distribution $N(0, B)$ as $n \to \infty$. We will prove an upper class law of the iterated logarithm for variables $N_k(n)$.

Theorem 3 *Let $\mathbb{E}d^2 < \infty$. Then, for any $k \in \mathbb{N}$ one has*

$$\limsup_{n\to\infty} \left| \frac{N_k(n) - nx^*_k}{\sqrt{2b_{k,k} n \ln\ln n}} \right| \le 1 \quad a.s.$$

2. Proof of the main result

For $x_1, ..., x_k \in \mathbb{R}_+$, $k \in \mathbb{N}$, define functions

$$h_0 = 1,\ h_1(x_1) = \sum_{m=1}^{\infty} \mathbb{P}(d = m) \left(x_1 \frac{1+\beta}{2+\beta}\right)^m, \quad h_k(x_1, ..., x_k) =$$

$$= \sum_{m=1}^{\infty} \mathbb{P}(d = m) \left(\left(\sum_{j=1}^{k} x_j \frac{(j+\beta)}{2+\beta} \right)^m - \left(\sum_{j=1}^{k-1} x_j \frac{(j+\beta)}{2+\beta} \right)^m \right),$$

$$f_k(x_1, ..., x_k) = h_{k-1}(x_1, ..., x_{k-1}) - h_k(x_1, ..., x_k),$$

$$g_k(x_1,...,x_k) = f(x_1,...,x_k) - x_k.$$

As it is shown in [1],

$$b_{1,1} = 1 - \sum_{m=1}^{\infty} \mathbb{P}(d=m)\left(x_1^*\frac{1+\beta}{2+\beta}\right)^m,$$

$$\mathbb{P}(N_1(n+1) - N_1(n) = 1|\mathcal{F}_n) = \sum_{m=1}^{\infty} \mathbb{P}(d=m)\left(1 - \left(\frac{N_1(n)(1+\beta)}{(2+\beta)n+\beta}\right)^m\right)$$

and

$$b_{k,k} = h_k(x_1^*,...,x_k^*) + h_{k-1}(x_1^*,...,x_{k-1}^*),$$

$$g_k(x_1^*,...,x_k^*) = 0, \tag{1}$$

$$\mathbb{P}\left(N_k(n+1) - N_k(n) = 1|\mathcal{F}_n\right) = h_{k-1}\left(\widetilde{Z}_1(n),...,\widetilde{Z}_k(n)\right), \tag{2}$$

$$\mathbb{P}\left(N_k(n+1) - N_k(n) = -1|\mathcal{F}_n\right) = h_k\left(\widetilde{Z}_1(n),...,\widetilde{Z}_k(n)\right) \tag{3}$$

for $k > 1$, where

$$\widetilde{Z}_i(n) = \frac{N_i(n)}{n + \frac{\beta}{2+\beta}} = Z_i(n)\frac{1}{1 + \frac{\beta}{n(2+\beta)}}, \quad i = 1,...,k.$$

Also there are random $\mathcal{F}_n$ - measurable sets $D_n^+(k,m), D_n^-(k,m) \subset [0,1]^m$, $m,k,n \in \mathbb{N}$, such that

$$\{N_k(n+1) - N_k(n) = 1\} = \{(U_n^1,...,U_n^{d_n}) \in D_n^+(k,d_n)\},$$

$$\{N_k(n+1) - N_k(n) = -1\} = \{(U_n^1,...,U_n^{d_n}) \in D_n^-(k,d_n)\}.$$

Therefore,

$$N_k(n+1) - N_k(n) = \mathbf{1}\{(U_n^1,...,U_n^{d_n}) \in D_n^+(k,d_n)\}$$

$$-\mathbf{1}\{(U_n^1,...,U_n^{d_n}) \in D_n^-(k,d_n)\}.$$

Hence, from (2), (3) and definition of h_k, we have for $k > 1$

$$|D_n^-(k,m)| = \left(\left(\sum_{j=1}^{k}\widetilde{Z}_j(n)\frac{(j+\beta)}{2+\beta}\right)^m - \left(\sum_{j=1}^{k-1}\widetilde{Z}_j(n)\frac{(j+\beta)}{2+\beta}\right)^m\right),$$

$$|D_n^+(k,m)| = \left(\left(\sum_{j=1}^{k-1}\widetilde{Z}_j(n)\frac{(j+\beta)}{2+\beta}\right)^m - \left(\sum_{j=1}^{k-2}\widetilde{Z}_j(n)\frac{(j+\beta)}{2+\beta}\right)^m\right)$$

and for $k = 1$ we get $|D_n^-(1,m)| = 0$, $|D_n^+(1,m)| = 1 - \left(\frac{N_1(n)(1+\beta)}{(2+\beta)n+\beta}\right)^m$. Let $\alpha_{1,m}^- = 0$, $\alpha_{1,m}^+ = \left(1 - \left(x_1^* \frac{(1+\beta)}{2+\beta}\right)^m\right)$ and for $k > 1$

$$\alpha_{k,m}^- = \left(\left(\sum_{j=1}^{k} x_j^* \frac{(j+\beta)}{2+\beta}\right)^m - \left(\sum_{j=1}^{k-1} x_j^* \frac{(j+\beta)}{2+\beta}\right)^m\right),$$

$$\alpha_{k,m}^+ = \left(\left(\sum_{j=1}^{k-1} x_j^* \frac{(j+\beta)}{2+\beta}\right)^m - \left(\sum_{j=1}^{k-2} x_j^* \frac{(j+\beta)}{2+\beta}\right)^m\right).$$

Introduce random $\mathcal{F}_n$-measurable sets $D_*^+(k,m,n)$ and $D_*^-(k,m,n)$ such that $|D_*^+(k,m,n)| = \alpha_{k,m}^+$, $|D_*^-(k,m,n)| = \alpha_{k,m}^-$. Moreover if $|D_n^+(k,m)| > \alpha_{k,m}^+$ then $D_*^+(k,m,n) \subset D_n^+(k,m)$, if $|D_n^+(k,m)| \leq \alpha_{k,m}^+$, then $D_n^+(k,m) \subset D_*^+(k,m,n)$, if $|D_n^-(k,m)| > \alpha_{k,m}^-$, then $D_*^-(k,m,n) \subset D_n^-(k,m)$ and if $|D_n^-(k,m)| \leq \alpha_{k,m}^-$, then $D_n^-(k,m) \subset D_*^-(k,m,n)$. Let

$$X_k(n) = \mathbf{1}\{(U_n^1, ..., U_n^{d_n}) \in D_*^+(k, d_n, n)\}$$
$$-\mathbf{1}\{(U_n^1, ..., U_n^{d_n}) \in D_*^-(k, d_n, n)\} - x_k^*.$$

Note that due to Remark 2 and formula (1) we have that $X_k(n)$, $n \in \mathbb{N}$ are i.i.d. random variables and

$$\mathbb{E}X_k(n) = g_k(x_1^*, ..., x_k^*) = 0.$$

Also $\mathbb{E}X_k(n)^2 = b_{k,k}$. Hence, by the law of the iterated logarithm for i.i.d. random variables

$$\limsup_{n\to\infty} \left| \frac{\sum_{i=1}^n X_k(i)}{\sqrt{2b_{k,k} n \ln\ln n}} \right| = 1 \quad a.s.$$

Introduce random variables

$$\epsilon_k(n) = \mathbf{1}\{(U_n^1, ..., U_n^{d_n}) \in D_n^+(k, d_n) \backslash D_*^+(k, d_n, n)\}$$
$$-\mathbf{1}\{(U_n^1, ..., U_n^{d_n}) \in D_*^+(k, d_n, n) \backslash D_n^+(k, d_n)\}$$
$$-\mathbf{1}\{(U_n^1, ..., U_n^{d_n}) \in D_n^-(k, d_n) \backslash D_*^-(k, d_n, n)\}$$
$$+\mathbf{1}\{(U_n^1, ..., U_n^{d_n}) \in D_*^-(k, d_n, n) \backslash D_n^-(k, d_n)\}.$$

Therefore,

$$N_k(n+1) - N_k(n) - x_k^* = X_k(n) + \epsilon_k(n).$$

To complete the proof we need to estimate $\sum_{i=1}^{n}\epsilon_k(i)$. We give the sketch of the proof. Define

$$S_i(n)=N_i(n+1)-N_i(n)-x_i^*,\ Y_k(n)=(S_1(n),...,S_k(n)),$$

$$F_k(x_1,...,x_k)=(f_1(x_1),...,f_k(x_1,...,x_k)),$$

$$A_k(n)=(\epsilon_1(n),...,\epsilon_k(n)),\ \widetilde{W}_k(n)=(\widetilde{Z}_1(n),...,\widetilde{Z}_k(n)).$$

We get

$$\mathbb{E}(Y_k(n)|\mathcal{F}_n)=F_k(\widetilde{Z}_1(n),...,\widetilde{Z}_k(n))-F_k(x_1^*,...,x_k^*).$$

Consequently,

$$\mathbb{E}(A_k(n)|\mathcal{F}_n)=\mathbb{E}(Y_k(n)|\mathcal{F}_n)=\bigtriangledown F_k\left(\widetilde{W}_k(n)-\rho_k^*\right)+O(||\widetilde{W}_k(n)-\rho_k^*||^2)$$

$$=\bigtriangledown F_k\left(\frac{1}{n}\left((N_1(0),...,N_k(0))+\sum_{i=1}^{n}Y_k(i)\right)\right)+O(||\widetilde{W}_k(n)-\rho_k^*||^2).$$

Since all eigenvalues of $\bigtriangledown F_k$ are negative, $A_k(n)$ could be decomposed as $A_k(n)=A_k^1(n)+A_k^2(n)$ (with $\epsilon_i(n)=\epsilon_i^1(n)+\epsilon_i^2(n)$, $i=1,...,k$) where a.s. $A_k^1(n)$ is of opposite sign to $\sum_{i=1}^{n}Y_k(i)$ and $\mathbb{E}(A_k^2(n)|\mathcal{F}_n)=0$. Therefore, by the law of the iterated logarithm for martingales (see, e.g., [2]) $\sum_{i=1}^{n}A_k^2(i)$ is of order $(n^{1/2}\ln\ln n)^{1/2}$. Thus,

$$|\sum_{i=1}^{n}S_k(i)|=|\sum_{i=1}^{n}S_k(i)+\epsilon_k^1(n)+\epsilon_k^2(n)+X_k(n)|$$

$$\leq|\sum_{i=1}^{n}S_k(i)+\epsilon_k^1(n)|+|\epsilon_k^2(n)|+|X_k(n)|\leq|\sum_{i=1}^{n}S_k(i)|+|\epsilon_k^2(n)|+|X_k(n)|$$

$$\leq...\leq|\sum_{i=1}^{n}\epsilon_k^2(i)|+|\sum_{i=1}^{n}X_k(i)|.$$

We come to the desire statement.

References

1. *Malyshkin Y.* Preferential attachment combined with random number of choices // ArXiv 1612.02229. — https://arxiv.org/abs/1612.02229.
2. *Stout W.* A martingale analoque of Kolmogorov's law of iterated logarithm // Z. Vahr. Verw. Geb. — 1970. — Vol. 15. — P. 279–290.