UDC 519.17, 519.21

NUMBER OF MAXIMAL ROOTED TREES IN PREFERENTIAL ATTACHMENT MODEL VIA STOCHASTIC APPROXIMATION[1]

Malyshkin Y.A.
Tver State University, Tver
Moscow Institute of Physics and Technology, Moscow

Received 24.11.2022, revised 21.06.2023.

We study the asymptotic behavior of the number of maximal trees in the preferential attachment model. In our model, we consider a sequence of graphs built by the following recursive rule. We start with the complete graph on $m+1$ vertices, $m > 1$. Then on the $n+1$ step, we add vertex $n+1$ and draw m edges from it to different vertices from $1, \dots, n$, chosen with probabilities proportional to their degrees plus some positive parameter β. We prove the convergence speed for the number of maximal trees in such a model using the stochastic approximation technique.

Keywords: random graphs, preferential attachment, stochastic approximation.

Vestnik TvGU. Seriya: Prikladnaya Matematika [Herald of Tver State University. Series: Applied Mathematics], 2023, № 2, 28–36, https://doi.org/10.26456/vtpmk688

Introduction

The number of subgraphs of a graph is studied for many different graph models (see, e.g., [2,3]). In the present paper, we are focused on the number of maximal subtrees (a subtree is *maximal* in G_n if all its non-leaf vertices are adjacent only to vertices of that tree) for a preferential attachment model. The number of maximal subgraphs relates to a local structure of the graph and could be used, e.g., to prove logical convergence laws for random graphs (see, e.g., [7]) While the expected number of subgraphs is often obtained using combinatorial arguments (see, e.g., [6]) we would use stochastic approximation (see [1, 8] for more details on stochastic approximation processes) to obtain result about the convergence rate. Note that a similar result for the uniform attachment model was obtained in [5].

Let us describe the preferential attachment graph model considered in the paper. We start with a complete graph G_m on m vertices. Then on each step, we construct a graph

[1]Работа выполнена при финансовой поддержке РФФИ (проект № 19-31-60021).

G_n by adding to G_{n-1} a new vertex and drawing m edges from it to different vertices, chosen among vertices of G_n with probabilities proportional to their degrees plus a parameter $\beta > 0$. Note that in such a model (see, e.g., Lemma 3 of [4]), the maximum degree $M(n)$ of G_n is $o(n^{\frac{1}{\tau-1}+\epsilon})$ for any $\epsilon > 0$ almost surely (i.e. $\lim_{n\to\infty} \frac{M(n)}{n^{\frac{1}{\tau-1}+\epsilon}} = 0$ almost surely), where $\tau = 3 + \frac{\beta}{m}$. In particular, $M(n) = o(n^{1/2-\epsilon})$ for some $\epsilon > 0$.

For a rooted tree T, let $N_T(n)$ be the number of vertices that are roots of maximal subtrees of G_n isomorphic to T. Note that the set $\mathcal{T}_{N,b}$ of all isomorphism classes of rooted trees with at most N vertices of depth b is finite. We would refer to a maximal subtree of G_n isomorphic to a tree T from that set as *having the type* T (i.e. when we talk about the type of a tree in G_n we assume it is rooted and maximal). Also, we call a tree T *max-admissible* if it could be a maximal subtree of G_n for large enough n. Let us formulate our main result.

Theorem 1. *For max-admissible tree T there is a constant $\rho_T \in (0, 1)$, such that for any $\delta > 0$*

$$N_T(n) = \rho_T n + o(n^{1/2+\delta}) \quad a.s.$$

We would prove this result by induction over b using results about stochastic approximation processes.

Let us first describe these results. An r-dimensional process $Z(n)$ with the corresponding filtration $\mathcal{F}_n$ is called a stochastic approximation process if it could be written in the following way

$$Z(n+1) - Z(n) = \frac{1}{n+1}\left(F(Z(n)) + E_{n+1} + R_{n+1}\right), \tag{1}$$

where E_n, R_n, and the function F satisfy the following conditions (we would provide stronger conditions that are needed for [1, Theorem 3.1.1] to hold). There exists $U \subset \mathbb{R}^r$ such that $Z_n \in U$ for all n almost surely and

- A1 The function $F : \mathbb{R}^r \to \mathbb{R}^r$ has a unique root θ in U, and its components are twice continuously differentiable in some neighborhood of U.
- A2 The derivative matrix of $F(x)$ exists, and its biggest eigenvalue does not exceed $-1/2$.
- A3 E_n is a martingale difference with respect to $\mathcal{F}_n$, $\sup_n \mathbb{E}(|E_{n+1}|^2|\mathcal{F}_n) < \infty$ almost surely and for some $\delta \in (0, 1/2)$, $R_n = O(n^{-\delta})$ almost surely (i.e. there exists a non-random constant C, such that $\limsup_{n\to\infty} \frac{|R_n|}{n^{-\delta}} \leq C$ almost surely).

We need the following result:

Theorem 2. *[1, Theorem 3.1.1] Under the above conditions, $Z(n) \to \theta$ a.s. with the convergence rate*

$$|Z(n) - \theta| = o(n^{-\delta}) \quad \textit{almost surely.}$$

1. Number of vertices of fixed degree

We prove the theorem by induction over b. To prove the case $b = 1$ we need to prove the convergence rate for the number $N_k(n)$ of vertices with degree k at time n

for $k \geq m$. Let fix $N \in \mathbb{N}$, $N \geq m$. Let $X_k(n) := N_k(n)/n$, $m \leq k \leq N$. Let define

$$\rho_k := \frac{(2+\beta)m^{k-m}}{m^2+2+\beta} \prod_{i=m+1}^{k} \frac{i-1}{mi+2+\beta}, \quad k = m, \dots, N. \tag{2}$$

For $b = 1$, the statement of Theorem 1 could be formulated as follow.

Lemma 1. *$X_k(n) \to \rho_k$ with rate $|X_k(n) - \rho_k| = o(n^{-1/2+\delta})$ for any $\delta > 0$ a.s.*

Proof. The probability to draw an edge to a given vertex of degree k at step $n+1$ equals to

$$1 - \prod_{i=1}^{m} \frac{n(2+\beta) - k - \sum_{j=1}^{i-1} d_j(n)}{n(2+\beta) - \sum_{j=1}^{i-1} d_j(n)} = 1 - \prod_{i=1}^{m} \left(1 - \frac{k}{n(2+\beta) - \sum_{j=1}^{i-1} d_j(n)}\right),$$

where $d_j(n)$ is the degree of the vertex joined by the j-th edge. Since $d_j(n) = o(n^{1/2-\epsilon})$ for some $\epsilon > 0$, we get that probability to draw and edge to a given vertex of a degree k equals

$$\frac{mk}{n(2+\beta)} + o\left(\frac{1}{n^{3/2+\epsilon}}\right). \tag{3}$$

Let $\mathcal{F}_n$ be the filtration that corresponds to the graphs G_n. We get

$$\mathbb{E}\left(N_m(n+1) - N_m(n) | \mathcal{F}_n\right) = 1 - \frac{m^2}{n(2+\beta)} N_m(n) + o\left(\frac{N_m(n)}{n^{3/2+\epsilon}}\right),$$
$$\mathbb{E}\left(N_k(n+1) - N_k(n) | \mathcal{F}_n\right) = \frac{m(k-1)}{n(2+\beta)} N_{k-1}(n) - \frac{mk}{n(2+\beta)} N_k(n) + o\left(\frac{1}{n^{1/2+\epsilon}}\right),$$

$k = m+1, \dots, N$. For $X_k(n)$ we get

$$\mathbb{E}\left(X_k(n+1) - X_k(n) | \mathcal{F}_n\right) = \frac{1}{n+1}\left(\mathbb{E}\left(N_k(n+1) - N_k(n) | \mathcal{F}_n\right) - X_k(n)\right). \tag{4}$$

Let us define functions

$$f_m(x_m, \dots, x_N) = 1 - \left(\frac{m^2}{2+\beta} + 1\right) x_m,$$
$$f_k(x_m, \dots, x_N) = \frac{m(k-1)}{2+\beta} x_{k-1} - \left(\frac{mk}{2+\beta} + 1\right) x_k, \quad k = m+1, \dots, N.$$

Then, for all $k \in [m, N]$,

$$\mathbb{E}\left(X_k(n+1) - X_k(n) | \mathcal{F}_n\right) = \frac{1}{n+1}\left(f_k\left(X_m(n), \dots, X_N(n)\right) + o\left(\frac{1}{n^{1/2+\epsilon}}\right)\right). \tag{5}$$

For the vector $Z(n) := (X_m(n), \dots, X_N(n))$ we get

$$Z(n+1) - Z(n) = \frac{1}{n+1}\left(F(Z(n)) + (n+1)(Z(n+1) - \mathbb{E}(Z(n+1)|\mathcal{F}_n)) + o\left(\frac{1}{n^{1/2+\epsilon}}\right)\right),$$

where $F(x_m,\dots,x_N)=(f_m(x_m,\dots,x_N),\dots,f_N(x_m,\dots,x_N))^t$. Set

$$E_{n+1}=(n+1)(Z(n+1)-\mathbb{E}(Z(n+1)|\mathcal{F}_n)),\quad R_{n+1}=o\left(\frac{1}{n^{1/2+\epsilon}}\right).$$

Let us find nulls of the system $F(x_m,\dots,x_N)=0$, i.e. the system

$$\left\{\begin{array}{rcll} 1-\frac{m^2}{2+\beta}x_m & = & x_m, & \\ \frac{m(k-1)}{2+\beta}x_{k-1}-\frac{mk}{2+\beta}x_k & = & x_k, & k=m+1,\dots,N. \end{array}\right. \tag{6}$$

We get

$$\begin{aligned} x_m &= \frac{2+\beta}{m^2+2+\beta}, \\ x_k &= \frac{m(k-1)}{mk+2+\beta}x_{k-1}, \quad k=m+1,\dots,N. \end{aligned}$$

Therefore for $k=m+1,\dots,N$

$$x_k=\frac{(2+\beta)m^{k-m}}{m^2+2+\beta}\prod_{i=m+1}^{k}\frac{i-1}{mi+2+\beta},$$

and, hence, the system (6) has a unique solution $x_k=\rho_k$, $k=m,\dots,N$. Let us check the conditions of Theorem 2. The non-zero partial derivatives of functions f_k, $k=m,\dots,N$, equals

$$\left\{\begin{array}{rcll} \frac{\partial f_m}{\partial x_m}(x_m,\dots,x_d) & = & -\frac{m^2}{2+\beta}-1, & \\ \frac{\partial f_k}{\partial x_{k-1}}(x_m,\dots,x_d) & = & \frac{m(k-1)}{2+\beta}, & k=m+1,\dots,N, \\ \frac{\partial f_k}{\partial x_k}(x_m,\dots,x_d) & = & -\frac{mk}{2+\beta}-1, & k=m+1,\dots,N. \end{array}\right. \tag{7}$$

Since diagonal elements exceed below-diagonals by 1, the largest real part of the eigenvalues of the derivative matrix equals -1. Hence, the process $Z(n)$ satisfies the conditions A1, A2 of Theorem 2. To check condition A3 we first recall that $R_{n+1}=o\left(\frac{1}{n^{1/2+\epsilon}}\right)$ for some $\epsilon>0$. At each step, we draw m edges, so we change the degrees of exactly m vertices while adding one new vertex. Therefore, $|N_k(n+1)-N_k(n)|\le m+1$ and $|X_k(n+1)-X_k(n)|\le\frac{m+1}{n}$. Hence, for E_{n+1} we get

$$\begin{aligned} |E_{n+1}| &\le (n+1)\left(|Z(n+1)-Z(n)|+|\mathbb{E}(Z(n+1)-Z(n)|\mathcal{F}_n)|\right) \\ &\le 2\frac{(n+1)(m+1)(N-m+1)}{n}, \end{aligned}$$

which results in condition $A3$. By Theorem 2, we get the statement of Lemma 1. □

2. Number of rooted trees

Now we finish the proof of Theorem 1 by proving the induction step over tree depth b. Let us fix $b>1$ and $N\in\mathbb{N}$ (we assume N is large enough so at least one achievable tree of depth b on N vertices exists). We assume that the statement of Theorem 1 and some auxiliary statements over the course of the proof are true for all maximal trees of depth at most $b-1$ on at most N vertices.

Proof. Let us define variables $X_T(n) := N_T(n)/n$ and vector $Z_b(n) := (X_{T_i}(n))$ over all rooted trees $T_i \in \mathcal{T}_{N,b}$ that could be maximal subtrees of G_n (there are only finitely many such trees). We suggest that the order of the elements of $Z_b(n)$ is defined in a way such that the addition of new branches (that preserves the depth of the tree) increases the order.

Note that

$$\mathbb{E}(X_T(n+1) - X_T(n)|\mathcal{F}_n) = \frac{1}{n+1}\left(\mathbb{E}(N_T(n+1) - N_T(n)|\mathcal{F}_n) - X_T(n)\right).$$

There are two ways to change $Z_b(n)$ at time $n+1$.

First, we could draw an edge to a tree of type $T \in \mathcal{T}_{N,b}$. This results in the decrease of N_T by 1 and a possible increase in one of the bigger components of $Z_b(n)$ (when the tree changes type to the type $T' \in \mathcal{T}_{N,b}$). For the latter to happen, we need to draw the rest of the edges to the roots of the maximal non-intersecting trees of depth at most $b-2$ of given types $T_1, ..., T_{m-1} \in \mathcal{T}_{N,b-2}$ (with bounded degrees since $\mathcal{T}_{N,b}$ contains trees on at most N vertices). Since the degrees of such trees are bounded, the probability to draw edges to intersecting trees is $O\left(\frac{1}{n}\right)$. Hence, the expected numbers of trees with type changes between $T \in \mathcal{T}_{N,b}$ and $T' \in \mathcal{T}_{N,b}$ (T' is bigger than T) is polynomial of $X_T(n), X_{T_1}(n), \dots, X_{T_{m-1}}(n)$ up to the term $o(n^{-1/2})$. Note that the expected number of trees of type T that changes type is polynomial of $X_T(n)$ up to the term $o(n^{-1/2})$ as well.

The second way to change $Z_b(n)$ is to create a maximal tree of type $T \in \mathcal{T}_{N,b}$ with root $n+1$. To do so we need to draw edges from $n+1$ to m roots of the maximal non-intersecting tree of depth $b-1$ of given types $T_1, ..., T_m$ (with bounded degrees). The probability of creating a maximal tree of type T this way is polynomial of $X_{T_1}(n), \dots, X_{T_m}(n)$. Note that the degree of the root of such trees would be equal to m, so they would be among the smallest trees from $T_i \in \mathcal{T}_{N,b}$, including the smallest achievable tree.

As result we get

$$\mathbb{E}(Z_b(n+1) - Z_b(n)|\mathcal{F}_n) = \frac{1}{n+1}\left(A_b Z_b(n) - Z_b(n) + Y_b + o(n^{-1/2})\right)$$

where $A_b = A_b(Z_1(n), \dots, Z_{b-2}(n))$ is a lower-triangular matrix with negative elements on the diagonal and non-negative under the diagonal and $Y_b = Y_b(Z_{b-1}(n))$ is a vector, such that the elements of both A_b and Y_b are polynomials of $X_{T_i}(n)$, where T_i are trees of depth at most $b-2$ (for A_b) or exactly $b-1$ (for Y_b). Let consider $F_b(Z_1, \dots, Z_b) := A_b Z_b(n) - Z_b(n) + Y_b$ (note that A_b and Y_b are functions of $Z_1, \dots, Z_{b-1}$ itself). Note that F_b is deterministic. By induction assumption, there is a unique solution of the system $F_i(z_1, \dots, z_i) = 0, i = 1, \dots, b-1$. Let us define $H_b(z_b) = F_b(z_1^*, \dots, z_{b-1}^*, z_b)$. Then $H_b(z_b) = 0$ is a system of linear equations with the unique root z_b^* since A_b is lower-triangular with negative elements on the diagonal. Now let us show that all components of z_b^* are positive. Recall that all elements under the diagonal of A_b are non-negative and each (except the first) row has at least one positive element outside the diagonal (if a tree is not the smallest possible, we could remove one vertex with its children from it to make it smaller). All components of $Y_b(z_{b-1}^*, \rho_d)$ are non-negative as well. Finally, the first element of Y_b is positive since the smallest max-admissible tree of depth b (which corresponds to the first coordinate of z_b) could be obtained by drawing edges from a new vertex to the smallest max-admissible trees of depth $b-1$ and the first coordinate of z^*_{b-1} is positive by the induction hypothesis.

Let us consider the vector $W_b(n) = (Z_1(n), \dots, Z_b(n))$. We get that

$$\mathbb{E}(W_b(n+1) - W_b(n)|\mathcal{F}_n) = \frac{1}{n+1}\Big((F_1, \dots, F_b) + o(n^{-1/2})\Big).$$

The derivative matrix of function $(F_1, \dots, F_b)(z_1, \dots, z_b)$ is of following form. Around the diagonal, it has clusters of derivatives of F_i with respect to z_i, which are lower-triangular (since $F_i = A_i z_i - z_i + Y_i$) with diagonal elements at most -1. Since F_i depends only on $z_1, \dots, z_i$, all elements above diagonal clusters are 0. Therefore the highest eigenvalue of the derivative matrix of $(F_1, \dots, F_b)$ is -1 (for all possible process values). Hence $W_b(n)$ satisfies condition A2 of Theorem 2. Since functions $(F_1, \dots, F_b)$ have second-order derivatives, condition A1 is satisfied as well. To check condition A3 note that if we take

$$E_{n+1} = (n+1)(W_b(n+1) - \mathbb{E}(W_b(n+1)|\mathcal{F}_n)),$$

then

$$\begin{aligned} R_{n+1} :&= (n+1)(W_b(n+1) - W_b(n)) - (F_1, \dots, F_b) - E_{n+1} \\ &= (n+1)\mathbb{E}(W_b(n+1) - W_b(n)|\mathcal{F}_n) - (F_1, \dots, F_b) = o(n^{-1/2}) \quad \text{a.s.} \end{aligned}$$

and

$$|E_{n+1}| \le (n+1)|W_b(n+1) - W_b(n)| + (n+1)|\mathbb{E}(W_b(n+1) - W_b(n)|\mathcal{F}_n)| \le C$$

for some constant C since the number of maximal trees (on at most N vertices) of depth b that the vertex $n+1$ could impact is bounded from above by a constant, which results in condition A3. Therefore, due to Theorem 2 $W_b(n)$ converges a.s. to $(z^*_1, \dots, z^*_b)$ with the rate $o(n^{-1/2+\delta})$ for any $\delta > 0$ almost surely. □

References

[1] Chen H.F., *Stochastic Approximation and Its Applications*. V. 64, Nonconvex Optimization and its Applications, Springer, New York, 2002, 360 pp., https://doi.org/10.1007/b101987.

[2] Frieze A., Karonski M., *Introduction to random graphs*, Cambridge University Press, 2016, 478 pp.

[3] Garavaglia A., *Preferential attachment models for dynamic networks*, Technische Universiteit Eindhoven, 2019, 305 pp.

[4] Malyshkin Y.A., “γ-variable first-order logic of preferential attachment random graphs”, *Discrete Applied Mathematics*, **314** (2022), 223–227.

[5] Malyshkin Y.A., “Number of maximal rooted trees in uniform attachment model via stochastic approximation”, *Vestnik TvGU. Seriya: Prikladnaya Matematika*, **3** (2022), 27–34.

[6] Malyshkin Y.A., Zhukovskii M.E., "MSO 0-1 law for recursive random trees", *Statistics and Probability Letters*, **173** (2021), 109061.

[7] Malyshkin Y.A., Zhukovskii M.E., *Logical convergence laws via stochastic approximation and Markov processes*, Preprint at https://arxiv.org/abs/2210.13437, 2022.

[8] Pemantle R., "A survey of random processes with reinforcement", *Probability Surveys*, **4** (2007), 1–79.

Citation

Malyshkin Y.A., "Number of maximal rooted trees in preferential attachment model via stochastic approximation", *Vestnik TvGU. Seriya: Prikladnaya Matematika [Herald of Tver State University. Series: Applied Mathematics]*, 2023, № 2, 28–36. https://doi.org/10.26456/vtpmk688

Author Info

1. **Malyshkin Yury Andreyevich**
   Associate Professor at the Department of Information Technology, Tver State University; senior researcher at Moscow Institute of Physics and Technology.

   Russia, 170100, Tver, 33 Zhelyabova str., TverSU. E-mail: yury.malyshkin@mail.ru

ИССЛЕДОВАНИЕ ЧИСЛА МАКСИМАЛЬНЫХ ДЕРЕВЬЕВ В МОДЕЛИ ПРЕДПОЧТИТЕЛЬНОГО ПРИСОЕДИНЕНИЯ С ПОМОЩЬЮ СТОХАСТИЧЕСКОЙ АППРОКСИМАЦИИ

Малышкин Ю.А.
Тверской государственный университет, г. Тверь
Московский физико-технический институт, г. Москва

Поступила в редакцию 24.11.2022, после переработки 21.06.2023.

В статье исследуется асимптотическое поведение числа максимальных деревьев в модели графов предпочтительного присоединения. В предлагаемой модели рассматривается последовательность графов, которая строится по следующему рекурсивному правилу. Мы начинаем построение с полного графа на $m+1$ вершине, $m > 1$. Затем на $n+1$-ом шаге мы добавляем вершину $n+1$ и проводим из нее m ребер в различные вершины, выбранные с вероятностями, пропорциональными их степеням плюс некоторый положительный параметр β. В статье получен результат о скорости сходимости числа максимальных деревьев в указанной модели с помощью стохастической аппроксимации.

Ключевые слова: случайные графы, предпочтительное присоединение, стохастическая аппроксимация.

Образец цитирования

Malyshkin Y.A. Number of maximal rooted trees in preferential attachment model via stochastic approximation // Вестник ТвГУ. Серия: Прикладная математика. 2023. № 2. С. 28–36. https://doi.org/10.26456/vtpmk688

Список литературы

[1] Chen H.F. Stochastic Approximation and Its Applications. Series: Nonconvex Optimization and its Applications. Vol. 64. New York: Springer, 2002. 360 p. https://doi.org/10.1007/b101987

[2] Frieze A., Karonski M. Introduction to random graphs. Cambridge University Press, 2016. 478 p.

[3] Garavaglia A. Preferential attachment models for dynamic networks. Technische Universiteit Eindhoven, 2019. 305 p.

[4] Malyshkin Y.A. γ-variable first-order logic of preferential attachment random graphs // Discrete Applied Mathematics. 2022. Vol. 314. Pp. 223–227.

[5] Malyshkin Y.A. Number of maximal rooted trees in uniform attachment model via stochastic approximation // Vestnik TvGU. Seriya: Prikladnaya Matematika. 2022. Vol. 3. Pp. 27–34.

[6] Malyshkin Y.A., Zhukovskii M.E. MSO 0-1 law for recursive random trees // Statistics and Probability Letters. 2021. Vol. 173. ID 109061.

[7] Malyshkin Y.A., Zhukovskii M.E. Logical convergence laws via stochastic approximation and Markov processes. Preprint at https://arxiv.org/abs/2210.13437. 2022.

[8] Pemantle R. A survey of random processes with reinforcement // Probability Surveys. 2007. Vol. 4. Pp. 1–79.