UDC 519.17, 519.21

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

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

Received 08.08.2022, revised 30.08.2022.

We study the asymptotic behavior of the number of maximal trees in a uniform 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, chosen uniformly from $1,\dots,n$. We prove the convergence speed for the number of maximal trees in such a model using the stochastic approximation technique.

Keywords: random graphs, uniform attachment, stochastic approximation.

Vestnik TvGU. Seriya: Prikladnaya Matematika [Herald of Tver State University. Series: Applied Mathematics], 2022, № 3, 27–34, https://doi.org/10.26456/vtpmk640

Introduction

The number of subgraphs of a graph is important in the understanding of the local structure and properties of the graph and was studied for various 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 uniform attachment model. While the expected number of subgraphs is often obtained using combinatorial arguments (see, e.g., [4]) we would use stochastic approximation (see [1,6] for more details on stochastic approximation processes) to obtain result about the convergence rate.

Let us describe the model of graphs $G_n = G_n(m)$ that we consider in the paper. We start with a complete graph G_m on m vertices. Then on each step, we construct a graph G_n by adding to G_{n-1} a new vertex and drawing m edges from it to different vertices, chosen uniformly among vertices of G_n.

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

[1]The present work was funded by RFBR, project number 19-31-60021.

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 $\limsup_{n\to\infty} \frac{|R_n|}{n^{-\delta}} \le C$ 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*

$$|Z(n) - \theta| = o(n^{-\delta}) \quad \textit{almost surely}\left(\textit{i.e.}\ \frac{|Z(n)-\theta|}{n^{-\delta}} \to 0 \textit{ almost surely}\right).$$

1. Number of vertices of fixed degree

First, let prove the case $b = 1$, which corresponds to the number $N_k(n)$ of vertices with degree k at time n for $k \ge m$ (the tree of depth 1 is a star and defined by the degree of its root). In our model at step $n+1$ probability to draw an edge to a given vertex equals to

$$1 - \frac{\binom{n-1}{m}}{\binom{n}{m}} = 1 - \frac{n-m}{n} = \frac{m}{n}. \tag{2}$$

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

$$\rho_k := \frac{m^{k-m}}{(m+1)^{k-m+1}}, \quad k = m, \ldots, N. \tag{3}$$

For $b = 1$, 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. Let $\mathcal{F}_n$ be the filtration that corresponds to the graphs G_n. We get

$$\mathbb{E}\left(N_m(n+1) - N_m(n)|\mathcal{F}_n\right) = 1 - \frac{m}{n}N_m(n),$$
$$\mathbb{E}\left(N_k(n+1) - N_k(n)|\mathcal{F}_n\right) = \frac{m}{n}\left(N_{k-1}(n) - N_k(n)\right), \quad k = m+1, \dots, N.$$

For $X_k(n)$ we get

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

Hence, if we define functions

$$f_m(x_m, \dots, x_N) = 1 - (m+1)x_m,$$
$$f_k(x_m, \dots, x_N) = mx_{k-1} - (m+1)x_k, \quad k = m+1, \dots, N,$$

we would get that for all $k \in [m, N]$,

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

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

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

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

$$E_{n+1} = (n+1)(Z(n+1) - \mathbb{E}(Z(n+1)|\mathcal{F}_n)), \quad R_{n+1} = 0.$$

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

$$\begin{cases} 1 - mx_m &= x_m, \\ m(x_{k-1} - x_k) &= x_k, \qquad k = m+1, \dots, N. \end{cases} \quad (6)$$

We get

$$x_m = \frac{1}{m+1},$$
$$x_k = \frac{m}{m+1}x_{k-1}, \quad k = m+1, \dots, N.$$

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

$$x_k = \frac{m^{k-m}}{(m+1)^{k-m+1}}.$$

Therefore the system (6) has a unique solution $x_k = \rho_k$, $k = m, \dots, N$. Let us check the conditions of Theorem 2. For non-zero partial derivatives of functions f_k, $k = m, \dots, N$, we would get:

$$\begin{cases} \frac{\partial f_m}{\partial x_m}(x_m, \dots, x_d) &= -m-1, \\ \frac{\partial f_k}{\partial x_{k-1}}(x_m, \dots, x_d) &= m, \qquad k = m+1, \dots, N, \\ \frac{\partial f_k}{\partial x_k}(x_m, \dots, x_d) &= -m-1, \qquad k = m+1, \dots, N. \end{cases} \quad (7)$$

Hence, the largest real part of the eigenvalues of the derivative matrix equals -1. Therefore the process $Z(n)$ satisfies the conditions A1, A2 of Theorem 2. To check condition A3 we first recall that $R_{n+1} = 0$. At each step, we draw m edges, so we change degrees of exactly m vertices while adding one new vertex. Hence, $|N_k(n+1) - N_k(n)| \leq m+1$ and $|X_k(n+1) - X_k(n)| \leq \frac{m+1}{n}$. Therefore, for E_{n+1} we get

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

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

2. Number of rooted trees

Now we prove Theorem 1 using induction over tree depth.

Proof. Let us fix $b > 1$ and large enough N and consider 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). Note that the case $b = 1$ refers to the number of stars and was already considered in Lemma 1. The order of the elements of $Z_b(n)$ (or, in other words, the order on the set of all max-admissible trees of depth b) 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 $N_T(n)$ at time $n+1$. We could draw an edge to a tree of type T or we could create a new copy of T rooted at $n+1$. Recall that due to equation (2) for each given vertex probability to draw an edge to it is $\frac{m}{n}$. In a rooted tree T, fix a non-leaf vertex u. Then the expected number (conditioned on G_n) of trees T' in G_n of type T such that an edge is drawn from $n+1$ to a vertex u' of T' and there exists an isomorphism of rooted trees $T \to T'$ sending u to u' equals

$$Cm\frac{N_T(n)}{n} = CmX_T(n),$$

where the constant $C = C(T,u)$ corresponds to the number of vertices that belong to the orbit of u under the action of the automorphism group.

The type of the maximal tree with root $n+1$ is defined by the types of trees of depth $b-1$, to which roots we draw m edges from vertex $n+1$. Note that, due to Lemma 5 of [5], the (conditional) probability to draw edges to trees that share a non-leaf vertex (thus creating a cycle) is $O(\frac{\ln^2 n}{n})$ almost surely and does not affect our argument. Hence, drawing a given edge to the root of a given tree does not impact (up to $O(\frac{\ln^2 n}{n})$ error term) probabilities to draw other edges to roots of other trees. Therefore, probability to create a tree of type T in the vertex $n+1$ is polynomial of

$X_{T_i}(n)$ up to $O(\frac{\ln^2 n}{n})$ error term, where T_i are max-admissible trees of depth $b-1$. To change a type of a given maximal tree in G_n (to another given type) of depth b we need to draw an edge to one of its vertices and draw the rest of the edges to the roots of trees of depth at most $b-2$ of given types (that depends on the given tree type). Such probability is polynomial of $X_{T_i}(n)$ up to a term $O(\frac{\ln^2 n}{n})$, where T_i are max-admissible trees of depth $b-2$.

Therefore

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

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 $F_b(Z_1,\dots,Z_b) := A_bZ_b(n)-Z_b(n)+Y_b$ (note that A_b and Y_b are functions of $Z_1,\dots,Z_{b-1}$ itself). Note that F_b is deterministic. We would use induction over b to prove that there is a unique solution of the system $F_i(z_1,\dots,z_i)=0, i=1,\dots,b$. We already established the existence of the unique (non-zero) root for the case $b=1$. Assume there are unique non-zero solutions $z_1^*,\dots,z_{b-1}^*$ of the systems $F_i(z_1,\dots,z_i)=0, i=1,\dots,b-1$. If we 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. Hence it is enough to show that the first element of Y_b is positive. It follows from the fact that 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

$$\mathbb{E}(W_b(n+1)-W_b(n)|\mathcal{F}_n)=\frac{1}{n+1}\left((F_1,\dots,F_b)+O\left(\frac{\ln^2 n}{n}\right)\right).$$

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 $F_i=A_iz_i-z_i+Y_i$) 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

$$R_{n+1} := (n+1)(W_b(n+1) - W_b(n)) - (F_1, \dots, F_b) - E_{n+1}$$
$$= (n+1)\mathbb{E}(W_b(n+1) - W_b(n)|\mathcal{F}_n) - (F_1, \dots, F_b) = O\left(\frac{\ln^2 n}{n}\right) \quad \text{a.s.}$$

and

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

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

[5] Malyshkin Y.A., Zhukovskii M.E., "γ-variable first-order logic of uniform attachment random graphs", *Discrete Mathematics*, **345**:5 (2022), 112802.

[6] 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 uniform attachment model via stochastic approximation", *Vestnik TvGU. Seriya: Prikladnaya Matematika [Herald of Tver State University. Series: Applied Mathematics]*, 2022, № 3, 27–34. https://doi.org/10.26456/vtpmk640

Information about Authors

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 РАВНОМЕРНОГО ПРИСОЕДИНЕНИЯ С ПОМОЩЬЮ СТОХАСТИЧЕСКОЙ АППРОКСИМАЦИИ

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

Поступила в редакцию 08.08.2022, после переработки 30.08.2022.

В статье исследуется ассимптотическое поведение числа максимальных деревьев в модели графов равномерного присоединения. В предлагаемой модели рассматривается последовательность графов, которая строится по следующему реккурсивному 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 разные вершины, выбранные равномерно из вершин $1,\ldots,n$. В статье получен результат о скорости сходимости числа максимальных деревьев в указанной модели с помощью стохастической аппроксимации.

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

Образец цитирования

Malyshkin Y.A. Number of maximal rooted trees in uniform attachment model via stochastic approximation // Вестник ТвГУ. Серия: Прикладная математика. 2022. № 3. С. 27–34. https://doi.org/10.26456/vtpmk640

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

[5] Malyshkin Y.A., Zhukovskii M.E. γ-variable first-order logic of uniform attachment random graphs // Discrete Mathematics. 2022. Vol. 345, № 5. ID 112802.

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