ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2024, Vol. 85, No. 11, pp. 1095–1107.

Investigation of Triangle Counts in Graphs Evolving by Clustering Attachment

M. Vaičiulis[*,a] **and N. M. Markovich**[**,b]

[*] *Institute of Data Science and Digital Technologies, Vilnius, Lithuania*
[**] *Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia*
e-mail: [a] *marijus.vaiciulis@mif.vu.lt,* [b] *nat.markovich@gmail.com*

Received January 18, 2024
Revised August 26, 2024
Accepted September 20, 2024

Abstract—The clustering attachment (CA) model proposed by Bagrow and Brockmann in 2013 may be used as an evolution tool for undirected random networks. A general definition of the CA model is introduced. Theoretical results are obtained for a new CA model that can be treated as the former's limit in the case of the model parameters $\alpha \to 0$ and $\epsilon = 0$. This study is focused on the triangle count of connected nodes at an evolution step n, an important characteristic of the network clustering considered in the literature. As is proved for the new model below, the total triangle count Δ_n tends to infinity almost surely as $n \to \infty$ and the growth rate of $E\Delta_n$ at an evolution step $n \geqslant 2$ is higher than the logarithmic one. Computer simulation is used to model sequences of triangle counts. The simulation is based on the generalized Pólya–Eggenberger urn model, a novel approach introduced here for the first time.

Keywords: clustering attachment, clustering coefficient, node weight, random graph, evolution, urn model

DOI: 10.31857/S0005117924110039

1. INTRODUCTION

Random graphs are used to model real-world networks [1]. These are graphs in which the node connectivity structure is random due to the dynamics (evolution) of networks over time. Random graphs can be described by a probabilistic distribution or random process that generates them [1–3]. One example of a random graph is the Erdös–Rényi graph, where an edge between any two different graph vertices appears with some fixed probability p, independent of the other pairs of vertices.

This paper is devoted to the clustering attachment (CA) rule proposed in [4]. It can be used as an evolution model of undirected local random networks when nodes belong to a certain community. The idea behind CA is that a newly appending node chooses existing nodes not with a high number of links but on the basis of their belonging to strongly connected groups of nodes (a typical example is the social behavior of people) and attaches to these nodes with new edges. CA is intended to model local networks, not leading to the emergence of nodes with a large number of links, in contrast to the preferential attachment (PA) implementing the so-called "rich-get-richer" models. In accordance with CA, a new node connects to $m \geqslant 2$ existing nodes. For example, in social networks, individuals may have friends not among popular people but in their neighborhood, forming close communities. CA may fit the development management model of a region without large metropolises or a model of society with an approximately equal distribution of benefits, under which the enrichment of individual nodes becomes impossible. CA can be used in medicine when modeling the evolution of brain neural networks and in biology when modeling communication in animal packs [4].

The model proposed in [4] describes the attachment of a new node to an existing node i with a probability proportional to its clustering coefficient $c_{i,n}$ at an evolution step n :

$$P_{CA}(i,n) \;\propto\; c_{i,t}^{\alpha} + \epsilon, \tag{1}$$

where the clustering coefficient of node $i \in V_n$ is defined by

$$c_{i,n} = \begin{cases} 0, & D_{i,n} = 0 \text{ or } D_{i,n} = 1 \\ 2\Delta_{i,n}/\left(D_{i,n}\left(D_{i,n}-1\right)\right), & D_{i,n} \geqslant 2, \end{cases}$$

$D_{i,n}$ is the degree of node i (the number of its links to other nodes), $\Delta_{i,n}$ is the number of triangles of node i, and $0 \leqslant \epsilon \leqslant 1$ and $\alpha \geqslant 0$ are model parameters.[1] As a rule, the connection of a newly appending node to existing nodes by the CA reduces its clustering coefficient [4]. Despite that CA leads to numerous new phenomena (community formation around new nodes, modularity bursts, and light-tailed distributions of node degrees, see [4]), it was not further developed. Many results in [4] were not rigorously established, such as the proof of light-tailed distributions of node degrees and the cluster structure of modularity; they are important for analyzing the connectivity of network nodes.

This paper aims to theoretically investigate a more general CA model than the original one [4] in the case of the model parameters $\alpha \to 0$ and $\epsilon = 0$. This model reduces to successively choosing $m \geqslant 2$ existing nodes included in the triangles by a newly appending node at each evolution step. Note that the numbers of nodes and edges in the graph are nonrandom, and randomness is introduced by choosing an existing node with a probability proportional to its weight.

A triangle of connected nodes is the most studied subgraph, which can be considered as a basic community. Perhaps, it was first introduced in [5]. Also, the triangle count serves to calculate the clustering coefficient, an important characteristic of the clustering structure of random graphs expressing the fraction of connected neighbors of a node. The limit behavior of the triangle count of nodes attracts the attention of many researchers; for example, see [6–8]. In this paper, we study the limit behavior of the total triangle counts Δ_n in an evolving CA graph as $n \to \infty$ and derive a lower bound for the expectation $E\Delta_n$ in the proposed CA model, which is a novel result.

The remainder of the paper is organized as follows. Section 2 describes the model under consideration. In Section 3 we discuss the successive choice of existing nodes by a newly appending node. The main results are presented in Section 4. The results of computer simulation are provided in Section 5. Section 6 summarizes the findings of this study. All proofs are given in the Appendices.

2. DESCRIPTION OF THE LIMIT MODEL

We denote by $G_n = (V_n, E_n)$, $n = 1, 2, \ldots$, a sequence of random graphs in which V_n and E_n are the sets of nodes and edges, respectively, and n may be interpreted as the discrete time since graphs are generated during the evolution. The evolution starts from an initial graph G_1 in which the numbers of nodes and edges are fixed. Throughout this paper, we adopt the original notations from [1], also denoting by $\#A$ the cardinality (the number of elements) of an arbitrary finite set A. For the graph G_n, we have

$$\#V_n = \#V_1 + n - 1, \qquad \#E_n = \#E_1 + m(n-1), \forall n \in \mathbb{N}, \tag{2}$$

where a natural number $m \geqslant 2$ is a parameter of the CA model.

The CA model defined below is more general than the original one [4].

[1] The notation $x \propto y$ means the existence of a nonzero constant C such that $x = Cy$.

Definition 1. The graph G_{n+1} is obtained from G_n according to the following two-part rule:
(a) the deterministic part: a new node $\#V_1 + n \in V_{n+1} \setminus V_n$ is appended to G_n;
(b) the stochastic part: each node $i \in V_n$ is equipped with the weight

$$p_{i,n} = \frac{f\left(c_{i,n}\right) + \epsilon}{\sum_{j \in V_n} \left(f\left(c_{j,n}\right) + \epsilon\right)}, \tag{3}$$

where $f : [0,1] \to [0,\infty)$ is a deterministic nondecreasing attachment function such that

$$f(0) = 0 \tag{4}$$

and $\epsilon \geqslant 0$ is a parameter of the CA model. The existing nodes $i_1, \ldots, i_m \in V_n$ are chosen by successive sampling[2] with probabilities proportional to their weights, and each new node $\#V_1 + n$ appends to each existing node by only one edge at each evolution step.

By Definition 1, the parameter ϵ is not necessarily bounded above by 1 due to normalization in formula (3), thereby differing from the CA model [4]. To eliminate the case of no evolution ($G_n = G_1$, $n \geqslant 2$) from consideration, we assume that the initial graph G_1 satisfies the condition

$$\#V_1 \geqslant m \quad \text{for } \epsilon > 0 \tag{5}$$

and

$$\#\tilde{V}_1 \geqslant m \quad \text{for } \epsilon = 0. \tag{6}$$

Here $\tilde{V}_n = \{i \in V_n : \ c_{i,n} > 0\}$, $n \geqslant 1$. From (2) and (5) it follows that $V_n \geqslant m$; for $\epsilon = 0$, condition (6) implies $\#\tilde{V}_n \geqslant m$ for each $n \in \mathbb{N}$. Hence, for $\epsilon \geqslant 0$ and $\forall n \in \mathbb{N}$, there exists at least one collection of nodes $i_1, \ldots, i_m \in V_n$ that can be chosen with a positive probability by successive sampling.

By replacing $D_{i,n}$ with $c_{i,n}$ in (3) and the definitional domain of the function $f(x)$ with $\{0\} \cup \mathbb{N}$, we obtain the definition of the PA model; see p. 5 in [9]. Substituting

$$f(x) = x^{\alpha}, \quad \alpha > 0, \tag{7}$$

into (3) yields the weights (1) for the CA model introduced in [4].

In view of (4), the weights (3) can be written as

$$p_{i,n}(\epsilon) = \frac{1}{\#V_n \epsilon + \sum_{j \in \tilde{V}_n} f\left(c_{j,n}\right)} \begin{cases} \epsilon, & i \in V_n \setminus \tilde{V}_n \\ f\left(c_{i,n}\right) + \epsilon, & i \in \tilde{V}_n. \end{cases} \tag{8}$$

The value of ϵ determines from which set the nodes $i_1, \ldots, i_m$ are chosen by successive sampling: this is the set $\tilde{V}_n$ for $\epsilon = 0$ and the set V_n for $\epsilon > 0$. If $\epsilon > 0$, the inequalities

$$\#V_n \geqslant \#\tilde{V}_n, \quad n \geqslant 1, \tag{9}$$

and formula (8) give the following inequalities for the weights:

$$\frac{1}{C_{f,\epsilon} \#V_n} \leqslant p_{i,n}(\epsilon) \leqslant \frac{C_{f,\epsilon}}{\#V_n}, \quad n \geqslant 1, \quad i \in V_n, \tag{10}$$

where $C_{f,\epsilon} = \left(f(1) + \epsilon\right)/\epsilon$. Due to (10), $p_{i,n} \to 0$ almost surely (a.s.) as $n \to \infty$. In the case $\epsilon = 0$, the behavior of the weights $p_{i,n}$, $i \in \tilde{V}_n$, becomes less obvious. Note that $p_{i,n}$, $i \in \tilde{V}_n$, strongly depend on the positive random value $\#\tilde{V}_n$, and the sequence $\#\tilde{V}_1, \#\tilde{V}_2, \ldots$ is formed randomly: $\#\tilde{V}_{n+1} = \#\tilde{V}_n + 1$ is true only if the successively chosen nodes $i_1, \ldots, i_m \in \tilde{V}_n$ contain at least one pair of nodes connected via an edge from the set E_n.

[2] For $m = 2$, the successive choice of nodes is described in Section 3.

Therefore, for $\epsilon = 0$, we consider the CA model with the attachment function

$$f(x) = \begin{cases} 0, & x = 0 \\ 1, & x > 0. \end{cases} \tag{11}$$

If the attachment function has the form (7) or (11), then $C_{f,\epsilon} = (1+\epsilon)/\epsilon$. Substituting (11) into (12) yields

$$p_{i,n}(\epsilon) = \begin{cases} \epsilon/\left(\#V_n\epsilon + \#\tilde{V}_n\right), & i \in V_n \setminus \tilde{V}_n, \quad V_n \setminus \tilde{V}_n \neq \emptyset \\ (1+\epsilon)/\left(\#V_n\epsilon + \#\tilde{V}_n\right), & i \in \tilde{V}_n. \end{cases} \tag{12}$$

Let us restrict the analysis to the case $\epsilon = 0$, in which the choice of a node from $V_n \setminus \tilde{V}_n$ is excluded a.s. and the weights of the nodes belonging to the set $\tilde{V}_n$ obey the uniform distribution:

$$p_{i,n}(0) = \begin{cases} 0, & i \in V_n \setminus \tilde{V}_n, \quad V_n \setminus \tilde{V}_n \neq \emptyset \\ 1/\#\tilde{V}_n, & i \in \tilde{V}_n. \end{cases} \tag{13}$$

The sequence of functions $f_n(x) = x^{1/n}$, $n = 1, 2, \ldots$, pointwise converges to $f(x)$ in (11). Therefore, the CA model with the attachment function (11) and the parameter $\epsilon = 0$ can be treated as the limit (as $\alpha \downarrow 0$) for the family of CA models with the attachment function (7) and $\epsilon = 0$.

3. SUCCESSIVE UNIFORM SAMPLING OF NODES

Let G_n be an observed graph. The objective of this section is to show that the sets $\tilde{E}_n$ and $\tilde{E}_n \cap E_n$ are enough to know for the evolution by the CA model. Due to (13), an appended node $\#V_1 + n$ can be attached with a positive probability only to existing nodes $\{i \in V_n : \Delta_{i,n} > 0\}$ involved in at least one triangle.

Let the sequence of sets $\mathcal{E}_1, \mathcal{E}_2, \ldots$ be such that for any $n \geqslant 1$, the pair $(V_n, \mathcal{E}_n)$ is a complete graph. We denote by W_n a pair of nodes $\{i_1, i_2\}$ from V_n that are chosen by successive sampling (random sampling without replacement). Successive sampling begins with choosing node i_1 from V_n with the probability (13). The probabilities of the remaining $\#V_n - 1$ nodes are then renormalized using the formula $p_{i,n}/(1 - p_{i_1,n})$. The process is repeated by choosing node i_2. The evolution by the CA model makes no distinction between the first and second nodes in a pair. Therefore, we will assume that W_n is a two-dimensional random vector with values in $\mathcal{E}_n$.

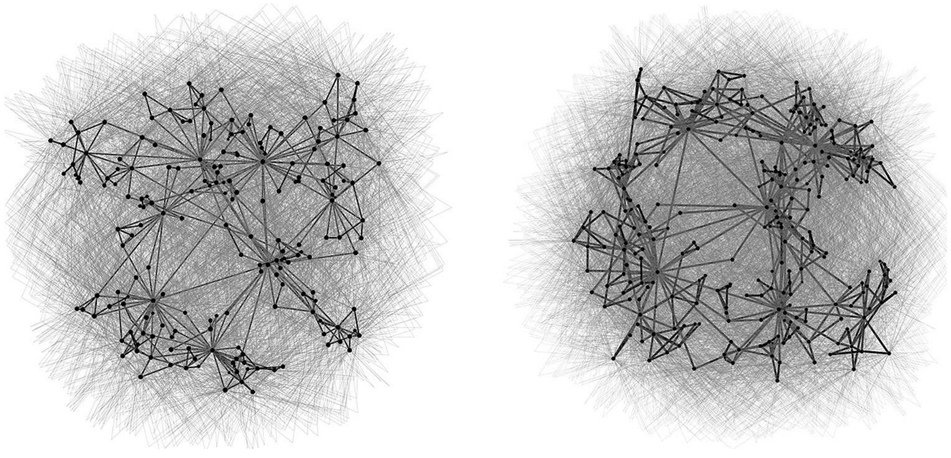

Fig. 1. The CA random graph (V_n, E_n) with $n = 5000$ (left) and $n = 10\,000$ (right): the nodes from $\tilde{V}_n$ are shown in black, the edges from $\tilde{E}_n \cap E_n$ in dark grey, and the rest of the graph in light grey.

Proposition 1. *The random vector W_n obeys the uniform distribution on the set $\tilde{E}_n$.*

According to Proposition 1, the attachment rule (b) with the attachment function (11) and the parameters $\epsilon = 0$ and $m = 2$ (see Definition 1) gives no preference to a pair of nodes with relatively large clustering coefficients. This means the absence of the "*rich-get-richer*" effect in the CA model under consideration.

The equality $P(W_n = \{i,j\}) = 0$, $i,j \in V_n \setminus \tilde{V}_n$, means that nodes from the set $V_n \setminus \tilde{V}_n$ take no part in forming the graph G_{n+1}: the newly appending nodes do not attach to such nodes. With partitioning the set V_n into the subsets $\tilde{V}_n$ and $V_n \setminus \tilde{V}_n$ (active and inactive nodes, respectively), the CA model can be used to generate various two-class communities.

Example 1. Figure 1 illustrates the classes mentioned above for one realization of the CA random graph (V_n, E_n) with $n \in \{5000, 10\,000\}$, where the initial graph G_1 is a quadrilateral with all diagonals except one:

$$V_1 = \{1,2,3,4\}, \quad E_1 = \{\{1,2\},\{1,3\},\{1,4\},\{2,3\},\{3,4\}\}.$$

4. THE MAIN RESULT

The total triangle count in the graph G_n is expressed through the triangle counts of its nodes:

$$\Delta_n = \frac{1}{3} \sum_{i \in \tilde{V}_n} \Delta_{i,n}.$$

Consider the CA model with the parameter $\epsilon = 0$ and the attachment function (11). As was noted in [4], an empirical study of the CA model with the parameter $m > 2$ requires bulky calculations due to complex combinatorics. Therefore, we restrict the analysis to the case $m = 2$.

Theorem 1. *Let G_n, $n = 1, 2, \ldots$, be a sequence of graphs generated by the CA model with the parameters $\epsilon = 0$ and $m = 2$ and the attachment function* (11). *Let the initial graph (V_1, E_1) be finite and satisfy condition* (6). *Then*

$$\Delta_n \to \infty \text{ a.s. as } \quad n \to \infty. \tag{14}$$

Let the sequence of sets $\tilde{E}_1, \tilde{E}_2, \ldots$ be such that for any $n \geqslant 1$, $\left(\tilde{V}_n, \tilde{E}_n\right)$ is a complete graph. In other words, $\tilde{E}_n$ is the set of all possible edges between nodes from $\tilde{V}_n$. Some edges from $\tilde{E}_n$ may not exist in the graph G_n. Corollary 1 gives a lower bound for the expectation $E\Delta_n$.

Corollary 1. *Under the assumptions of Theorem 1,*

$$E\Delta_n - \Delta_1 > \frac{\#\left(\tilde{E}_1 \cap E_1\right)}{3\#\tilde{E}_1} \ln(n-1), \quad n \geqslant 2. \tag{15}$$

5. COMPUTER SIMULATION

In this section, we consider simulation of the sequence Δ_n, $n \geqslant 1$. There is no need to simulate the sequence of graphs G_n, $n \geqslant 1$, to investigate the growth of the sample mean of Δ_n: it suffices to simulate the sequence of sets $\tilde{E}_n$, $n \geqslant 1$.

Let us interpret the unordered pairs of nodes from the set $\tilde{E}_n$ in terms of the generalized Pólya–Eggenberger urn model. For each $n \geqslant 1$, we divide the set $\tilde{E}_n$ into two subsets, $\tilde{E}_n \cap E_n$ and $\tilde{E}_n \setminus \left(\tilde{E}_n \cap E_n\right)$. Each unordered pair of nodes from $\tilde{E}_n \cap E_n$ is interpreted as a white ball whereas a pair of nodes belonging to the set $\tilde{E}_n \setminus \left(\tilde{E}_n \cap E_n\right)$ as a black ball. Hence, the urn (or the set $\tilde{E}_n$) contains $\#\left(\tilde{E}_n \cap E_n\right)$ white and $\#\left(\tilde{E}_n \setminus \left(\tilde{E}_n \cap E_n\right)\right)$ black balls.

At each step $n \geqslant 1$, a random ball is drawn uniformly from the urn (random sampling with replacement, see Proposition 1). The color of the ball is inspected and the urn is replenished according to the following rule. If a white ball is drawn, then two white balls and $\left((1/2)\left(1+\sqrt{1+8\#\tilde{E}_n}\right)-2\right)$ black balls are added in the urn. The cardinalities $\#\tilde{V}_n$ and $\#\tilde{E}_n$ are related by $\#\tilde{V}_n\left(\#\tilde{V}_n-1\right)=2\#\tilde{E}_n$, see (A.1); therefore, $\#\tilde{V}_n+1$ balls are added in the urn. If a black ball is drawn, the content of the urn remains unchanged.

Note that the number of added black and/or white balls is fixed in the classical Pólya–Eggenberger urn model; for example, see [10, p. 437]. In the current model, the number of added black balls depends on $\#\tilde{E}_n$. Other generalizations of the Pólya–Eggenberger urn model can be found in [12].

Since the pairs of nodes from $\tilde{E}_n$ are uniformly distributed (see Proposition 1), we provide a simple algorithm for simulating the sequence Δ_n, $n \geqslant 1$.

Algorithm 1.

1. The initial step. Using an initial graph G_1, calculate Δ_1, $\#\tilde{V}_1$, $\#\left(\tilde{E}_1 \cap E_1\right)$ and $\#\left(\tilde{E}_1 \setminus \left(\tilde{E}_1 \cap E_1\right)\right)$.

2. The evolutionary step. For any $n \geqslant 1$, simulate the value of the discrete random variable ξ_n uniformly distributed on the set $\{1,\ldots,\#\tilde{E}_n\}$. If $\xi_n > \#\left(\tilde{E}_n \cap E_n\right)$, then

$$\#\left(\tilde{E}_{n+1} \cap E_{n+1}\right) = \#\left(\tilde{E}_n \cap E_n\right),$$
$$\#\left(\tilde{E}_{n+1} \setminus \left(\tilde{E}_{n+1} \cap E_{n+1}\right)\right) = \#\left(\tilde{E}_n \setminus \left(\tilde{E}_n \cap E_n\right)\right).$$

If $\xi_n \leqslant \#\left(\tilde{E}_n \cap E_n\right)$, then

$$\#\left(\tilde{E}_{n+1} \cap E_{n+1}\right) = \#\left(\tilde{E}_n \cap E_n\right) + 2,$$
$$\#\left(\tilde{E}_{n+1} \setminus \left(\tilde{E}_{n+1} \cap E_{n+1}\right)\right) = \#\tilde{E}_{n+1} - \#\left(\tilde{E}_{n+1} \cap E_{n+1}\right),$$

where

$$\#\tilde{E}_{n+1} := \#\left(\tilde{E}_n \cap E_n\right) + \#\left(\tilde{E}_n \setminus \left(\tilde{E}_n \cap E_n\right)\right) + (1/2)\left(1+\sqrt{1+8\#\tilde{E}_{n+1}}\right).$$

In both cases,

$$\Delta_{n+1} = \Delta_1 - \#\tilde{V}_1 + (1/2)\left(1+\sqrt{1+8\#\tilde{E}_{n+1}}\right).$$

Note that when implementing this algorithm, one needs to store the four values: Δ_1, $\#\tilde{V}_1$, $\#\left(\tilde{E}_n \cap E_n\right)$, and $\#\left(\tilde{E}_n \setminus \left(\tilde{E}_n \cap E_n\right)\right)$.

For each initial graph G_1, 100 independent sequences $\Delta_1^{(j)},\ldots,\Delta_N^{(j)}$ of length $N=10^6$ were simulated; the upper index (j) denotes the sequence number. The following complete graphs were selected as the initial ones: 1) the triangle $G_{1,1}$, 2) the complete graph $G_{1,2}$ with 17 vertices, and 3) the complete graph $G_{1,3}$ with 51 vertices.

The sequence of the sample means $\bar{\Delta}_n(G_{1,\ell})$, $1 \leqslant n \leqslant N$, was calculated by the formula

$$\bar{\Delta}_n(G_{1,\ell}) = \frac{1}{100}\sum_{j=1}^{100}\Delta_n^{(j)}$$

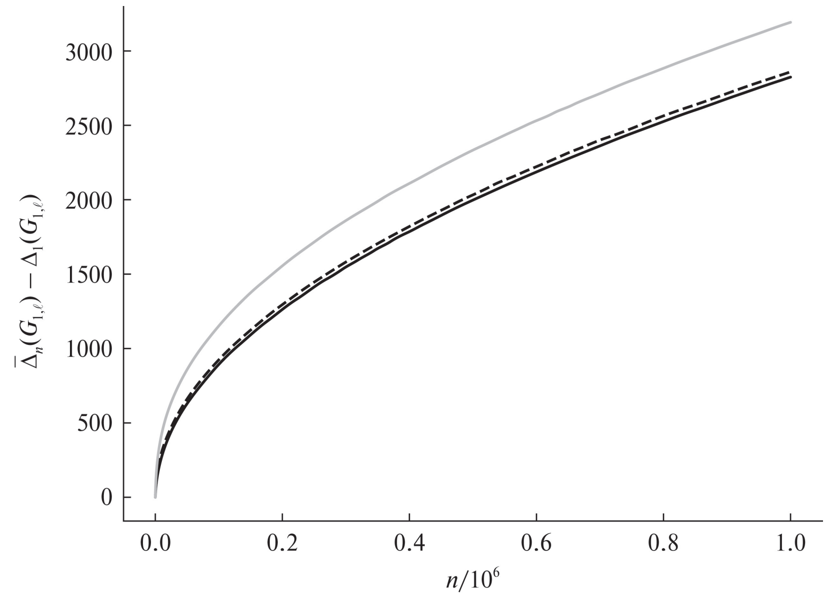

Fig. 2. The plots of $\left\{\left(n, \bar{\Delta}_n(G_{1,\ell}) - \Delta_1(G_{1,\ell})\right),\ 0 \leqslant n \leqslant 10^6\right\}$, where the curves with $\ell \in \{1,2,3\}$ are shown in black, dotted black, and light grey.

for each initial graph $G_{1,\ell}$. Figure 2 shows the plots of $\{(n, \bar{\Delta}_n(G_{1,\ell}) - \Delta_1(G_{1,\ell})), 1 \leqslant n \leqslant N\}$, where $\ell \in \{1,2,3\}$ and $N = 10^6$. The calculations were carried out in 152.88 s.

The least squares method was applied to approximate the discrete data

$$\left\{\left(n, \bar{\Delta}_n(G_{1,\ell}) - \Delta_1(G_{1,\ell})\right),\ 2 \times 10^5 \leqslant n \leqslant N\right\},$$
$$\left\{\left(n, \bar{\Delta}_n(G_{1,\ell}) - \Delta_1(G_{1,\ell})\right),\ 5 \times 10^5 \leqslant n \leqslant N\right\}$$

by the function $\varphi(n) = c_1 + c_2 n^{c_3}$. The estimates of c_1, c_2, and c_3 are presented in the table below.

The estimates of the parameters c_1, c_2, and c_3

$2 \times 10^5 \leqslant n \leqslant 10^6$			
Initial graph	$\hat{c}_1$	$\hat{c}_2$	$\hat{c}_3$
$G_{0,1}$	1	2.827	0.499
$G_{0,2}$	680	3.179	0.492
$G_{0,3}$	20 825	6.564	0.447

$5 \times 10^5 \leqslant n \leqslant 10^6$			
Initial graph	$\hat{c}_1$	$\hat{c}_2$	$\hat{c}_3$
$G_{0,1}$	1	2.831	0.499
$G_{0,2}$	680	3.234	0.491
$G_{0,3}$	20 825	6.133	0.452

Based on the simulation, we have the following findings of this study.

1. The simulation results in Fig. 2 are sufficiently close to the theoretical ones. Indeed, $E\Delta_n$ and the empirical mean $\bar{\Delta}_n$ grow with increasing n; see Corollary 1 and Fig. 2.
2. The value $\bar{\Delta}_n$ grows as $C\sqrt{n}$, where $2 \times 10^5 \leqslant n \leqslant 10^6$; see the table. This simulation result does not contradict Corollary 1. Note that this simulation result does not imply the same (or close) growth rate of $\bar{\Delta}_n$ for $n > 10^6$.
3. Due to the limited computer memory, difficulties arose when simulating the triangle count Δ_n for $n > 10^6$.
4. The plots of $(n, \bar{\Delta}_n(G_{1,1}) - \Delta_1(G_{1,1}))$ and $(n, \bar{\Delta}_n(G_{1,2}) - \Delta_1(G_{1,2}))$ are close to each other, whereas the plot of $(n, \bar{\Delta}_n(G_{1,3}) - \Delta_1(G_{1,3}))$ has a shift along the vertical axis relative to the other two plots for $0 \leqslant n \leqslant 10^6$, which is perhaps due to the size of the graph $G_{1,3}$. It can be hypothesized that the influence of a relatively large initial graph should be eliminated with a large n.

6. CONCLUSIONS

By analogy with [9], where a rather wide definition of PA evolution models was given, this paper has introduced a new generalized CA model. The class of attachment functions covered by Definition 1 is wide enough to simulate the evolution of many random systems and networks in which newly appending nodes choose existing nodes not with a high number of links but on the basis of their belonging to strongly connected groups of nodes.

The theoretical properties of the limit CA model have been investigated. In this model, new nodes connect to two existing nodes by successive choice (random sampling without replacement) with probabilities inversely proportional to the number of nodes involved in triangles. The new model can be treated as the limit of the original CA model [4] as the parameter $\alpha \to 0$ in the case $\epsilon = 0$ and the attachment function (11). Due to the specifics of the weights (3), the cases $\epsilon > 0$ and $\epsilon = 0$ must be studied separately.

The main results of the paper concern the case $\epsilon = 0$. According to Theorem 1, the total triangle count Δ_n tends to infinity a.s. as the evolution step $n \to \infty$. The initial graph must contain at least m nodes as part of triangles to start the evolution. Corollary 1 shows that the growth rate of the expected triangle count $E\Delta_n$ is higher than the logarithmic one for an evolution step $n \geqslant 2$. The simulation results do not contradict the theoretical considerations. Refining the upper and lower bounds for $E\Delta_n$ will be the aim of further research.

ACKNOWLEDGMENTS

N.M. Markovich acknowledges the support of the Russian Science Foundation, project no. 24-21-00183. The authors are grateful to M.S. Ryzhov for computer simulation and the reviewers for careful reading of the manuscript and helpful remarks.

APPENDIX A

Proof of Proposition 1. Let $\{i,j\} \in \tilde{E}_n$. Consider for each node $i \in \tilde{V}_n$ the random event $A_{n,i} \equiv \{\text{node } i \text{ is chosen}\}$. By the total probability law,

$$\begin{aligned} P\left(W_n = \{i,j\}\right) &= P(A_{n,j}|A_{n,i})P(A_{n,i}) + P(A_{n,i}|A_{n,j})P(A_{n,j}) \\ &= \frac{2}{\#\tilde{V}_n\left(\#\tilde{V}_n - 1\right)} = \frac{1}{\#\tilde{E}_n}. \end{aligned} \tag{A.1}$$

Using (A.1), we obtain

$$P\left(W_n \in \tilde{E}_n\right) = \sum_{\{i_1,i_2\} \in \tilde{E}_n} P\left(W_n = \{i_1, i_2\}\right) = 1.$$

Hence, it follows that $P\left(W_n \in \mathcal{E}_n \setminus \tilde{E}_n\right) = 0$, and consequently, $P\left(W_n = \{i,j\}\right) = 0$ for any pair of nodes $\{i,j\} \in \mathcal{E}_n \setminus \tilde{E}_n$.

APPENDIX B

We represent the set of edges E_n, $n \geqslant 1$, as the set of unordered pairs of nodes. For any $n \geqslant 1$ we divide the set $\tilde{E}_n$ into the subsets $\tilde{E}_n \cap E_n$ and $\tilde{E}_n \setminus \left(\tilde{E}_n \cap E_n\right)$. The subset $\tilde{E}_n \cap E_n$ consists of the unordered pairs of nodes connected via edges in the graph G_n. Let $B_n = \{W_n \in \tilde{E}_n \cap E_n\}$, $n \geqslant 1$, be a random event that an unordered pair of nodes W_n chosen by successive sampling is connected via an edge from E_n.

We denote by B^c the event complementary to B. Let $n, k \in \mathbb{N}$ and $\ell \in \mathbb{N} \cup \{0\}$ be such that $n \geqslant 2$, $1 \leqslant k < n$, and $0 \leqslant \ell \leqslant n-k$. The event $(B_k \cap \cdots \cap B_{n-1})_\ell$ is the union of $(n-k)!/(\ell!(n-k-\ell)!)$ nonintersecting random events such that ℓ events B_j are replaced by their complementary ones in the intersection $B_k \cap \cdots \cap B_{n-1}$. If the random event $(B_k \cap \cdots \cap B_{n-1})_\ell$ occurs, then ℓ pairs of nodes among $W_k, \ldots, W_{n-1}$ are not connected, whereas the remaining $(n-k-\ell)$ pairs are connected. For example, for $k=1$ and $n=4$ we have

$$\begin{aligned}
&\ell = 0: \ (B_1 \cap B_2 \cap B_3)_0 = B_1 \cap B_2 \cap B_3,\\
&\ell = 1: \ (B_1 \cap B_2 \cap B_3)_1 = (B_1 \cap B_2 \cap B_3^c) \cup (B_1 \cap B_2^c \cap B_3) \cup (B_1^c \cap B_2 \cap B_3),\\
&\ell = 2: \ (B_1 \cap B_2 \cap B_3)_2 = (B_1 \cap B_2^c \cap B_3^c) \cup (B_1^c \cap B_2 \cap B_3^c) \cup (B_1^c \cap B_2^c \cap B_3),\\
&\ell = 3: \ (B_1 \cap B_2 \cap B_3)_3 = B_1^c \cap B_2^c \cap B_3^c.
\end{aligned}$$

By Proposition 1,

$$P(B_n) = \sum_{\{i_1, i_2\} \in \tilde{E}_n \cap E_n} \frac{1}{\#\tilde{E}_n} = \frac{\#\left(\tilde{E}_n \cap E_n\right)}{\#\tilde{E}_n}. \tag{B.1}$$

Let $n \geqslant k-1$ and $k \geqslant 1$. If the random event $(B_k \cap \cdots \cap B_{n-1})_\ell$ occurs, then

$$\begin{aligned}
\#\left(\tilde{E}_n \cap E_n\right) &= \#\left(\tilde{E}_k \cap E_k\right) + 2(n-\ell-k),\\
\#\tilde{E}_n &= \#\tilde{E}_k + (n-\ell-k)\#\tilde{V}_k + (1/2)(n-\ell-k)(n-\ell-k-1).
\end{aligned}$$

In this case, from (B.1) it follows that

$$P\left(B_n \middle| (B_k \cap \cdots \cap B_{n-1})_\ell\right) = p_{n-\ell}(G_k),$$

where

$$p_n(G_k) = \frac{\#\left(\tilde{E}_k \cap E_k\right) + 2(n-k)}{\#\tilde{E}_k + (n-k)\#\tilde{V}_k + (1/2)(n-k)(n-k-1)}. \tag{B.2}$$

Now we establish the inequalities that will be used for proving Theorem 1.

Lemma 1. *Under the assumptions of Theorem 1,*

$$p_n(G_k) > 0, \tag{B.3}$$

$$p_n(G_k) > p_{n+1}(G_k) \tag{B.4}$$

for any natural numbers $k \leqslant n$.

Proof of Lemma 1. By Proposition (6), $\tilde{E}_k \neq \emptyset$ and $\tilde{E}_k \cap E_k \neq \emptyset$ for any natural number k. Therefore, (B.2) implies (B.3).

Next, based on (B.2), inequality (B.4) is reduced to

$$\#\tilde{V}_k \left\{\#\left(\tilde{E}_k \cap E_k\right) - \#\tilde{V}_k + 1\right\} + (n-k)\#\left(\tilde{E}_k \cap E_k\right) + (n-k)(n-k+1) > 0.$$

It suffices to show that, for any initial graph satisfying (6),

$$\#\left(\tilde{E}_k \cap E_k\right) > \#\tilde{V}_k - 1. \tag{B.5}$$

We will prove (B.5) by induction. Consider the graph $\left(\tilde{V}_1, \tilde{E}_1 \cap E_1\right)$. The minimum number of edges in $\tilde{E}_1 \cap E_1$ depends on $\#\tilde{V}_1$. If $\#\tilde{V}_1 = 3$, then the graph with the minimum number of edges is a triangle for which $\#\left(\tilde{E}_1 \cap E_1\right) = 3$. Hence, (B.5) holds for $k=1$ in the case $\#\tilde{V}_1 = 3$.

If $\#\tilde{V}_1 > 3$, then the graph with the minimum number of edges consists of $\#\tilde{V}_1 - 2$ triangles with one common side (see Example 1 for $\#\tilde{V}_1 = 4$). For such graphs, we have $\#\left(\tilde{E}_1 \cap E_1\right) = \#\tilde{V}_1 + 1$. This means that (B.5) with $k = 1$ and $\#\tilde{V}_1 > 3$ is valid as well.

Now we make the inductive hypothesis that (B.5) is true for k. Under this hypothesis, it is necessary to show (B.5) for $(k+1)$. If the random event B_k occurs, then $\#\left(\tilde{E}_{k+1} \cap E_{k+1}\right) = \#\left(\tilde{E}_k \cap E_k\right) + 2$ and $\#\tilde{V}_{k+1} = \#\tilde{V}_k + 1$. Using (B.5), we obtain

$$\#\left(\tilde{E}_{k+1} \cap E_{k+1}\right) - \#\tilde{V}_{k+1} + 1 = \left\{\#\left(\tilde{E}_k \cap E_k\right) - \#\tilde{V}_k + 1\right\} + 1 > 0.$$

The case in which the event B_k does not occur can be considered by analogy.

Lemma 2. *Under the assumptions of Theorem 1,*

$$P\left(B_n\right) > P\left(B_n | B_1 \cap \cdots \cap B_{n-1}\right), \quad n = 2, 3, \ldots. \tag{B.6}$$

Proof of Lemma 2. By the total probability law,

$$P(B_n) = \sum_{\ell=0}^{n-1} P\left(B_n | \left(B_1 \cap \cdots \cap B_{n-1}\right)_\ell\right) P\left(\left(B_1 \cap \cdots \cap B_{n-1}\right)_\ell\right).$$

From Lemma 1 it follows that

$$P\left(B_n | B_1 \cap B_2 \cap \cdots \cap B_{n-1}\right) < P\left(B_n | \left(B_1 \cap B_2 \cap \cdots \cap B_{n-1}\right)_\ell\right)$$

for any $1 \leqslant \ell \leqslant n-1$. With this inequality, we obtain

$$P(B_n) > P\left(B_n | B_1 \cap \cdots \cap B_{n-1}\right) \sum_{\ell=0}^{n-1} P\left(\left(B_1 \cap \cdots \cap B_{n-1}\right)_\ell\right).$$

Since the last sum is 1, we finally arrive at (B.6).

Theorem 1 is proved by verifying the conditions of the Borel–Cantelli lemma; see below.

Lemma 3 [10, p. 79]. *Let random events $C_1, C_2, \ldots$ defined on the same probability space satisfy the correlation condition: for any natural numbers u and v such that $u \neq v$, the random events C_u and C_v are negative correlated or uncorrelated. If $\sum_{n=1}^{\infty} P(C_n) = \infty$, then $P\{C_n$ i.m.$\} = 1$.*[3]

Lemma 4. *Under the assumptions of Theorem 1, random events $B_1, B_2, \ldots$ satisfy the correlation condition:*

$$P\left(B_u \cap B_v\right) \leqslant P\left(B_u\right) P\left(B_v\right). \tag{B.7}$$

Proof of Lemma 4. If $\min\{u, v\} = 1$, then assumption (6) implies $0 < \#\left(\tilde{E}_1 \cap E_1\right) \leqslant \#\left(\tilde{E}_1\right)$, see the proof of Lemma 1. In this case, from (B.1) it follows that $P(B_1) > 0$. If $\min\{u, v\} > 1$, then combining (B.3) and (B.6) gives $P\left(B_{\min\{u,v\}}\right) > p_{\min\{u,v\}}(G_1) > 0$. Thus, we have proved that $P\left(B_{\min\{u,v\}}\right) > 0$ for any natural numbers u and v. With the conditional probability formula, the correlation condition (B.7) can be written as

$$P\left(B_{\max\{u,v\}}\right) - P\left(B_{\max\{u,v\}} | B_{\min\{u,v\}}\right) \geqslant 0. \tag{B.8}$$

It suffices to show inequality (B.8) for $\max\{u, v\} = n + k$ and $\min\{u, v\} = n$, where n and k are any natural numbers.

[3] Here, $\{C_n$ i.m.$\}$ denotes the event consisting in the occurrence of infinitely many events from $C_1, C_2, \ldots$ [11].

If $k = 1$, then applying the total probability formula together with (B.4) yields

$$P(B_{n+1}) - P(B_{n+1}|B_n) = (1 - p_1(G_n))\,(p_1(G_n) - p_2(G_n)) \geqslant 0.$$

Here the equality holds for $\#\left(\tilde{E}_n \cap E_n\right) = \#\left(\tilde{E}_n\right)$.

Now let $k \geqslant 2$. We state that

$$P(B_{n+k}) = J + \sum_{\ell=0}^{k-1} p_{n+k-\ell-1}(G_n) P(B_n^c \cap (B_{n+1} \cap \cdots \cap B_{n+k-1})_\ell), \tag{B.9}$$

$$P(B_{n+k}|B_n) = J + \sum_{\ell=0}^{k-1} p_{n+k-\ell}(G_n) P(B_n^c \cap (B_{n+1} \cap \cdots \cap B_{n+k-1})_\ell), \tag{B.10}$$

where

$$J = \sum_{\ell=0}^{k-1} p_{n+k-\ell}(G_n) P\left(B_n \cap (B_{n+1} \cap \cdots \cap B_{n+k-1})_\ell\right).$$

From (B.9) and (B.10) it immediately follows that the difference $P\left(B_{n+k}\right) - P\left(B_{n+k}|B_n\right)$ is

$$\begin{aligned}
&\sum_{\ell=0}^{k-1} \left\{p_{n+k-\ell-1}(G_n) - p_{n+k-\ell}(G_n)\right\} P\left(B_n^c \cap (B_{n+1} \cap \cdots \cap B_{n+k-1})_\ell\right) \\
&= \sum_{\ell=0}^{k-1} \left\{p_{n+k-\ell-1}(G_n) - p_{n+k-\ell}(G_n)\right\} \left\{1 - p_{n-\ell}(G_n)\right\} P\left((B_{n+1} \cap \cdots \cap B_{n+k-1})_\ell\right).
\end{aligned}$$

Due to (B.3) and (B.4), the latter sum is positive.

Let us prove (B.9). Applying the total probability formula gives

$$\begin{aligned}
P\left(B_{n+k}\right) &= P\left(B_{n+k}|\left(B_n \cap \cdots \cap B_{n+k-1}\right)_0\right) P\left(\left(B_n \cap \cdots \cap B_{n+k-1}\right)_0\right) \\
&+ \sum_{\ell=1}^{k-1} P\left(B_{n+k}|\left(B_n \cap \cdots \cap B_{n+k-1}\right)_\ell\right) P\left(\left(B_n \cap \cdots \cap B_{n+k-1}\right)_\ell\right) \\
&+ P\left(B_{n+k}|\left(B_n \cap \cdots \cap B_{n+k-1}\right)_k\right) P\left(\left(B_n \cap \cdots \cap B_{n+k-1}\right)_k\right).
\end{aligned}$$

Using the identity

$$\begin{aligned}
\left(B_n \cap \cdots \cap B_{n+k-1}\right)_\ell = &\left\{B_n \cap \left(B_{n+1} \cap \cdots \cap B_{n+k-1}\right)_\ell\right\} \\
&\cup \left\{B_n^c \cap \left(B_{n+1} \cap \cdots \cap B_{n+k-1}\right)_{\ell-1}\right\},
\end{aligned}$$

we obtain

$$\begin{aligned}
P\left(B_{n+k}\right) = &\Bigg\{P\left(B_{n+k}|B_n \cap \cdots \cap B_{n+k-1}\right) P\left(B_n \cap \cdots \cap B_{n+k-1}\right) \\
&+ \sum_{\ell=1}^{k-1} P\left(B_{n+k}|\left(B_n \cap \cdots \cap B_{n+k-1}\right)_\ell\right) P\left(B_n \cap \left(B_{n+1} \cdots \cap B_{n+k-1}\right)_\ell\right)\Bigg\} \\
&+ \sum_{\ell=1}^{k-1} P\left(B_{n+k}|\left(B_n \cap \cdots \cap B_{n+k-1}\right)_\ell\right) P\left(B_n^c \cap \left(B_{n+1} \cap \cdots \cap B_{n+k-1}\right)_{\ell-1}\right) \\
&+ P\left(B_{n+k}|B_n^c \cap \cdots \cap B_{n+k-1}^c\right) P\left(B_n^c \cap \cdots \cap B_{n+k-1}^c\right).
\end{aligned}$$

In view of (B.2), the sum in curly brackets is J. Replacing the summation variable with $s = \ell - 1$, we obtain

$$\sum_{\ell=1}^{k-1} P\left(B_{n+k} | \left(B_n \cap \cdots \cap B_{n+k-1}\right)_\ell\right) P\left(B_n^c \cap \left(B_{n+1} \cdots \cap B_{n+k-1}\right)_{\ell-1}\right)$$
$$= \sum_{s=0}^{k-2} P\left(B_{n+k} | \left(B_n \cap \cdots \cap B_{n+k-1}\right)_{s+1}\right) P\left(B_n^c \cap \left(B_{n+1} \cdots \cap B_{n+k-1}\right)_s\right)$$
$$= \sum_{s=0}^{k-1} P\left(B_{n+k} | \left(B_n \cap \cdots \cap B_{n+k-1}\right)_{s+1}\right) P\left(B_n^c \cap \left(B_{n+1} \cdots \cap B_{n+k-1}\right)_s\right)$$
$$-P\left(B_{n+k} | \left(B_n \cap \cdots \cap B_{n+k-1}\right)_k\right) P\left(B_n^c \cap \left(B_{n+1} \cdots \cap B_{n+k-1}\right)_{k-1}\right).$$

Using (B.2) once more, we finally arrive at (B.9).

Identity (B.10) can be verified similarly. To this end, it suffices to apply the total probability formula and the identity

$$\left(B_{n+1} \cap \cdots \cap B_{n+k-1}\right)_\ell = \left\{B_n \cap \left(B_{n+1} \cap \cdots \cap B_{n+k-1}\right)_\ell\right\}$$
$$\cup \left\{B_n^c \cap \left(B_{n+1} \cap \cdots \cap B_{n+k-1}\right)_\ell\right\}.$$

Proof of Theorem 1. The difference of the total triangle counts in the graphs G_n and G_1, as well as the difference of the numbers of nodes involved in the triangles, can be expressed through the indicators of random events B_j as follows:

$$\Delta_n - \Delta_1 = \sum_{j=1}^{n-1} I\left\{B_j\right\}, \tag{B.11}$$

$$\#\tilde{V}_n - \#\tilde{V}_1 = \sum_{j=1}^{n-1} I\left\{B_j\right\}, \quad n \geqslant 2, \tag{B.12}$$

where $I\{\cdot\}$ denotes the indicator of a corresponding event. From (B.11) and (B.12) it follows that the triangle count Δ_n is a linear function of the cardinality of the set $\tilde{V}_n$:

$$\Delta_n = \#\tilde{V}_n + \left(\Delta_1 - \#\tilde{V}_1\right), \quad n \geqslant 2. \tag{B.13}$$

Due to (B.11), the convergence (14) is immediate from

$$\sum_{j=1}^{n-1} I\left\{B_j\right\} \to \infty \text{ a.s. as } \quad n \to \infty,$$

which is equivalent to

$$P(B_n \text{ i.m.}) = 1.$$

Finally, we prove that the sequence $B_1, B_2, \ldots$ satisfies the assumptions of the Borel–Cantelli lemma (see Lemma 3). One of them has been verified in Lemma 4. It suffices to establish

$$\sum_{n=1}^{\infty} P\left(B_n\right) = \infty. \tag{B.14}$$

By Lemmas 1 and 2,

$$\sum_{n=1}^{\infty} P(B_n) > \sum_{n=3}^{\infty} p_n(G_1).$$

Note that (B.2) implies

$$p_n(G_1) > \frac{\#\left(\tilde{E}_1 \cap E_1\right) + 2(n-1)}{n\#\tilde{E}_1 + (1/2)(n-1)(n-2)}. \tag{B.15}$$

The inequality $\#\left(\tilde{E}_1 \cap E_1\right) \leqslant \#\tilde{E}_1$ can be used to show that the right-hand side of (B.15) exceeds $\#\left(\tilde{E}_1 \cap E_1\right) / \left(\left(3\#\tilde{E}_1\right)(n-2)\right)$ for $n \geqslant 3$. Hence,

$$\sum_{n=1}^{\infty} P(B_n) > \frac{\#\left(\tilde{E}_1 \cap E_1\right)}{3\#\tilde{E}_1} \lim_{N\to\infty} H_N,$$

where $H_N = \sum_{n=1}^{N} 1/n$. The relation (B.14) follows by applying the inequality $H_N > \ln(N)$ and the limit relation $\lim_{N\to\infty} \ln(N) = \infty$.

To establish Corollary 1, we note that (B.11) implies

$$E(\Delta_n) - \Delta_1 = \sum_{j=1}^{n-1} P(B_j) > \sum_{j=3}^{n+1} p_j(G_1), \quad n \geqslant 2.$$

From the proof of (B.14) it follows that the latter sum is greater than $(\#(\tilde{E}_1 \cap E_1)/(3\#\tilde{E}_1)) \ln(n-1)$. Therefore, we arrive at the desired inequality (15).

REFERENCES

1. Van der Hofstad, R., *Random Graphs and Complex Networks*, Cambridge: Cambridge University Press, 2017.
2. Bollobas, B., *Random Graphs*, 2nd. ed., Cambridge: Cambridge University Press, 2001.
3. Newman, M.E.J., *Networks*, Oxford: Oxford University Press, 2018.
4. Bagrow, J.P. and Brockmann, D., Natural Emergence of Clusters and Bursts in Network Evolution, *Phys. Rev. X*, 2013, no. 3, p. 021016.
5. Newman, M.E.J., Random Graphs with Clustering, *Phys. Rev. Lett.*, 2009, vol. 103, no. 5, p. 058701.
6. Liu, Q. and Dong, Z., Limit Laws for the Number of Triangles in the Generalized Random Graphs with Random Node Weights, *Stat. Probab. Lett.*, 2020, vol. 161, art. no. 108733.
7. Bobkov, S.G., Danshina, M.A., and Ulyanov, V.V., Rate of Convergence to the Poisson Law of the Numbers of Cycles in the Generalized Random Graphs, in *Operator Theory and Harmonic Analysis. Springer Proceedings in Mathematics and Statistics*, Karapetyants, A.N., Pavlov, I.V., and Shiryaev, A.N., Eds., Springer: Cham, 2021, vol. 358, pp. 109–133.
8. Garavaglia, A. and Stegehuis, C., Subgraphs in Preferential Attachment Models, *Advances in Applied Probability*, 2019, vol. 51, pp. 898–926.
9. Wang, T. and Resnick, S.I., Consistency of Hill Estimators in a Linear Preferential Attachment Model, *Extremes*, 2019, vol. 22, pp. 1–28.
10. Fristedt, B. and Gray, L., *A Modern Approach to Probability Theory*, Boston: Birkhäuser, 1997.
11. Shiryaev, A.N., *Probability-1*, New York: Springer, 2016.
12. Chen, M.-R. and Kuba, M., On Generalized Pólya Urn Models, *J. Appl. Probab.*, 2013, vol. 50, pp. 1169–1186.

This paper was recommended for publication by A.I. Lyakhov, a member of the Editorial Board