

Стохастические системы

© 2024 г. М. ВАЙЧЮЛИС, д-р мат. наук (marijus.vaiciulis@mif.vu.lt)
(Институт науки данных и цифровых технологий
Вильнюсского университета, Литва),
Н.М. МАРКОВИЧ, д-р физ.-мат. наук (nat.markovich@gmail.com)
(Институт проблем управления им. В.А. Трапезникова РАН, Москва)

ИССЛЕДОВАНИЕ КОЛИЧЕСТВА ТРЕУГОЛЬНИКОВ В ГРАФАХ, ЭВОЛЮЦИОНИРУЮЩИХ КЛАСТЕРНЫМ ПРИСОЕДИНЕНИЕМ¹

Модель кластерного присоединения (КП), предложенная Багроу и Брокманном (2013 г.), может быть использована как инструмент эволюции ненаправленных случайных сетей. В статье вводится обобщенное определение модели КП. Теоретические результаты получены для новой модели КП, которую можно рассматривать как предел прежней, когда параметр модели α стремится к нулю, а параметр $\epsilon = 0$. Предметом исследования является количество треугольников связанных узлов в графе на шаге эволюции n – важная характеристика кластеризации сети. Доказано, что количество треугольников стремится к бесконечности с вероятностью единица для предложенной модели эволюции при $n \rightarrow \infty$, а скорость роста среднего количества треугольников $E\Delta_n$ на шаге эволюции $n \geq 2$ выше логарифмической. Компьютерное моделирование использовано для моделирования последовательностей количества треугольников. Данное моделирование основано на обобщенной модели урн Пойа–Эггенбергера, что предложено впервые.

Ключевые слова: кластерное присоединение, коэффициент кластеризации, вес узла, случайный граф, эволюция, модель урн.

DOI: 10.31857/S0005231024110034, **EDN:** YMDKBI

1. Введение

Случайные графы используются для моделирования реальных сетей (real-world networks) [1]. Это графы, в которых структура соединения узлов случайна ввиду динамики (эволюции) сетей во времени. Случайные графы могут быть описаны вероятностным распределением или случайным процессом, которые их генерируют [1–3]. Примером случайного графа служит граф Эрдеша–Реньи, где ребро между любыми двумя различными вершинами графа появляется с некоторой фиксированной вероятностью p , независимо от других пар вершин.

¹ Исследование Маркович Н.М. выполнено за счет Российского научного фонда (грант № 24-21-00183). Авторы выражают благодарность М.С. Рыжову за проведенное компьютерное моделирование, а также анонимным рецензентам за ценные замечания, позволившие существенно улучшить статью.

Статья посвящена кластерному присоединению (КП), предложенному в [4], которое может быть использовано в качестве модели эволюции ненаправленных локальных случайных сетей, когда узлы принадлежат какому-то сообществу. Идея КП в том, что вновь присоединенный новый узел выбирает существующие узлы не с большим числом связей, а по принципу их принадлежности к сильно связанным группам узлов (что характерно, в частности, для социального поведения людей) и присоединяется к этим узлам новыми ребрами. КП предназначен для моделирования локальных сетей и не ведет к появлению узлов с большим числом связей, в отличие от предпочтительного присоединения (ПП), которое реализует так называемый принцип «*богатый-становится-богаче*» («rich-get-richer»). Новый узел, согласно КП, присоединяется к $m \geq 2$ существующим узлам. Например, в социальных сетях индивиды могут иметь друзей не среди популярных людей, а в своем окружении, формируя тесные сообщества. КП может соответствовать модели управления развитием региона с отсутствием больших мегаполисов или модели общества с примерно равным распределением благ, при котором обогащение отдельных узлов невозможно. КП может быть использовано в медицине при моделировании эволюции нейронных сетей мозга, в биологии при описании коммуникаций в стаях животных [4].

В [4] предлагается модель присоединения нового узла к существующему узлу i с вероятностью, пропорциональной его коэффициенту кластеризации $c_{i,n}$ на шаге эволюции n :

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

где коэффициент кластеризации узла $i \in V_n$ определяется как

$$c_{i,n} = \begin{cases} 0, & D_{i,n} = 0 \text{ или } D_{i,n} = 1, \\ 2\Delta_{i,n} / (D_{i,n}(D_{i,n} - 1)), & D_{i,n} \geq 2, \end{cases}$$

$D_{i,n}$ – степень узла i (число его связей с другими узлами), $\Delta_{i,n}$ – количество треугольников, к которым он принадлежит, $0 \leq \epsilon \leq 1$, $\alpha \geq 0$ – параметры модели.² КП нового узла к существующим узлам, как правило, снижает коэффициент кластеризации узла [4]. Несмотря на то, что КП ведет к многочисленным новым феноменам таким, как формирование сообщества вокруг новых узлов, всплески модулярности и распределения с легкими хвостами для количества связей узлов, см. [4], КП не получил дальнейшего развития. Многие утверждения в [4] не были строго доказаны, как, например, подтверждение легких хвостов распределения количества связей узлов, исследование кластерной структуры модулярности, что имеет большее значение для анализа связности узлов в сети.

Цель статьи – теоретическое исследование более общей, чем в [4], модели КП для случая, когда один из ее параметров $\alpha \rightarrow 0$, а другой параметр $\epsilon = 0$.

² $x \propto y$ означает, что существует ненулевая константа C такая, что $x = Cy$.

Эта модель сводится к последовательному выбору $m \geq 2$ существующих узлов, входящих в треугольники, вновь присоединенным новым узлом на каждом шаге эволюции. При этом количество узлов и ребер в графе неслучайны, а случайность привносится выбором существующего узла с вероятностью, пропорциональной его весу.

Треугольник соединенных узлов – наиболее изучаемый подграф, который может быть рассмотрен как базовое сообщество. Возможно, впервые он был предложен к изучению в [5]. Количество треугольников также задействовано в вычислении коэффициента кластеризации. Этот коэффициент является важной характеристикой кластерной структуры случайных графов, так как он выражает долю соседних узлов узла, соединенных ребрами. Предельное поведение количества треугольников узлов привлекает внимание многих ученых, см., например, [6–8]. Цель работы – исследовать асимптотическое поведение общего количества треугольников в графе Δ_n при $n \rightarrow \infty$ и получить оценку снизу для математического ожидания $E\Delta_n$ для предлагаемой модели КП, что является новым результатом.

Структура статьи следующая. Описание модели приводится в разделе 2. В разделе 3 обсуждается последовательный выбор существующих узлов новым узлом. Основные результаты содержатся в разделе 4. В разделе 5 приводятся результаты компьютерного моделирования. Раздел 6 содержит выводы. Доказательства приведены в Приложениях.

2. Описание предельной модели

Пусть $G_n = (V_n, E_n)$, $n = 1, 2, \dots$ – последовательность случайных графов, где V_n – множество узлов, E_n – множество ребер, а множество номеров графов может быть интерпретировано как дискретное время, поскольку графы генерируются в ходе эволюции. Эволюция начинается с исходного графа G_1 , в котором число узлов и ребер фиксировано. Применим обозначения из [1]. Обозначим через $\#A$ мощность (число элементов) произвольного конечного множества A . Для графа G_n имеем

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

где натуральное число $m \geq 2$ – параметр модели КП.

Приведем более общее, чем в [4], определение модели КП.

Определение 1. Граф G_{n+1} образуется из G_n по правилу из двух частей:

а) *детерминированная часть:* новый узел $\#V_1 + n \in V_{n+1} \setminus V_n$ добавляется к G_n ;

б) *стохастическая часть:* каждый узел $i \in V_n$ снабжается весом

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

где $f : [0, 1] \rightarrow [0, \infty)$ – детерминированная и неубывающая функция присоединения, удовлетворяющая условию

$$(4) \quad f(0) = 0,$$

$\epsilon \geq 0$ – параметр модели КП. Существующие узлы $i_1, \dots, i_m \in V_n$ выбираются последовательным выбором³ с вероятностями, пропорциональными их весам, и каждый из них присоединяется только одним ребром к новому добавленному узлу $\#V_1 + n$ на каждом шаге эволюции.

По определению 1 параметр ϵ не обязательно ограничен сверху единицей в силу нормализации в формуле (3) и этим отличается от модели КП в [4]. Чтобы исключить случай отсутствия эволюции $G_n = G_1$, $n \geq 2$ из рассмотрения, предположим, что начальный граф G_1 удовлетворяет условию:

$$(5) \quad \#V_1 \geq m$$

при $\epsilon > 0$, и

$$(6) \quad \#\tilde{V}_1 \geq m$$

при $\epsilon = 0$. Здесь $\tilde{V}_n = \{i \in V_n : c_{i,n} > 0\}$, $n \geq 1$. Из (2) и (5) следует $V_n \geq m$, а при $\epsilon = 0$ из (6) следует $\#\tilde{V}_n \geq m$ для каждого $n \in \mathbb{N}$. Отсюда для $\epsilon \geq 0$ и $\forall n \in \mathbb{N}$ существует хотя бы один набор узлов $i_1, \dots, i_m \in V_n$, который может быть выбран с положительной вероятностью методом последовательного выбора.

Заменив в (3) $c_{i,n}$ на $D_{i,n}$ и область определения функции $f(x)$ на $\{0\} \cup \mathbb{N}$, получим определение модели ПП, см. стр. 5 в [9]. Подставив

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

в (3), получим веса (1) для введенной в [4] модели КП.

Учитывая (4), веса (3) могут быть переписаны как

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

Величина ϵ определяет, из какого множества узлов выбираются узлы i_1, \dots, i_m последовательным выбором: это множество \tilde{V}_n при $\epsilon = 0$ и множество V_n при $\epsilon > 0$. При $\epsilon > 0$, используя неравенства

$$(9) \quad \#V_n \geq \#\tilde{V}_n, \quad n \geq 1,$$

из (8) получим неравенства для весов

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

³ Описание последовательного выбора узлов при $m = 2$ дано в разделе 3.

где $C_{f,\epsilon} = (f(1) + \epsilon)/\epsilon$. Из (10) следует, что $p_{i,n} \rightarrow 0$ почти наверное при $n \rightarrow \infty$. В случае $\epsilon = 0$ поведение весов $p_{i,n}$, $i \in \tilde{V}_n$ менее очевидно. Заметим, что $p_{i,n}$, $i \in \tilde{V}_n$ сильно зависят от положительной случайной величины $\#\tilde{V}_n$, а последовательность $\#\tilde{V}_1, \#\tilde{V}_2, \dots$ формируется случайным образом: $\#\tilde{V}_{n+1} = \#\tilde{V}_n + 1$ верно лишь в случае, когда между последовательно выбранными узлами $i_1, \dots, i_m \in \tilde{V}_n$ найдется хотя бы одна пара узлов, связанная ребром из множества E_n .

Поэтому при $\epsilon = 0$ рассмотрим модель КП с функцией присоединения

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

Если функция присоединения имеет вид (7) или (11), то $C_{f,\epsilon} = (1 + \epsilon)/\epsilon$. Подставив (11) в (12), получим

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

Ограничимся случаем $\epsilon = 0$, при котором выбор узла из $V_n \setminus \tilde{V}_n$ исключен с вероятностью 1, а веса узлов, принадлежащих множеству \tilde{V}_n , равномерно распределены:

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

Последовательность функций $f_n(x) = x^{1/n}$, $n = 1, 2, \dots$ сходится к $f(x)$ в (11) поточечно. Поэтому модель КП с функцией присоединения (11) и $\epsilon = 0$ может быть рассмотрена как предельная (при $\alpha \downarrow 0$) для семейства моделей КП с функцией присоединения (7) и параметром $\epsilon = 0$.

3. Последовательный равномерный выбор узлов

Пусть G_n – наблюдаемый граф. Цель данного раздела показать, что для эволюции моделью КП достаточно знать множества \tilde{E}_n и $\tilde{E}_n \cap E_n$. Из (13) следует, что добавленный узел $\#V_1 + n$ с положительной вероятностью может быть присоединен только к существующим узлам $\{i \in V_n : \Delta_{i,n} > 0\}$, т.е. принадлежащим хотя бы одному треугольнику.

Пусть последовательность множеств $\mathcal{E}_1, \mathcal{E}_2, \dots$ такова, что для любого $n \geq 1$ пара (V_n, \mathcal{E}_n) представляет собой полный граф. Обозначим через W_n пару узлов $\{i_1, i_2\}$ из V_n , выбранных методом последовательного выбора (т.е. случайного выбора без возвращения). Последовательный выбор начинается с выбора узла i_1 из V_n с вероятностью (13). Вероятности оставшихся $\#V_n - 1$ узлов затем нормируются как $p_{i,n}/(1 - p_{i_1,n})$. Процесс повторяется, выбирается узел i_2 . Для эволюции по модели КП не имеет значения, какой узел будет выбран первым в паре. Поэтому далее будем предполагать, что W_n – двумерный случайный вектор со значениями в \mathcal{E}_n .

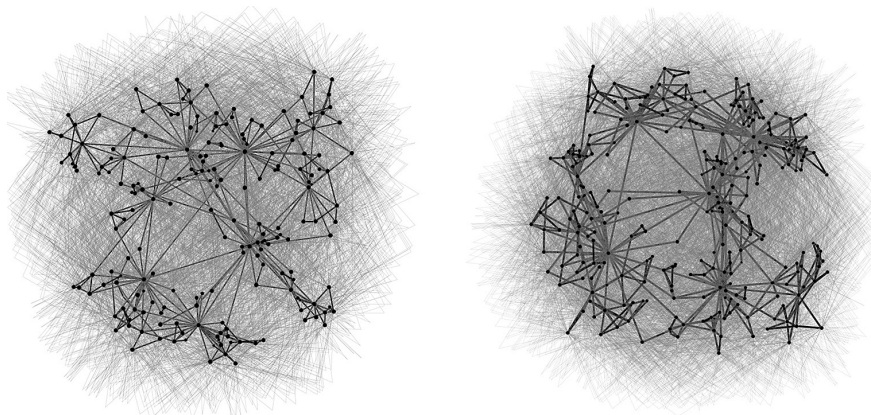


Рис. 1. Реализация случайного графа (V_n, E_n) , сгенерированного моделью КП при $n = 5000$ (слева) и $n = 10\,000$ (справа): узлы из \tilde{V}_n показаны черным цветом, ребра из $\tilde{E}_n \cap E_n$ темно-серым цветом, остальная часть графа светло-серая.

Предложение 1. Случайный вектор W_n равномерно распределен на множестве \tilde{E}_n .

Из предложения 1 следует, что правило присоединения б) из определения 1 с функцией присоединения (11) и параметрами $\epsilon = 0$, $m = 2$ не отдает предпочтения паре узлов с относительно большими коэффициентами кластеризации. Это означает отсутствие эффекта «богатый-становится-богаче» в рассматриваемой модели КП.

Равенство $P(W_n = \{i, j\}) = 0$, $i, j \in V_n \setminus \tilde{V}_n$ означает, что узлы из множества $V_n \setminus \tilde{V}_n$ не принимают участия в формировании графа G_{n+1} , поскольку вновь добавляемые узлы не присоединяются к таким узлам. Разбиение множества V_n на подмножества активных узлов \tilde{V}_n и неактивных узлов $V_n \setminus \tilde{V}_n$ позволяет использовать модель КП для моделирования различных сообществ из двух классов.

Пример 1. На рис. 1 показаны упомянутые выше классы для одной реализации модели КП (V_n, E_n) при $n \in \{5000, 10\,000\}$, где начальный граф G_1 – четырехугольник со всеми диагоналями, кроме одной:

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

4. Основной результат

Количество треугольников в графе G_n выражается через количество треугольников узлов:

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

Рассмотрим модель КП с параметром $\epsilon = 0$ и функцией присоединения (11). В [4] отмечено, что эмпирическое исследование модели КП с параметром

$m > 2$ требует громоздких вычислений из-за сложной комбинаторики. Поэтому ограничимся случаем $m = 2$.

Теорема 1. Пусть G_n , $n = 1, 2, \dots$ – последовательность графов, порожденная моделью КП с параметрами $\epsilon = 0$, $m = 2$ и функцией присоединения (11). Пусть начальный граф (V_1, E_1) конечен и удовлетворяет условию (6). Тогда

$$(14) \quad \Delta_n \rightarrow \infty, \quad n \rightarrow \infty$$

с вероятностью 1.

Пусть последовательность множеств $\tilde{E}_1, \tilde{E}_2, \dots$ такая, что для любого $n \geq 1$ $(\tilde{V}_n, \tilde{E}_n)$ – полный граф, т.е. \tilde{E}_n – множество всех возможных ребер между узлами из \tilde{V}_n . Некоторые ребра из \tilde{E}_n могут не существовать в графе G_n . Ниже в следствии 1 дается оценка снизу математического ожидания $E\Delta_n$.

Следствие 1. Пусть выполнены условия теоремы 1. Тогда

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

5. Компьютерное моделирование

Рассмотрим моделирование последовательности Δ_n , $n \geq 1$. Чтобы исследовать рост выборочного среднего Δ_n моделирование последовательности графов G_n , $n \geq 1$ не требуется. Достаточно смоделировать последовательность множеств \tilde{E}_n , $n \geq 1$.

Будем рассматривать неупорядоченные пары узлов из множества \tilde{E}_n , как обобщенную модель урны Пойа–Эггенбергера. Для каждого $n \geq 1$ множество \tilde{E}_n разделим на два подмножества: $\tilde{E}_n \cap E_n$ и $\tilde{E}_n \setminus (\tilde{E}_n \cap E_n)$. Каждая неупорядоченная пара узлов из $\tilde{E}_n \cap E_n$ интерпретируется как белый шар, а пара узлов, принадлежащих множеству $\tilde{E}_n \setminus (\tilde{E}_n \cap E_n)$, – как черный шар. Следовательно, урна (или множество \tilde{E}_n) содержит $\#(\tilde{E}_n \cap E_n)$ белых и $\#(\tilde{E}_n \setminus (\tilde{E}_n \cap E_n))$ черных шаров.

На каждом шаге $n \geq 1$ из урны извлекается (с возвращением) равновероятно один шар (см. предложение 1). Исходя из цвета шара, урна пополняется по следующему правилу. Если извлечен белый шар, то в урну добавляется два белых и $\left((1/2) \left(1 + \sqrt{1 + 8\#\tilde{E}_n} \right) - 2 \right)$ черных шара. Мощности $\#\tilde{V}_n$ и $\#\tilde{E}_n$ связаны соотношением $\#\tilde{V}_n (\#\tilde{V}_n - 1) = 2\#\tilde{E}_n$ (см. (П.1.1)), и поэтому в урну добавлен $\#\tilde{V}_n + 1$ шар. Если извлечен черный шар, то содержимое урны не меняется.

Отметим, что количество добавленных черных и/или белых шаров фиксировано в классической модели урны Пойа–Эггенбергера (см., например, [10, с. 437]), а количество добавленных черных шаров в рассматриваемом здесь случае зависит от $\#\tilde{E}_n$. Другие обобщения модели урны Пойа–Эггенбергера можно найти в [12].

Имея в виду, что пары узлов из \tilde{E}_n распределены равномерно (см. предложение 1), приведем простой алгоритм для моделирования последовательности Δ_n , $n \geq 1$.

Алгоритм 1.

1. Начальный шаг. По начальному графу G_1 вычислить Δ_1 , $\#\tilde{V}_1$, $\#(\tilde{E}_1 \cap E_1)$ и $\#(\tilde{E}_1 \setminus (\tilde{E}_1 \cap E_1))$.

2. Эволюционный шаг. Для любого $n \geq 1$ сгенерировать значение дискретной случайной величины ξ_n , равномерно распределенной на множестве $\{1, \dots, \#\tilde{E}_n\}$. Если $\xi_n > \#(\tilde{E}_n \cap E_n)$, то

$$\begin{aligned} \#(\tilde{E}_{n+1} \cap E_{n+1}) &= \#(\tilde{E}_n \cap E_n), \\ \#(\tilde{E}_{n+1} \setminus (\tilde{E}_{n+1} \cap E_{n+1})) &= \#(\tilde{E}_n \setminus (\tilde{E}_n \cap E_n)). \end{aligned}$$

Если $\xi_n \leq \#(\tilde{E}_n \cap E_n)$, то

$$\begin{aligned} \#(\tilde{E}_{n+1} \cap E_{n+1}) &= \#(\tilde{E}_n \cap E_n) + 2, \\ \#(\tilde{E}_{n+1} \setminus (\tilde{E}_{n+1} \cap E_{n+1})) &= \#\tilde{E}_{n+1} - \#(\tilde{E}_{n+1} \cap E_{n+1}), \end{aligned}$$

где

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

В обоих случаях

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

Заметим, что при реализации алгоритма на каждом шаге необходимо хранить четыре величины: Δ_1 , $\#\tilde{V}_1$, $\#(\tilde{E}_n \cap E_n)$ и $\#(\tilde{E}_n \setminus (\tilde{E}_n \cap E_n))$.

Для каждого начального графа G_1 сгенерировано 100 независимых последовательностей $\Delta_1^{(j)}, \dots, \Delta_N^{(j)}$ длины $N = 10^6$, где индекс (j) обозначает номер последовательности. Следующие полные графы выбраны в качестве начальных: 1) треугольник $G_{1,1}$; 2) полный граф $G_{1,2}$ с числом вершин 17; 3) полный граф $G_{1,3}$ с числом вершин 51. С использованием формулы

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

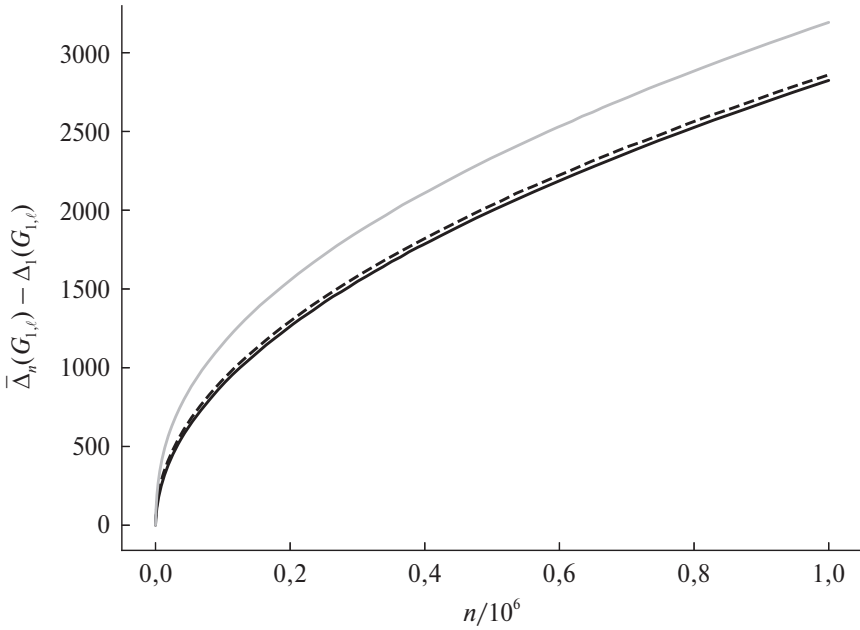


Рис. 2. Графики $\{(n, \bar{\Delta}_n(G_{1,\ell}) - \Delta_1(G_{1,\ell})), 0 \leq n \leq 10^6\}$, где кривые с $\ell \in \{1, 2, 3\}$ представлены соответственно черным, пунктирным черным и светло-серым цветом.

составлена последовательность выборочных средних $\bar{\Delta}_n(G_{1,\ell})$, $1 \leq n \leq N$ для каждого начального графа $G_{1,\ell}$. Графики $\{(n, \bar{\Delta}_n(G_{1,\ell}) - \Delta_1(G_{1,\ell})), 1 \leq n \leq N\}$, где $\ell \in \{1, 2, 3\}$ и $N = 10^6$, представлены на рис. 2. Время выполнения расчетов составило 152,88 секунды.

Метод наименьших квадратов применяется для аппроксимации дискретных данных

$$\begin{aligned} & \{(n, \bar{\Delta}_n(G_{1,\ell}) - \Delta_1(G_{1,\ell})), 2 \times 10^5 \leq n \leq N\}, \\ & \{(n, \bar{\Delta}_n(G_{1,\ell}) - \Delta_1(G_{1,\ell})), 5 \times 10^5 \leq n \leq N\} \end{aligned}$$

функцией $\varphi(n) = c_1 + c_2 n^{c_3}$. Оценки c_1 , c_2 и c_3 представлены в таблице.

Оценки параметров c_1 , c_2 и c_3

$2 \times 10^5 \leq n \leq 10^6$			
Начальный граф	\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
$5 \times 10^5 \leq n \leq 10^6$			
Начальный граф	\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

Из моделирования получены следующие выводы.

1. Результаты моделирования, представленные на рис. 2, достаточно близки к теоретическим. Действительно, $E\Delta_n$ и эмпирическое среднее $\bar{\Delta}_n$ возрастают с увеличением n , см. следствие 1 и рис. 2.
2. $\bar{\Delta}_n$ растет, как $C\sqrt{n}$, где $2 \times 10^5 \leq n \leq 10^6$, см. таблицу. Этот результат моделирования не противоречит следствию 1. Отметим, что из этого результата не следует, что $\bar{\Delta}_n$ имеет ту же (или близкую) скорость роста при $n > 10^6$.
3. При моделировании количества треугольников Δ_n при $n > 10^6$ возникли сложности, связанные с ограниченностью оперативной памяти компьютера.
4. Графики $(n, \bar{\Delta}_n(G_{1,1}) - \Delta_1(G_{1,1}))$ и $(n, \bar{\Delta}_n(G_{1,2}) - \Delta_1(G_{1,2}))$ близки, в то время как график $(n, \bar{\Delta}_n(G_{1,3}) - \Delta_1(G_{1,3}))$ имеет смещение по вертикальной оси относительно двух других графиков при $0 \leq n \leq 10^6$, что вероятно связано с размером графа $G_{1,3}$. Можно предположить, что влияние относительно большого начального графа должно быть устранено при большом n .

6. Заключение

Аналогично [9], где предложено общее определение моделей эволюции ПП, в настоящей работе введено новое обобщенное определение модели КП. Класс функций присоединения, представленный в определении 1, достаточно широк, чтобы с помощью модели КП можно было моделировать эволюцию во многих случайных системах и сетях, где вновь присоединяемые узлы более вероятно присоединяются к тесно связанным группам узлов, а не к узлам с большим количеством связей.

Исследованы теоретические свойства предельной модели КП, в которой новые узлы присоединяются методом последовательного выбора (случайного выбора без возвращения) к существующим двум узлам с вероятностями, обратно пропорциональными числу узлов, входящих в треугольники. Данная модель может быть представлена, как предел модели КП, предложенной в [4], при $\alpha \rightarrow 0$, $\epsilon = 0$ и функции присоединения (11). В силу специфики весов (3) случаи $\epsilon > 0$ и $\epsilon = 0$ должны быть исследованы отдельно.

Основные результаты статьи относятся к случаю $\epsilon = 0$. В теореме 1 доказано, что количество треугольников Δ_n стремится к бесконечности почти наверное при стремлении шага эволюции n к бесконечности. При этом начальный граф должен содержать не менее t узлов, входящих в треугольники, чтобы начать эволюцию. Следствие 1 показывает, что скорость роста среднего количества треугольников $E\Delta_n$ выше логарифмической для шага эволюции $n \geq 2$. Результаты моделирования не противоречат теоретическим результатам. Уточнение верхней и нижней границ для $E\Delta_n$ – цель дальнейшего исследования.

Доказательство предложения 1. Пусть $\{i, j\} \in \tilde{E}_n$. Для каждого узла $i \in \tilde{V}_n$ рассмотрим случайное событие $A_{n,i} \equiv \{\text{выбран узел } i\}$. По закону полной вероятности

$$(П.1.1) \quad \begin{aligned} P(W_n = \{i, j\}) &= P(A_{n,j}|A_{n,i})P(A_{n,i}) + P(A_{n,i}|A_{n,j})P(A_{n,j}) = \\ &= \frac{2}{\#\tilde{V}_n(\#\tilde{V}_n - 1)} = \frac{1}{\#\tilde{E}_n}. \end{aligned}$$

Используя (П.1.1), получаем

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

Отсюда следует $P(W_n \in \mathcal{E}_n \setminus \tilde{E}_n) = 0$ и, соответственно, $P(W_n = \{i, j\}) = 0$ для любой пары узлов $\{i, j\} \in \mathcal{E}_n \setminus \tilde{E}_n$.

Представим множество ребер E_n , $n \geq 1$ как множество неупорядоченных пар узлов. Для любого $n \geq 1$ разделим множество \tilde{E}_n на подмножества: $\tilde{E}_n \cap E_n$ и $\tilde{E}_n \setminus (\tilde{E}_n \cap E_n)$. $\tilde{E}_n \cap E_n$ состоит из неупорядоченных пар узлов, соединенных ребрами в графе G_n . Пусть $B_n = \{W_n \in \tilde{E}_n \cap E_n\}$, $n \geq 1$ – случайное событие, такое что неупорядоченная пара узлов W_n , выбранных последовательным выбором, связана ребром из E_n .

Обозначим через B^c событие, противоположное к B . Пусть $n, k \in \mathbb{N}$ и $\ell \in \mathbb{N} \cup \{0\}$ такие, что $n \geq 2$, $1 \leq k < n$, $0 \leq \ell \leq n - k$. Событие $(B_k \cap \dots \cap B_{n-1})_\ell$ обозначает объединение $(n - k)! / (\ell!(n - k - \ell)!)$ непересекающихся случайных событий таких, что ℓ событий B_j заменены на их противоположные в пересечении $B_k \cap \dots \cap B_{n-1}$. Если происходит случайное событие $(B_k \cap \dots \cap B_{n-1})_\ell$, то ℓ пар узлов среди W_k, \dots, W_{n-1} не соединены, а оставшиеся $n - k - \ell$ пар соединены. Например, при $k = 1$, $n = 4$ имеем

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

Согласно предложению 1

$$(П.2.1) \quad P(B_n) = \sum_{\{i_1, i_2\} \in \tilde{E}_n \cap E_n} \frac{1}{\#\tilde{E}_n} = \frac{\#(\tilde{E}_n \cap E_n)}{\#\tilde{E}_n}.$$

Пусть $n \geq k - 1$ и $k \geq 1$. Если произошло случайное событие $(B_k \cap \dots \cap B_{n-1})_\ell$, то

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

Тогда из (П.2.1) получим

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

где

$$(П.2.2) \quad p_n(G_k) = \frac{\# \left(\tilde{E}_k \cap E_k \right) + 2(n - k)}{\# \tilde{E}_k + (n - k) \# \tilde{V}_k + (1/2)(n - k)(n - k - 1)}.$$

Докажем неравенства, которые будут использованы при доказательстве теоремы 1.

Лемма 1. Пусть предположения теоремы 1 выполнены. Для любых натуральных чисел $k \leq n$

$$(П.2.3) \quad p_n(G_k) > 0,$$

$$(П.2.4) \quad p_n(G_k) > p_{n+1}(G_k).$$

Доказательство леммы 1. Из предположения (6) следует, что $\tilde{E}_k \neq \emptyset$ и $\tilde{E}_k \cap E_k \neq \emptyset$ для любого натурального k . В связи с этим (П.2.3) следует из (П.2.2).

Далее, с использованием (П.2.2) неравенство (П.2.4) сводится к следующему:

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

Достаточно показать, что для любого начального графа, удовлетворяющего (6), выполняется неравенство

$$(П.2.5) \quad \# \left(\tilde{E}_k \cap E_k \right) > \# \tilde{V}_k - 1.$$

Используем метод математической индукции для доказательства (П.2.5). Рассмотрим граф $(\tilde{V}_1, \tilde{E}_1 \cap E_1)$. Минимальное число ребер в $\tilde{E}_1 \cap E_1$ зависит от $\# \tilde{V}_1$. Если $\# \tilde{V}_1 = 3$, то граф с минимальным числом ребер – это треугольник, для которого $\# \left(\tilde{E}_1 \cap E_1 \right) = 3$. Поэтому (П.2.5) выполнено для $k = 1$ в случае $\# \tilde{V}_1 = 3$. Если $\# \tilde{V}_1 > 3$, то граф с минимальным числом ребер состоит из $\# \tilde{V}_1 - 2$ треугольников с одной общей стороной (см. пример 1 для $\# \tilde{V}_1 = 4$). Для таких графов выполнено $\# \left(\tilde{E}_1 \cap E_1 \right) = \# \tilde{V}_1 + 1$. Это означает, что (П.2.5) с $k = 1$ и $\# \tilde{V}_1 > 3$ снова выполнено.

Выдвинем гипотезу индукции, что (П.2.5) выполнено. Покажем, что (П.2.5) выполнено, когда k заменено на $k+1$. Если случайное событие B_k произошло, то имеем $\#(\tilde{E}_{k+1} \cap E_{k+1}) = \#(\tilde{E}_k \cap E_k) + 2$ и $\#\tilde{V}_{k+1} = \#\tilde{V}_k + 1$. Используя (П.2.5), получим

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

Случай, когда событие B_k не происходит, может быть рассмотрен аналогично.

Лемма 2. При условиях теоремы 1

$$(П.2.6) \quad P(B_n) > P(B_n | B_1 \cap \dots \cap B_{n-1}), \quad n = 2, 3, \dots$$

Доказательство леммы 2. По закону полной вероятности

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

Из леммы 1 следует, что для любого $1 \leq \ell \leq n-1$

$$P(B_n | B_1 \cap B_2 \cap \dots \cap B_{n-1}) < P(B_n | (B_1 \cap B_2 \cap \dots \cap B_{n-1})_\ell).$$

Используя последнее неравенство, получаем

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

Отсюда следует утверждение (П.2.6), так как последняя сумма равна 1.

Доказательство теоремы 1 – это проверка условий нижеприведенной леммы Борелля–Кантелли.

Лемма 3 [10, с. 79]. Пусть на одном вероятностном пространстве определены случайные события C_1, C_2, \dots , удовлетворяющие условию коррелированности: при любых натуральных u и v таких, что $u \neq v$, случайные события C_u и C_v являются отрицательно коррелированными или некоррелированными. Если $\sum_{n=1}^{\infty} P(C_n) = \infty$, то $P\{C_n \text{ б.ч.}\} = 1$.⁴

Лемма 4. Пусть условия теоремы 1 выполнены. Тогда случайные события B_1, B_2, \dots удовлетворяют условию коррелированности, т.е.,

$$(П.2.7) \quad P(B_u \cap B_v) \leq P(B_u) P(B_v).$$

Доказательство леммы 4. Если $\min\{u, v\} = 1$, то предположение (6) влечет $0 < \#(\tilde{E}_1 \cap E_1) \leq \#\tilde{E}_1$, см. доказательство леммы 1. Тогда из (П.2.1) следует, что $P(B_1) > 0$. Если $\min\{u, v\} > 1$, то комбинируя

⁴ Через $\{C_n \text{ б.ч.}\}$ обозначается событие, состоящее в том, что произойдет бесконечно много событий из C_1, C_2, \dots [11].

(П.2.3) и (П.2.6), получаем $P(B_{\min\{u,v\}}) > p_{\min\{u,v\}}(G_1) > 0$. Доказано, что $P(B_{\min\{u,v\}}) > 0$ для любых натуральных u и v . Применив определение условной вероятности, перепишем условие коррелированности (П.2.7) следующим образом:

$$(П.2.8) \quad P(B_{\max\{u,v\}}) - P(B_{\max\{u,v\}}|B_{\min\{u,v\}}) \geq 0.$$

Достаточно доказать, что неравенство (П.2.8) выполняется при $\max\{u, v\} = n + k$ и $\min\{u, v\} = n$, где n и k – любые натуральные числа.

Если $k = 1$, то применив формулу полной вероятности вместе с (П.2.4), получим

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

Здесь равенство верно при $\#(\tilde{E}_n \cap E_n) = \#(\tilde{E}_n)$.

Пусть теперь $k \geq 2$. Утверждаем, что

$$(П.2.9) \quad P(B_{n+k}) = J + \sum_{\ell=0}^{k-1} p_{n+k-\ell-1}(G_n) P(B_n^c \cap (B_{n+1} \cap \dots \cap B_{n+k-1})_\ell),$$

$$(П.2.10) \quad P(B_{n+k}|B_n) = J + \sum_{\ell=0}^{k-1} p_{n+k-\ell}(G_n) P(B_n^c \cap (B_{n+1} \cap \dots \cap B_{n+k-1})_\ell),$$

где

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

Из (П.2.9), (П.2.10) немедленно следует, что разность $P(B_{n+k}) - P(B_{n+k}|B_n)$ равна

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

Из (П.2.3), (П.2.4) следует, что последняя сумма положительная.

Докажем (П.2.9). Применяя формулу полной вероятности, получим

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

Используя тождество

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

получим

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

Применив (П.2.2), получим, что сумма в фигурных скобках равна J . Заменяв переменную суммирования на $s = \ell - 1$, получим

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

Применяя (П.2.2) еще раз, получим доказательство (П.2.9).

Тождество (П.2.10) может быть проверено похожим образом. Для этого достаточно применить формулу полной вероятности и тождество

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

Доказательство теоремы 1. Разность количества треугольников в графах G_n и G_1 , так же как и разность числа узлов, входящих в треугольники, могут быть выражены через индикаторы случайных событий B_j следующим образом:

$$(П.2.11) \quad \Delta_n - \Delta_1 = \sum_{j=1}^{n-1} I\{B_j\},$$

$$(П.2.12) \quad \#\tilde{V}_n - \#\tilde{V}_1 = \sum_{j=1}^{n-1} I\{B_j\}, \quad n \geq 2,$$

где $I\{\cdot\}$ обозначает индикатор события. Из (П.2.11), (П.2.12) следует, что количество треугольников Δ_n линейно связано с мощностью множества \tilde{V}_n :

$$(П.2.13) \quad \Delta_n = \#\tilde{V}_n + \left(\Delta_1 - \#\tilde{V}_1\right), \quad n \geq 2.$$

В виду (П.2.11) утверждение (14) следует из

$$\sum_{j=1}^{n-1} I\{B_j\} \rightarrow \infty, \quad n \rightarrow \infty$$

с вероятностью единица, что эквивалентно,

$$P(B_n \text{ б.ч.}) = 1.$$

Докажем, что последовательность B_1, B_2, \dots удовлетворяет условиям леммы Борелля–Кантелли (см. лемму 3). Одно из них проверено в лемме 4. Достаточно доказать, что

$$(П.2.14) \quad \sum_{n=1}^{\infty} P(B_n) = \infty.$$

В силу лемм 1, 2,

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

Заметим, что из (П.2.2) следует

$$(П.2.15) \quad p_n(G_1) > \frac{\#(\tilde{E}_1 \cap E_1) + 2(n-1)}{n\#\tilde{E}_1 + (1/2)(n-1)(n-2)}.$$

Используя неравенство $\#(\tilde{E}_1 \cap E_1) \leq \#\tilde{E}_1$, можно доказать, что правая часть (П.2.15) превышает $\#(\tilde{E}_1 \cap E_1) / \left((3\#\tilde{E}_1)(n-2)\right)$ при $n \geq 3$. Следовательно,

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

где $H_N = \sum_{n=1}^N 1/n$. Соотношение (П.2.14) вытекает путем применения неравенства $H_N > \ln(N)$ и соотношения $\lim_{N \rightarrow \infty} \ln(N) = \infty$.

Для доказательства следствия 1 заметим, что из (П.2.11) следует неравенство

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

Из доказательства (П.2.14) следует, что последняя сумма больше, чем $\left(\#(\tilde{E}_1 \cap E_1) / (3\#\tilde{E}_1)\right) \ln(n-1)$. Этим завершается доказательство неравенства (15).

СПИСОК ЛИТЕРАТУРЫ

1. *Van der Hofstad R.* Random Graphs and Complex Networks. Cambridge: Cambridge University Press, 2017. V. 1.
2. *Bollobas B.* Random Graphs. Cambridge: Cambridge University Press, 2001. 2nd ed.
3. *Newman M.E.J.* Networks. Oxford University Press: Oxford, New York, 2018.
4. *Bagrow J.P., 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. No. 103. I. 5. P. 058701.
6. *Qun Liu, Zhishan Dong.* Limit laws for the number of triangles in the generalized random graphs with random node weights // *Stat. Probab. Lett.* 2020. No. 161. P. 108733.
7. *Bobkov S.G., Danshina M.A., 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*, V. 358. Eds. Karapetyants, A.N., Pavlov, I.V., Shiryayev, A.N. Springer: Cham, 2021. P. 109–133.
8. *Garavaglia A., Stegehuis C.* Subgraphs in preferential attachment models // *Advances in Applied Probability.* 2019. No. 51. P. 898–926.
9. *Wang T., Resnick S.I.* Consistency of Hill estimators in a linear preferential attachment model // *Extremes.* 2019. No. 22. P. 1–28.
10. *Fristedt B., Gray L.* A Modern Approach to Probability Theory. Boston: Birkhäuser, 1997.
11. *Ширяев А.Н.* Вероятность. М.: Наука, 1989.
12. *Chen M.-R., Kuba M.* On generalized Pólya urn models // *J. Appl. Probab.* 2013. No. 50. P. 1169–1186.

Статья представлена к публикации членом редколлегии А.И. Ляховым.

Поступила в редакцию 18.01.2024

После доработки 26.08.2024

Принята к публикации 20.09.2024