

# ИССЛЕДОВАНИЕ ЗАВИСИМОСТЕЙ И РАСПРЕДЕЛЕНИЙ В СЛУЧАЙНЫХ СЕТЯХ ДЛЯ СМЕШАННЫХ МОДЕЛЕЙ ЭВОЛЮЦИИ И ПРИ УДАЛЕНИИ УЗЛОВ <sup>1</sup>

Маркович Н. М.<sup>2</sup>, Рыжов М. С.<sup>3</sup>, Кулик М. Р.<sup>4</sup>.

( ФГБУН Институт проблем управления  
им. В.А. Трапезникова РАН, Москва)

*Изучается эволюция случайной сети моделями предпочтительного (preferential attachment), кластерного (clustering attachment) и смешанного присоединений для формирования связей вновь присоединенных узлов с существующими узлами. Рассматриваются стратегии удаления узла на каждом шаге эволюции сети: 1) без удаления узлов и связей; 2) удаление наименее влиятельного узла среди наиболее «старых», где в качестве меры влиятельности узла используется его пейджранг; 3) удаление узла с вероятностью, обратно пропорциональной числу его связей. Для этих стратегий удаления моделированием сравниваются зависимости двух характеристик случайных сетей: числа связей узлов и числа их треугольников (т.е. троек связанных узлов, в которые узел вовлечен) и поведение кластерных коэффициентов узлов. Оценивается тяжесть хвоста распределения для числа связей и треугольников. Смешанное кластерно-предпочтительное присоединение предлагается впервые.*

Ключевые слова: эволюция сети, предпочтительное и кластерное присоединения, смешанное присоединение, число связей узла, число треугольников узла, удаление узла, кластерный коэффициент.

## 1. Введение

Рассматривается эволюция ненаправленных случайных сетей и описывающих их графов в предположении разных стратегий удаления узлов с их связями. Случайные графы – статистические модели, имеющие множество приложений от нейронных и транспортных сетей до анализа социально-информационных

<sup>1</sup> Исследование выполнено за счет гранта Российского научного фонда №24-21-00183.

Авторы признательны анонимному рецензенту за ценные замечания.

<sup>2</sup> Наталья Михайловна Маркович, д.ф.-м.н., г.н.с. (nat.markovich@gmail.com).

<sup>3</sup> Максим Сергеевич Рыжов, н.с. (maksim.ryzhov@frtk.ru).

<sup>4</sup> Михаил Ростиславович Кулик, студент, техник (mishakulik2002@yandex.ru).

сетей. Обозначим граф на шаге эволюции  $t$  как  $G_t = (V_t, E_t)$ ,  $V_t$  – множество вершин,  $E_t$  – множество ребер,  $\|\cdot\|$  – мощность множества (число элементов). Эволюция, т.е. присоединение новых узлов новыми связями к существующим узлам, начинается с начального графа  $G_0$ . Рост числа узлов и связей в реальных сетях происходит по различным моделям эволюции таким, как предпочтительное присоединение (ПП, preferential attachment) различных типов [3, 6, 11, 12, 28], кластерное присоединение (КП, clustering attachment) [7, 25, 26] и их смеси [4]. Смена модели присоединения происходит из-за внешних факторов, например, появления новой области знаний в сети цитирования. Параметры модели могут меняться во времени [4].

Важными характеристиками эволюции сети во времени  $t$  являются распределение числа связей узлов (node degree) и кластерный коэффициент (clustering coefficient) узла, зависящий как от числа связей, так и от числа треугольников, к которым данный узел принадлежит. Под треугольником понимаются три узла, связанные между собой. Число треугольников определяет локальную связанность узлов в сети. Хвостовой индекс (ХИ) или его обратная величина – индекс экстремальной величины (ИЭВ), показывающие тяжесть хвоста распределения числа связей, могут сменой своего знака отражать и смену модели эволюции. В работе смена модели эволюции предполагается случайной, происходящей с заданной вероятностью.

Известно, что ПП ведет к степенному закону (power-law) распределения числа связей (так называемому свойству scale-free) [11] и к свойству «маленького мира» (small-world) [17], с. 238. В моделях ПП кластерный коэффициент сильно отличается для разных графов. Геометрическим моделям ПП свойственна высокая кластерность [1, 20, 21]. Случайная величина (сл.в.) имеет степенное распределение, если

$$(1) \quad P(X = k) \sim Ck^{-(1+\iota)}, \iota > 0, \quad k \rightarrow \infty,$$

для некоторых положительных констант  $C$  и  $\iota$ .<sup>1</sup> Дискретное распределение (1) относится к классу распределений с тяжелым хвостом, а именно, с регулярно меняющейся хвостовой функцией  $\bar{F}(x) := 1 - F(x)$  ( $F(x)$  – функция распределения):

$$(2) \quad \bar{F}(x) = \ell(x)x^{-\iota} \text{ для любого } x > 0,$$

где  $\ell(x)$  медленно меняющаяся функция, т.е. такая, что  $\lim_{x \rightarrow \infty} \ell(tx)/\ell(x) = 1 \forall t > 0$  [24]. Чем меньше ХИ  $\iota$ , тем тяжелее хвост.

Некоторые сети, например транспортные, имеют распределения числа связей с легкими хвостами [22]. Легкие хвосты стремятся к нулю при  $x \rightarrow \infty$  медленнее, чем хвост экспоненциального распределения. Их ИЭВ  $\gamma = 1/\iota < 0$  или  $\gamma = 0$ . Примерами распределений с легкими хвостами с  $\gamma = 0$  служат экспоненциальное и нормальное распределения, а с  $\gamma < 0$  – равномерное и бета-распределения. Определение ИЭВ приводится в [15].

В случайных сетях, например сетях цитирования, многие старые узлы теряют свою значимость, присоединение к ним новых узлов (цитирование новыми статьями) не происходит, и они оказываются «забытыми», изолированными, фактически удаляются из сети.

Цель работы – изучение влияния смены моделей присоединения, естественного для динамики реальных сетей, и стратегий удаления узлов на тяжесть хвоста распределений числа связей и треугольников, на их зависимость и на поведение средних кластерных коэффициентов узлов. Влияние начальных графов, с которых начинается эволюция, на указанные характеристики, а также детекция сохранения геометрии сети в ходе эволюции требуют исследования.

В работе [19] получены ХИ степенных распределений для числа связей узлов ненаправленных графов при равномерном, т.е. с равными вероятностями, удалении узла или(и) связи на каждом

<sup>1</sup>  $a_n \sim b_n$  означает, что последовательности  $a_n$  и  $b_n$  асимптотически равны, т.е.,  $a_n/b_n \rightarrow 1, n \rightarrow \infty$ .

шаге эволюции ПП. В нашей работе рассматривается неравномерное удаление узлов.

Тестирование степени зависимости (полная, сильная, слабая зависимость и независимость) между двумя случайными векторами, например, числом входящих и выходящих связей узлов графов, полученных с помощью ПП, предложено в работе [36]. Полная зависимость соответствует концентрации двумерных наблюдений на луче с положительным наклоном; сильная зависимость – концентрации наблюдений в правильном конусе в квадранте положительной полуоси  $R_+^2$ ; слабая зависимость – рассеянию во всем квадранте  $R_+^2$ ; независимость – концентрации вдоль осей  $X$  и  $Y$ . Строгие тесты построены в [36] в предположении, что наблюдения каждого вектора независимы, одинаково распределены и имеют распределение с тяжелым хвостом, что не всегда отвечает контексту случайных сетей. В рассматриваемой нами зависимости числа треугольников и числа связей, распределения этих сл.в. могут иметь как тяжелый, так и легкий хвост в зависимости от моделей эволюции и стратегий удаления узлов с их связями. Эти зависимости сравниваются средствами моделирования для трех моделей эволюции: ПП, КП и смешанном кластерно-предпочтительном (КПП) присоединениях. КПП предлагается впервые.

В работе изучаются три стратегии удаления на каждом шаге эволюции: 1) без удаления существующих узлов или связей; 2) удаление наименее влиятельного узла среди наиболее «старых», где в качестве меры влиятельности узла используется пейджранг узла; 3) удаление узла с вероятностью, обратно пропорциональной числу его связей.

Статья имеет следующее содержание. В разделе 2 приводится краткий обзор моделей эволюции сетей. В разделе 3 описываются стратегии удаления маловлиятельных узлов. Раздел 4 содержит применяемые подходы и методы статистического анализа. Разделы 5 и 6 содержат описание эксперимента и результаты моделирования для сравнения ИЭВ и зависимости числа треугольников и числа связей узлов, поведения среднего кластерного ко-

эффицента для ПП, КП и КПП. В разделе 7 приводится анализ реальных сетей. Раздел 8 содержит выводы.

## 2. Обзор моделей эволюции случайных графов

ПП моделирует присоединение новых узлов с большей вероятностью к наиболее влиятельным узлам, так как вероятность присоединения пропорциональна числу связей узла. Новый узел присоединяется к существующему узлу  $i$  ненаправленного графа с вероятностью

$$(3) \quad P_{PA}(i, t) = d_{i,t} / \sum_{s \in V_t} d_{s,t},$$

где  $d_{i,t}$  – число связей узла  $i \in V_t$ . ПП был предложен для обоснования тяжелых хвостов распределений характеристик влиятельности узлов таких, как число их связей и пейджранги,<sup>1</sup> что характерно для веб-графов. Поэтому ПП интенсивно изучается в литературе, см. обзор [24]. Для направленных графов предложены три схемы ПП с учетом направления присоединяемых ребер и возможности создания новых ребер между существующими узлами, что позволяет строить несколько ребер между узлами и петли узлов [34]. ПП выражается в стремительном росте связей небольшого числа узлов.

КП предложен в [7] для моделирования локальных сетей, когда присоединение нового узла происходит лишь к узлам, вовлеченным в треугольники. Например, в социальных сетях индивиды могут иметь друзей не среди популярных людей, а в сво-

<sup>1</sup> Определение вектора пейджранга  $R = (R_1, \dots, R_n)^T \in (0, \infty)^n$  предложено в [9] как единственное решение системы линейных уравнений

$$(4) \quad R_i = c \sum_{j:(j,i) \in E} \frac{R_j}{D_j} + (1-c)q_i, \quad i \in \{1, \dots, n\},$$

где сумма берется по числу страниц  $j$ , имеющих входящие связи в страницу  $i$  (число входящих связей узла  $i$ ),  $D_j$  – число выходящих связей страницы  $j$ ,  $c \in (0, 1)$  – коэффициент демпфирования, обычно равный 0,85 в Гугле,  $q = (q_1, q_2, \dots, q_n)$  – вектор персонализации такой, что  $0 \leq q_i \leq 1$ ,  $\sum_{i=1}^n q_i = 1$ . В несколько иной форме система линейных уравнений приводится в [5].

ем окружении, формируя тесные сообщества. Вместо гигантских узлов с большими числом связей, как для ПП, КП ведет к тесно связанным сообществам. В [7] показано без доказательства, что КП порождает легкие хвосты распределений числа связей узлов. КП моделирует эволюцию социальных сетей и сетей, где новые узлы присоединяются не к узлам с большим числом связей, а к сильно связанным сообществам. Вероятность присоединения нового узла к узлу  $i \in V_t$  определяется как

$$(5) \quad P_{CA}(i, t) = \left( c_{i,t}^\alpha + \epsilon \right) / \left( \sum_{s \in V_t} c_{s,t}^\alpha + \epsilon \|V_t\| \right), \quad \alpha > 0,$$

$$(6) \quad P_{CA}(i, t) = \frac{\mathbb{I}\{c_{i,t} > 0\} + \epsilon}{\sum_{j \in V_t} \mathbb{I}\{c_{j,t} > 0\} + \|V_t\| \epsilon}, \quad \alpha = 0,$$

где

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

– кластерный коэффициент,  $0 \leq c_{i,t} \leq 1$ ,  $\Delta_{i,t}$  – число треугольников, в которые вовлечен узел  $i$  на шаге эволюции  $t$ ,  $\alpha, \epsilon \geq 0$  – параметры модели, [24]. Модель КП не позволяет моделировать параллельные связи между парами узлов и петли узлов.

Предложим модель КПП, где новый узел присоединяется к существующим узлам с вероятностями, соответствующими моделям ПП или КП, что больше соответствует реальности. Примером может служить просмотр сайтов своих коллег или каких-то ученых с большим индексом цитирования. На шаге эволюции  $t$  вероятность присоединения нового узла к узлу  $i \in V_t$  следующая:

$$P_{CPA}(i, t) = \begin{cases} P_{PA}(i, t) & \text{с вероятностью } p, \\ P_{CA}(i, t) & \text{с вероятностью } 1 - p. \end{cases}$$

$p \in (0, 1)$  – параметр модели КПП.

Изучим поведение среднего кластерного коэффициента узлов сети:

$$(7) \quad \bar{c}_t = \frac{1}{\|V_t\|} \sum_{i \in V_t} c_{i,t}$$

для моделей ПП, КП и КПП.  $\bar{C}_t$  показывает тенденцию узлов к формированию треугольников и наличие геометрии сетей. В геометрической сети узел не только имеет вес, но и некоторую позицию в  $d$ -мерном метрическом пространстве (геолокацию). Вероятность соединения двух узлов в геометрическом графе является возрастающей функцией от весов узлов и убывающей функцией от расстояния между ними [27]. Если  $\bar{C}_t$  не убывает с возрастанием числа узлов  $\|V_t\|$ , то есть геометрия в сети [27].

### 3. Стратегии удаления маловлиятельных узлов

Если на каждом шаге эволюции добавляется и удаляется по одному узлу, то общее число узлов в сети остается неизменным. В нашем моделировании предполагается присоединение нового узла к существующим двумя связями. Так как узел  $i$  удаляется вместе с его связями, а их случайное число  $d_{i,t}$ , то число связей в сети на каждом таком шаге эволюции меняется на величину  $2 - d_{i,t}$ .

#### 3.1. УДАЛЕНИЕ МАЛОВЛИЯТЕЛЬНЫХ УЗЛОВ С БОЛЬШИМ «ВРЕМЕНЕМ ЖИЗНИ» В СЕТИ

Опишем алгоритм выбора наименее влиятельного узла среди наиболее «старых» узлов сети. В качестве меры влиятельности узла используем его пейджеранг. Для вычисления пейджеранга  $R_i$  случайно выбранной страницы  $i \in V$  веб-графа  $G = (V, E)$  используем итерационный метод

$$(8) \quad \widehat{R}_i^{(n,0)} = 1, \quad \widehat{R}_i^{(n,k)} = \sum_{j:(j,i) \in E} \frac{c}{D_j} \widehat{R}_j^{(n,k-1)} + (1 - c), \quad k \in N,$$

предложенный в [13] для заданного равномерного вектора  $q_i = 1/n$ ,  $1 \leq i \leq n = \|V\|$ . Итерация (8) продолжается, пока разность между двумя последовательными итерациями  $|\widehat{R}_i^{(n,k)} - \widehat{R}_i^{(n,k-1)}|$  не становится достаточно малой, что достигается вследствие быстрой сходимости метода после небольшого числа итераций  $k$ .

Узлы сети на шаге эволюции  $t$  располагаются в порядке убывания «времени жизни»  $\{T_i\}$ ,  $i = 1, 2, \dots, t$ , измеряемом в числе шагов, прошедших с момента присоединения  $i$ -го узла  $t_{0,i}$  до  $t$ , т.е.  $T_i = t - t_{0,i}$ .

Поскольку «молодые» узлы могут еще не набрать большой пейджранг, то кандидаты на удаление могут быть рассмотрены среди узлов с «временами жизни»  $T_i > t - \epsilon_0$ ,  $0 < \epsilon_0 < t$ . Среди узлов-кандидатов выбирается узел с наименьшим пейджрангом и удаляется. Константа  $\epsilon_0$  – параметр метода удаления «старого» узла и определения «молодых» узлов, не подлежащих удалению.

### 3.2. УДАЛЕНИЕ МАЛОВЛИЯТЕЛЬНЫХ УЗЛОВ НЕЗАВИСИМО ОТ «ВОЗРАСТА»

Рассмотрим удаление узла  $i$  с вероятностью, обратно пропорциональной числу его связей:

$$(9) \quad P_d(i, t) = \left( 1 + \sum_{s \in V_t, s \neq i} \frac{d_{i,t} + \eta}{d_{s,t} + \eta} \right)^{-1}, \quad \eta > 0,$$

величина которой тем больше, чем меньше  $d_{i,t}$ . В случае, если все узлы изолированные и  $d_{s,t} = 0$ ,  $\forall s \in V_t$ , они будут убираться с равными вероятностями  $1/\|V_t\|$ .

Подход позволяет удалить узел вне зависимости от его времени пребывания в сети. Число связей  $\{d_{i,t}\}$  в (9) можно заменить на пейджранги.

## 4. Применяемые методы статистического анализа

### 4.1. НЕПАРАМЕТРИЧЕСКОЕ ОЦЕНИВАНИЕ ИНДЕКСА ЭКСТРЕМАЛЬНОЙ ВЕЛИЧИНЫ

Для распределений  $\{d_{i,t}\}$  и  $\{\Delta_{i,t}\}$  оцениваются ИЭВ  $\gamma$  состоятельными непараметрическими оценками моментов (moment, M) [16], УН [8, 23] и смешанных моментов (mixed moment, MM) [18].  $\iota = 1/\gamma$  определяет ХИ. Выбранные оценки предназначены для  $\gamma \in R$ .  $\gamma \leq 0$  указывает на легкий хвост распределения, а  $\gamma > 0$  на тяжелый по теореме Фишера – Типпета – Гнеденко [8, 23].



Для последовательности сл.в.  $\{X_n\}_{n \geq 1}$  обозначим соответствующие порядковые статистики, расположенные в порядке возрастания, как  $X_{i,n}$ ,  $1 \leq i \leq n$ . Используем статистику

$$\hat{\varphi}_n(k) = \frac{M_n^{(1)}(k) - L_n^{(1)}(k)}{\left(L_n^{(1)}(k)\right)^2},$$

$$L_n^{(1)}(k) = 1 - \frac{1}{k} \sum_{i=1}^k \frac{X_{n-k,n}}{X_{n-i+1,n}}, \quad M_n^{(1)}(k) = \frac{1}{k} \sum_{i=1}^k \ln \frac{X_{n-i+1,n}}{X_{n-k,n}}.$$

Оценка  $MM$  определяется как

$$\hat{\gamma}^{MM}(n, k) = \frac{\hat{\varphi}_n(k) - 1}{1 + 2 \min(\hat{\varphi}_n(k) - 1, 0)}.$$

Оценка  $UH$  определяется как

$$\hat{\gamma}^{UH}(n, k) = (1/k) \sum_{i=1}^k \ln UH_i - \ln UH_{k+1},$$

где  $UH_i = X_{(n-i)} M_n^{(1)}(i)$ .

Оценка моментов определяется как

$$\hat{\gamma}^M(n, k) = M_n^{(1)}(k) + 1 - 0.5 \left(1 - (M_n^{(1)}(k))^2 / S_{n,k}\right)^{-1},$$

где  $S_{n,k} = (1/k) \sum_{i=1}^k (\ln X_{(n-i+1)} - \ln X_{(n-k)})^2$  [16].

Число наибольших порядковых статистик  $k$  в этих оценках будем выбирать методом бутстреп [23], где число бутстреп-подвыборок, т.е. выборки с возвращением,  $B = 100$ .

Слабая сходимость оценки Хилла  $M_n^{(1)}(k)$  доказана для ненаправленных графов в работе [35].

Оцениванию  $\gamma$  в случайных графах мешает повтор наблюдений, так как число связей и треугольников дискретные сл.в. Это сокращает число различающихся наблюдений и делает их зависимыми. Указанные оценки  $\gamma$  требуют достаточно больших

объемов выборок для большей точности и построены в предположении, что наблюдаемые сл.в. независимы и одинаково распределены. Корреляции и небольшой объем выборки увеличивают дисперсию оценок. Бутстреп-подвыборки с одинаковыми наблюдениями исключены из рассмотрения.

#### 4.2. ВЫЧИСЛЕНИЕ КОРРЕЛЯЦИИ РАССТОЯНИЙ

Корреляция расстояний (distance correlation,  $dCor$ ), в отличие от коэффициента корреляции Спирмена, показывает степень линейной или нелинейной зависимости двух случайных величин или векторов  $(X, Y)$ , где  $(X_i, Y_i), i = 1, 2, \dots, n$ , – пары их наблюдений с вещественными значениями.

Напомним определение. Пусть  $a_{ij} = \|X_i - X_j\|$  и  $b_{ij} = \|Y_i - Y_j\|$ , где  $\|\cdot\|$  обозначает евклидову норму. Соответствующие матрицы определяются как  $(A_{i,j})_{i,j=1}^n$  и  $(B_{i,j})_{i,j=1}^n$ ,

$$\begin{aligned} A_{i,j} &= a_{i,j} - \frac{1}{n} \sum_{l=1}^n a_{il} - \frac{1}{n} \sum_{k=1}^n a_{kj} + \frac{1}{n^2} \sum_{k,l=1}^n a_{kl}, \\ B_{i,j} &= b_{i,j} - \frac{1}{n} \sum_{l=1}^n b_{il} - \frac{1}{n} \sum_{k=1}^n b_{kj} + \frac{1}{n^2} \sum_{k,l=1}^n b_{kl}. \end{aligned}$$

Тогда оценка квадрата выборочной ковариации расстояний определяется как  $\mathcal{V}_n^2(X, Y) = \frac{1}{n^2} \sum_{i,j=1}^n A_{i,j} B_{i,j}$ . Выборочная корреляция расстояний определена как

(10)

$$\mathcal{D}_n^2(X, Y) = \begin{cases} \frac{\mathcal{V}_n^2(X, Y)}{\sqrt{\mathcal{V}_n^2(X, X) \mathcal{V}_n^2(Y, Y)}} & \text{при } \mathcal{V}_n^2(X, X) \mathcal{V}_n^2(Y, Y) > 0, \\ 0 & \text{при } \mathcal{V}_n^2(X, X) \mathcal{V}_n^2(Y, Y) = 0. \end{cases}$$

Ее свойства следующие:  $\mathcal{D}_n(X, Y) \in [0, 1]$ ;  $\mathcal{D}_n(X, Y) = 0$  тогда и только тогда, когда  $X$  and  $Y$  независимы;  $\mathcal{D}_n(X, Y) = 1$  указывает на их зависимость  $Y = A + bCX$  для некоторого вектора  $A$ , скаляра  $b$ , ортонормальной матрицы  $C$ . Для вычислений используем оценку из пакета 'dcor' [31].

## 5. Описание экспериментальных данных и применяемых методов

### 5.1. ЭВОЛЮЦИЯ БЕЗ УДАЛЕНИЯ УЗЛОВ И СВЯЗЕЙ

В качестве начального графа, с которого начинается эволюция ПП, КП и КПП, выберем треугольник связанных узлов, минимально возможный граф, позволяющий строить эволюцию по КП и КПП. Предполагаем, что на каждом шаге эволюции  $t$  новый узел соединяется двумя связями с узлами из  $V_t$ , которые выбираются с вероятностью соответствующей модели. Для КПП возможно, что один из существующих узлов выбирается по ПП, второй по КП или оба узла – одной моделью.

КП берется с параметрами  $(\alpha, \epsilon) = \{(0, 0), (1, 0), (1, 1)\}$ . ПП не содержит параметров, см. (3). КПП использует параметры КП. Моделируются по 20 графов, эволюционирующих по КП, ПП, КПП, для усреднения по этим графам среднего кластерного коэффициента  $\bar{C}_t$  (7), общего числа связей  $\|E_t\|$  и треугольников:

$$(11) \quad \tilde{C}_t = \frac{1}{20} \sum_{j=1}^{20} \bar{C}_t^{(j)}, \quad \|\tilde{E}_t\| = \frac{1}{20} \sum_{j=1}^{20} \|E_t^{(j)}\|, \quad \tilde{\Delta}_t = \frac{1}{20} \sum_{j=1}^{20} \Delta_t^{(j)}$$

для каждого шага эволюции  $t$ ,  $t \in \{0, 10^4\}$ .

Общее число треугольников на шаге  $t$  вычисляется как

$$\Delta_t = \frac{1}{3} \sum_{i \in \tilde{V}_t} \Delta_{i,t},$$

где  $\tilde{V}_t$  число узлов, вовлеченных в треугольники, т.е. для которых  $c_{i,t} > 0$ . По  $\{\Delta_{i,t}\}$  и  $\{d_{i,t}\}$  узел каждого из 20 графов оценивается ИЭВ  $\gamma$ . Вычисляется средняя оценка

$$(12) \quad \tilde{\gamma}_t = \frac{1}{20} \sum_{j=1}^{20} \hat{\gamma}_t^{(j)}(n, k_j),$$

где  $\hat{\gamma}_t^{(j)}(n, k_j)$  любая оценка  $\gamma$  из раздела 4.1.

### 5.2. ЭВОЛЮЦИЯ С УДАЛЕНИЕМ УЗЛОВ С ИХ СВЯЗЯМИ

При удалении узла выбор треугольника в качестве начального графа может привести к невозможности развития графа. КП

делает возможным присоединение только к узлам, входящим хотя бы в один треугольник. При удалении одной из вершин треугольника, присоединение к оставшимся двум какого-то нового узла невозможно. ПП позволяет новому узлу присоединиться к двум вершинам и образовать новый треугольник, но убирание вершины треугольника приводит к закичиванию и невозможности роста графа.

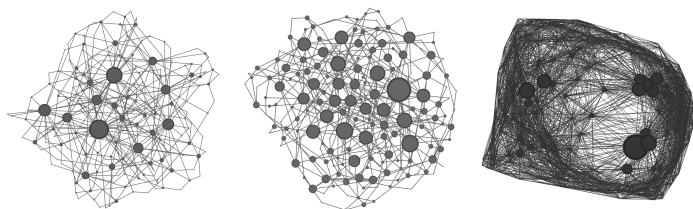


Рис. 1. Начальные графы с  $n = 200$  узлами, полученные по ПП (рис. 1а), по КП  $(\alpha, \epsilon) = (1, 1)$  (рис. 1б) и методом построения геометрических графов с  $r = 0,25$  (рис. 1в), где размер изображенных узлов пропорционален числу связей

Для стратегий удаления узлов рассмотрим три типа графов в качестве начальных: 1) геометрический граф; 2) граф, порождаемый ПП; 3) граф, порождаемый КП с параметрами  $(\alpha, \epsilon) = (1, 1)$ , см. рис. 1. Напомним, что геометрический граф ненаправленный, в котором  $n$  узлов расположены равномерно в квадрате с ребром единица, где ребро соединяет два узла только если расстояние между ними меньше чем  $r$  [29].

## 6. Основные результаты

### 6.1. СЕТЬ БЕЗ УДАЛЕНИЯ УЗЛОВ И СВЯЗЕЙ

На рис. 2 (левый столбец) показана степенная зависимость числа треугольников  $\{\Delta_{i,t}\}$  от числа связей  $\{d_{i,t}\}$ ,  $i \in V_t$ , в графах, эволюционирующих по ПП, КП и КПП, для фиксированного

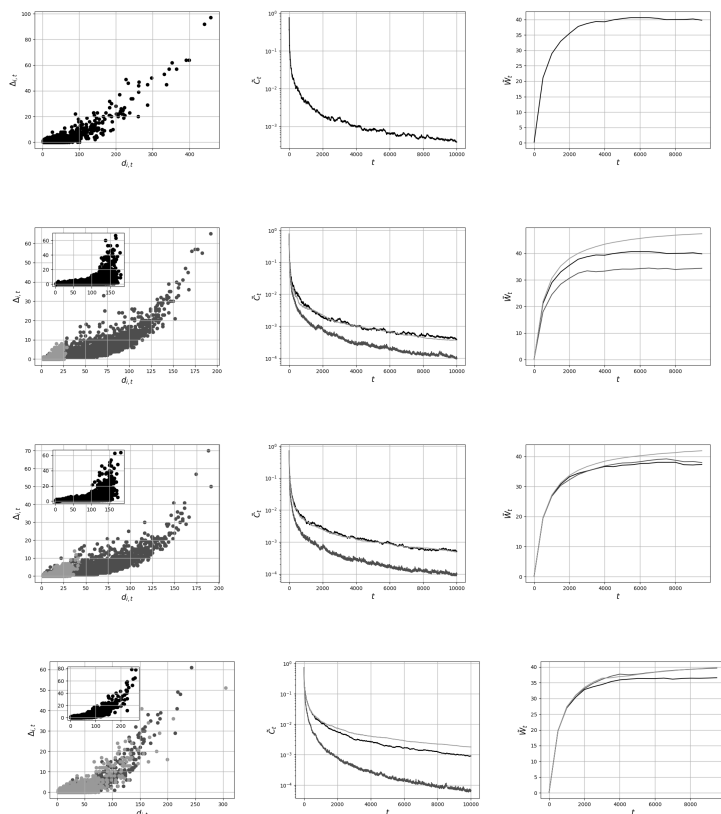


Рис. 2.  $\{\Delta_{i,t}\}$  против  $\{d_{i,t}\}$ ,  $i \in V_t$ , для узлов графов, полученных после  $t = 10^4$  шагов эволюции по ПП, КП, КПП с  $p = 0,2$  и КПП с  $p = 0,8$ , с первого по четвёртый столбец соответственно; для КП и КПП взяты параметры  $(\alpha, \epsilon) \in \{(0, 0), (1, 0), (1, 1)\}$  (чёрные, тёмно-серые, серые точки). Усреднённый по 20 графам средний кластерный коэффициент  $\bar{C}_t$  (средний столбец) и взвешенное число треугольников  $\bar{W}_t$  (правый столбец), см. (11) и (16), для соответствующих моделей присоединения (чёрные, тёмно-серые, серые линии). Эволюция проводится без удаления узлов и связей из начального графа треугольника

момента времени  $t$ , т.е.

$$(13) \quad \Delta_{i,t} = d_{i,t}^\beta, \quad \beta < 1.$$

Из неравенства  $\Delta_{i,t} < d_{i,t}$  очевидно следует, что хвост распределения  $\Delta_{i,t}$  более легкий, чем у  $d_{i,t}$ . Действительно, для ПП известно (см., например, [34]), что  $d_{i,t}$  имеет степенное распределение (2) с некоторым ХИ  $\iota$  и ИЭВ  $\gamma = 1/\iota$ . Тогда имеем

$$P\{\Delta_{i,t} > x\} = P\{d_{i,t}^\beta > x\} = \ell(x)x^{-\iota/\beta} = \ell(x)x^{-1/(\beta\gamma)}, \forall x > 0.$$

Отсюда  $\gamma_\Delta = \beta\gamma < \gamma$ , а  $\iota_\Delta > \iota$ . По свойству распределений с регулярно меняющимися хвостами  $d_{i,t}$  имеет более тяжелый хвост, чем  $\Delta_{i,t}$ .

Для КП, напротив, в [7] показано с помощью эвристик и моделирования, что  $d_{i,t}$  имеет распределение с легким (экспоненциальным) хвостом. В силу (13) имеем

$$(14) \quad P\{\Delta_{i,t} > x\} = P\{d_{i,t} > x^{1/\beta}\} < \exp(-\lambda x), \quad \beta < 1,$$

для некоторого  $\lambda > 0$ , т.е.  $\Delta_{i,t}$  также имеет более легкий хвост, чем  $d_{i,t}$ . На рис. 2 (средний столбец) показаны убывающие по времени последовательности усредненных кластерных коэффициентов  $\tilde{C}_t$ ,  $t \geq 0$ . Для КПП и КП параметры  $(\alpha, \epsilon) = \{(0, 0), (1, 1)\}$  приводят к  $\tilde{C}_t$ , близкому к случаю ПП, поскольку присоединение к новым узлам происходит фактически равновероятно согласно (5), (6), независимо от наличия треугольников. Для  $(\alpha, \epsilon) = (1, 0)$   $\tilde{C}_t$  убывает быстрее остальных случаев. Убывание  $\tilde{C}_t$  может указывать на отсутствие геолокации узлов и слабой тенденции узлов к кластеризации. Однако, как отмечено в [27], убывание  $\tilde{C}_t$  может быть медленным, порядка  $t^{\iota-2} \ln t$ , если хвостовой индекс  $\iota$  характеристики узла близок к 2. Для геометрических неоднородных случайных графов  $\tilde{C}_t$  не убывает с возрастом  $t$  [10]. Для КП с ростом  $\alpha$  хвост распределения временных интервалов между формированием новых треугольников становится все более тяжелым [7]. Поскольку новый узел

присоединяется к узлам с  $c_{i,t} > 0$ , т.е. с треугольниками, а треугольники возникают у малой части узлов, то появление гигантских узлов, как в случае ПП, маловероятно, что влияет на тяжесть хвоста распределения  $\{d_{i,t}\}$ . Для ПП вероятность присоединения пропорциональна числу связей, которое растет со временем для небольшого числа узлов.

В таблице 1 для проверки тяжести хвоста распределений приводятся оценки ИЭВ  $\hat{\gamma}$  для числа связей  $\{d_{i,t}\}$  и взвешенного числа треугольников  $\{W_{i,t}\}$  узла  $i \in V_t$  вместо  $\{\Delta_{i,t}\}$ . Здесь и далее для оценивания  $\gamma$  использовались выборки размера  $10^4$  узлов, полученные для 20 графов.

Выборка  $\{\Delta_{i,t}\}$  содержит небольшое число уникальных значений, так как много узлов имеют одинаковое число треугольников. Это не позволяет качественно оценить  $\gamma$  для  $\{\Delta_{i,t}\}$ . Число уникальных значений  $\{\Delta_{i,t}\}$  и  $\{d_{i,t}\}$  для моделей ПП, КП, КПП и соответствующих графов на рис. 2 показано в таблице 2. В силу свойств эволюционирующих сетей, в отношении  $\{\Delta_{i,t}\}$  наблюдаются вырожденные распределения. Преодолеть это можно, заменив число треугольников узла  $i$  на взвешенное число треугольников

$$W_{i,t} = \sum_{j,k \in V_t} 1_{\{(i,j,k)=\Delta\}} / (d_{i,t}d_{j,t}d_{k,t}),$$

где  $1_{\{(i,j,k)=\Delta\}}$  – индикаторная функция события, что узлы  $i, j, k$  образуют треугольник;  $d_{i,t}, d_{j,t}, d_{k,t}$  – число связей его вершин, суммирование проводится по всем возможным парам связанных соседей узла  $i$ . В [27] подобное взвешивание вводится для распознавания геометрических графов по значениям статистики

$$(15) \quad W_t = \sum_{i,j,k \in V_t} 1_{\{(i,j,k)=\Delta\}} / (d_{i,t}d_{j,t}d_{k,t}).$$

Отмечается, что в негеометрических моделях ПП треугольник обычно формируется между узлами с большим числом связей. В геометрических графах большое число треугольников формируется у близких по геолокации узлов с небольшим числом свя-

зей. Поэтому для негеометрических графов  $W_n$  ограничено, а для геометрических растёт вместе с числом узлов  $n = \|V\|$ .

Для ПП  $\hat{\gamma} > 0$ , что соответствует тяжёлым хвостам распределений для  $\{d_{i,t}\}$  и  $\{W_{i,t}\}$ . Для КПП и КП  $\hat{\gamma}$  могут быть отрицательными или близкими к 0 для  $\{d_{i,t}\}$ , что означает лёгкие хвосты распределений. Однако с ростом  $p$  (доли ПП в КПП) для  $\{d_{i,t}\}$  оценки  $\hat{\gamma}$  становятся положительными, что говорит о тяжёлых хвостах и доминировании ПП над КП. Для  $\{W_{i,t}\}$  всегда  $\gamma > 0$ , и для КПП хвост распределения  $\{W_{i,t}\}$  наиболее тяжёлый, независимо от пропорций КП и ПП в смеси и величин параметров  $(\alpha, \epsilon)$ .

Таблица 1. Средние по выборкам 20 графов оценки  $\tilde{\gamma}_t$  (12) для  $\{d_{i,t}\}$  и  $\{W_{i,t}\}$ ,  $i \in V_t$ ; значения  $dCor$  для  $\{d_{i,t}\}$ ,  $\{\Delta_{i,t}\}$ ,  $i \in V_t$  со среднеквадратическими отклонениями в скобках при эволюции без удаления узлов из начального графа треугольника до  $t = 10^4$

| Модель эволюции | $(\alpha, \epsilon)$ | Оценка MM         |                  | Оценка M          |                   | Оценка УН         |                  | dCor              |
|-----------------|----------------------|-------------------|------------------|-------------------|-------------------|-------------------|------------------|-------------------|
|                 |                      | $\{d_{i,t}\}$     | $\{W_{i,t}\}$    | $\{d_{i,t}\}$     | $\{W_{i,t}\}$     | $\{d_{i,t}\}$     | $\{W_{i,t}\}$    |                   |
| ПП              |                      | 0,48<br>(0,058)   | 0,62<br>(0,161)  | 0,493<br>(0,031)  | 0,283<br>(0,103)  | 0,458<br>(0,031)  | 0,378<br>(0,142) | 0,402<br>(<0,001) |
| КП              | (0, 0)               | -0,204<br>(0,145) | 0,895<br>(0,189) | -0,685<br>(0,125) | 0,265<br>(0,117)  | -0,404<br>(0,113) | 0,354<br>(0,197) | 0,896<br>(<0,001) |
|                 | (1, 0)               | -0,002<br>(0,012) | 0,631<br>(0,161) | 0,038<br>(0,072)  | 0,222<br>(0,125)  | 0,027<br>(0,064)  | 0,29<br>(0,102)  | 0,922<br>(<0,001) |
|                 | (1, 1)               | 0,03<br>(0,065)   | 0,688<br>(0,172) | -0,159<br>(0,091) | 0,323<br>(0,086)  | -0,058<br>(0,061) | 0,475<br>(0,162) | 0,192<br>(<0,001) |
| КПП             | (0, 0)               | -0,237<br>(0,074) | 1,359<br>(0,058) | -0,427<br>(0,102) | 0,441<br>(0,002)  | -0,336<br>(0,118) | 0,719<br>(0,043) | 0,898<br>(<0,001) |
|                 | (1, 0)               | 0,003<br>(0,071)  | 1,349<br>(0,053) | 0,043<br>(0,078)  | 0,442<br>(0,0031) | 0,031<br>(0,088)  | 0,737<br>(0,013) | 0,923<br>(<0,001) |
|                 | $p = 0,2$<br>(1, 1)  | 0,032<br>(0,057)  | 1,362<br>(0,067) | 0,014<br>(0,068)  | 0,443<br>(0,003)  | 0,02<br>(0,052)   | 0,731<br>(0,015) | 0,222<br>(<0,001) |
| КПП             | (0, 0)               | 0,222<br>(0,087)  | 1,357<br>(0,012) | 0,288<br>(0,04)   | 0,44<br>(0,002)   | 0,317<br>(0,044)  | 0,732<br>(0,027) | 0,72<br>(<0,001)  |
|                 | (1, 0)               | 0,253<br>(0,087)  | 1,367<br>(0,019) | 0,12<br>(0,088)   | 0,443<br>(0,001)  | 0,162<br>(0,048)  | 0,729<br>(0,02)  | 0,779<br>(<0,001) |
|                 | $p = 0,8$<br>(1, 1)  | 0,353<br>(0,068)  | 1,364<br>(0,057) | 0,369<br>(0,054)  | 0,443<br>(0,003)  | 0,352<br>(0,043)  | 0,722<br>(0,003) | 0,351<br>(<0,001) |



Таблица 2. Число различных значений  $\{\Delta_{i,t}\}$  и  $\{d_{i,t}\}$ ,  $i \in V_t$ , для графов, эволюционирующих по ПП, КП, КПП без удалений узлов (стратегия 1), с удалением наименее влиятельных «старых» узлов (стратегия 2), с удалением узлов с вероятностью (9) (стратегия 3) на шаге эволюции  $t = 10^4$

| Модель эволюции   | $(\alpha, \epsilon)$ | Стратегия 1   |                    | Стратегия 2   |                    | Стратегия 3   |                    |
|-------------------|----------------------|---------------|--------------------|---------------|--------------------|---------------|--------------------|
|                   |                      | $\{d_{i,t}\}$ | $\{\Delta_{i,t}\}$ | $\{d_{i,t}\}$ | $\{\Delta_{i,t}\}$ | $\{d_{i,t}\}$ | $\{\Delta_{i,t}\}$ |
| ПП                |                      | 182           | 47                 | 67            | 974                | 26            | 6                  |
| КП                | (0,0)                | 173           | 50                 | 67            | 1022               | 28            | 18                 |
|                   | (1,0)                | 156           | 46                 | 70            | 1071               | 23            | 14                 |
|                   | (1,1)                | 27            | 7                  | 67            | 980                | 9             | 3                  |
| КПП,<br>$p = 0,2$ | (0,0)                | 168           | 48                 | 72            | 1082               | 29            | 14                 |
|                   | (1,0)                | 160           | 42                 | 71            | 1042               | 23            | 12                 |
|                   | (1,1)                | 41            | 13                 | 68            | 989                | 12            | 3                  |
| КП,<br>$p = 0,8$  | (0,0)                | 194           | 48                 | 67            | 982                | 32            | 12                 |
|                   | (1,0)                | 157           | 41                 | 68            | 992                | 24            | 10                 |
|                   | (1,1)                | 129           | 31                 | 70            | 1059               | 18            | 3                  |

Положительная корреляция между числом связей узла и числом треугольников очевидна из рис. 2 (см. левый столбец). Оценим меру зависимости  $\{\Delta_{i,t}\}$  и  $\{d_{i,t}\}$  для моделей эволюции с помощью  $dCor$  (10).

В таблице 1 показано, что имеется зависимость в данных для моделей КП и всех видов моделей КПП с  $(\alpha, \epsilon) \in \{(0, 0), (1, 0)\}$ , максимальное значение  $dCor$  наблюдается для КПП с  $p = 0,2$  и  $(\alpha, \epsilon) = (1, 0)$ , где  $dCor$  оказался равен 0,923, что показывает сильную зависимость в данных. Обратный эффект наблюдается у оставшихся моделей. Видна тенденция: при переходе от КП к КПП и от КПП с  $p = 0,2$  к КПП с  $p = 0,8$   $dCor$  для  $(\alpha, \epsilon) = (1, 1)$  растет, однако каждое значение заметно меньше, чем любое из описанных выше, что говорит о малой зависимости в данных. Действительно, модель КП с  $(\alpha, \epsilon) = (1, 1)$  соответствует фактически равновероятному выбору существующего узла новым узлом по (5).

Рассмотрим усреднённое взвешенное число треугольников

$$(16) \quad \tilde{W}_t = \frac{1}{20} \sum_{j=1}^{20} W_t^{(j)},$$

где  $W_t^{(j)}$  – взвешенное число треугольников графа  $j$ , см. (15). Рис. 2 (правый столбец) показывает ограниченность (невозрастание)  $W_t$  с ростом  $t$ , что вместе с убыванием  $\tilde{C}_t$  означает отсутствие геометрии в графе [27].

### 6.2. СЕТЬ С УДАЛЕНИЕМ НАИМЕНЕЕ ВЛИЯТЕЛЬНЫХ «СТАРЫХ» УЗЛОВ

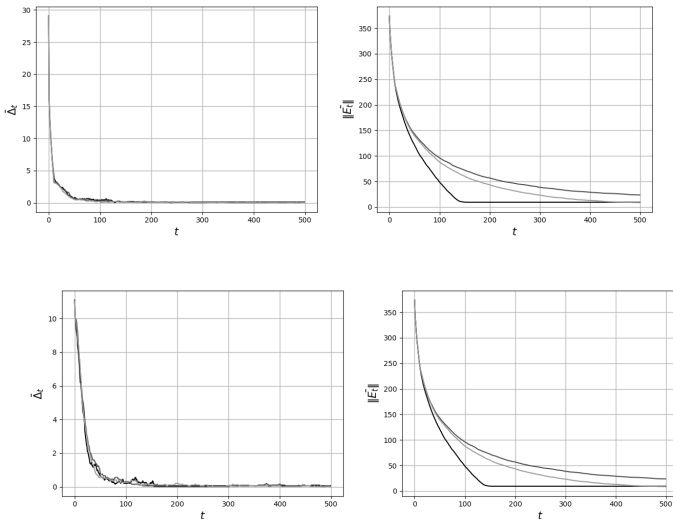


Рис. 3. Динамика во времени усреднённого по 20 графам общего числа треугольников  $\bar{\Delta}_t$  и числа связей графа  $\|E_t\|$  (11) для ПП (чёрные линии), КП  $((\alpha, \epsilon) = (1, 1))$  (тёмно-серые линии) и КПП  $(p = 0,2, (\alpha, \epsilon) = (1, 1))$  (серые линии) при начальных графах ПП (верхняя строка) и КП (нижняя строка), см. раздел 5.2. Эволюция с удалением маловлиятельных «старых» узлов

Проведем аналогичный анализ, что и в разделе 6.1, но с удалением одного из «старых» узлов с наименьшим пейджрангом на каждом шаге эволюции при присоединении нового узла. Перед эволюцией каждому узлу начального графа в произвольном порядке присваивается «возраст» – натуральное число от 200 до 1 – который будет возрастать с каждым шагом эволюции на единицу.

Выбран «пограничный возраст» узла – параметр  $\epsilon_0 = 100$ , больше которого узел считается «старым» и может быть удален, см. раздел 3.1. В качестве начальных можно выбрать графы, описанные в секции 5.2.

Таблица 3. Средние по выборкам 20 графов оценки  $\tilde{\gamma}_t$  (12) и значения  $dCor$  для  $\{d_{i,t}\}$  и  $\{\Delta_{i,t}\}$ ,  $i \in V_t$  со среднеквадратическими отклонениями в скобках при эволюции из геометрического начального графа (с числом узлов  $10^4$  и  $r = 0,25$ ) с удалением «старых» узлов с наименьшими пейджрангами до  $t = 10^4$

| Модель эволюции  | $(\alpha, \epsilon)$ | Оценка ММ         |                    | Оценка М          |                    | Оценка УН         |                    | dCor              |
|------------------|----------------------|-------------------|--------------------|-------------------|--------------------|-------------------|--------------------|-------------------|
|                  |                      | $\{d_{i,t}\}$     | $\{\Delta_{i,t}\}$ | $\{d_{i,t}\}$     | $\{\Delta_{i,t}\}$ | $\{d_{i,t}\}$     | $\{\Delta_{i,t}\}$ |                   |
| ПП               |                      | -0,478<br>(0,182) | -0,416<br>(0,214)  | -0,566<br>(0,252) | -0,854<br>(0,266)  | -0,431<br>(0,144) | -0,458<br>(0,187)  | 0,979<br>(<0,001) |
|                  | (0, 0)               | -0,568<br>(0,169) | -0,394<br>(0,12)   | -0,928<br>(0,356) | -0,713<br>(0,242)  | -0,489<br>(0,16)  | -0,509<br>(0,146)  | 0,981<br>(<0,001) |
|                  | (1, 0)               | -0,553<br>(0,121) | -0,316<br>(0,166)  | -0,771<br>(0,241) | -0,65<br>(0,233)   | -0,498<br>(0,166) | -0,491<br>(0,144)  | 0,979<br>(<0,001) |
| КП               | (1, 1)               | -0,502<br>(0,214) | -0,354<br>(0,176)  | -0,736<br>(0,353) | -0,666<br>(0,372)  | -0,431<br>(0,161) | -0,384<br>(0,176)  | 0,984<br>(<0,001) |
|                  | (0, 0)               | -0,374<br>(0,19)  | -0,349<br>(0,157)  | -0,476<br>(0,233) | -0,601<br>(0,352)  | -0,344<br>(0,139) | -0,411<br>(0,153)  | 0,979<br>(<0,001) |
|                  | (1, 0)               | -0,425<br>(0,196) | -0,336<br>(0,171)  | -0,415<br>(0,212) | -0,642<br>(0,297)  | -0,37<br>(0,136)  | -0,469<br>(0,224)  | 0,980<br>(<0,001) |
| КПП<br>$p = 0,2$ | (1, 1)               | -0,398<br>(0,193) | -0,352<br>(0,13)   | -0,554<br>(0,243) | -0,668<br>(0,255)  | -0,415<br>(0,158) | -0,44<br>(0,221)   | 0,983<br>(<0,001) |
|                  | (0, 0)               | -0,498<br>(0,191) | -0,323<br>(0,16)   | -0,545<br>(0,248) | -0,648<br>(0,248)  | -0,388<br>(0,168) | -0,423<br>(0,239)  | 0,980<br>(<0,001) |
|                  | (1, 0)               | -0,441<br>(0,164) | -0,312<br>(0,119)  | -0,519<br>(0,102) | -0,538<br>(0,231)  | -0,435<br>(0,164) | -0,427<br>(0,176)  | 0,979<br>(<0,001) |
| КПП<br>$p = 0,8$ | (1, 1)               | -0,481<br>(0,122) | -0,382<br>(0,129)  | -0,644<br>(0,133) | -0,526<br>(0,237)  | -0,405<br>(0,171) | -0,458<br>(0,148)  | 0,982<br>(<0,001) |

Однако стратегия, описанная в разделе 3.1, ведет к быстрой деградации сети, если начальный граф выбран по ПП или КП, см. рис. 3. А именно, «старые» влиятельные узлы, сильно связанные со «старыми» невлиятельными узлами, теряют связи с ними вследствие их удаления. В результате быстро растет число изолированных узлов. Сеть деградирует. Ввиду отсутствия числа связей и треугольников у изолированных узлов, их распределе-

ния становятся вырожденными.

При использовании геометрического графа в качестве начального изолированные узлы не появляются. Поэтому в дальнейшем будем использовать геометрические начальные графы.

На рис. 4 (левый столбец) зависимость числа треугольников и числа связей узлов степенная (13), но с  $\beta > 1$ , и мало отличается при различных моделях эволюции. Ввиду отсутствия теоретических результатов для распределения числа узлов и треугольников для ПП и КП для рассматриваемой стратегии удаления узла, сравнение ИЭВ  $\{\Delta_{i,t}\}$  и  $\{d_{i,t}\}$  проведем средствами моделирования.

На рис. 4 (средний столбец) средний кластерный коэффициент  $\tilde{C}_t$ , увеличиваясь в начале эволюции, затем стабилизируется около некоторого ненулевого значения с течением времени. Это, согласно [27], означает сохранение геометрической структуры графа даже при последовательных удалениях узлов. При большей доле ПП в смешанной эволюции,  $\tilde{C}_t$  увеличивается и стабилизируется около значений, близких для разных значений параметров КП ( $\alpha, \epsilon$ ), рис. 4 нижняя строка, центр. Для ПП (рис. 4, первая строка, центр)  $\tilde{C}_t$  имеет тенденцию к некоторому снижению после увеличения, что означает снижение влияния изначальной геолокации узлов. При КП или его доминировании в КПП (рис. 4, вторая и третья строки, центр) наблюдается меньшее стабильное значение  $\tilde{C}_t$  при  $(\alpha, \epsilon) = (1, 1)$ . Поскольку  $0 \leq c_{i,t} \leq 1 \forall i, t$ , то  $\epsilon = 1$  доминирует в значении вероятности присоединения (5), что означает выбор новым узлом какого-то узла среди существующих, близкий к равномерному.

При присоединении и удалении узла на каждом шаге эволюции общее число узлов не меняется. Для имитации возрастания числа узлов в графе так, как это происходит в сети без удаления, рассматривается последовательность начальных графов с возрастающим числом узлов  $n$ . Рис. 4 (правый столбец) показывает убывание с ростом числа узлов начального графа  $n$  усредненного по 20 графам взвешенного числа треугольников  $\bar{W}_n$  для всех моделей эволюции. Вместе с убыванием среднего кластер-

ного коэффициента  $\tilde{C}_t$  это означает исчезновение изначальной геолокации узлов.

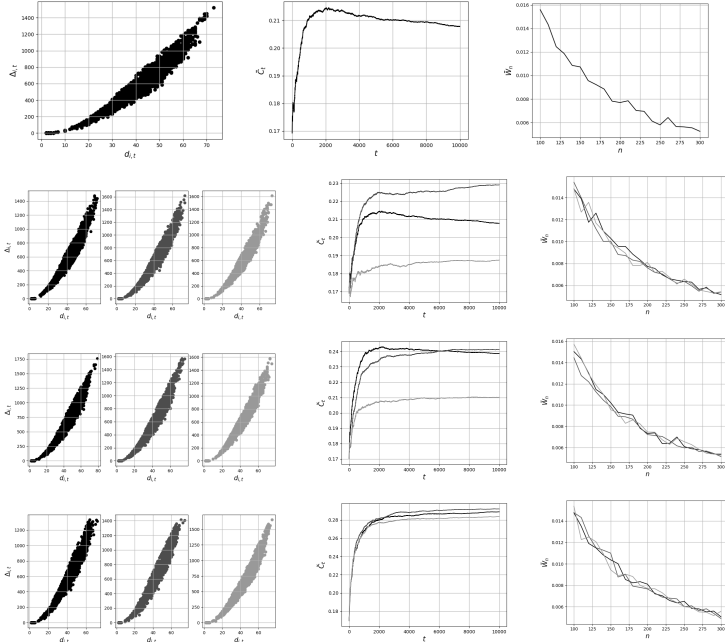


Рис. 4.  $\{\Delta_{i,t}\}$  против  $\{d_{i,t}\}$ ,  $i \in V_t$  для узлов графов после  $t = 10^4$  шагов эволюции по ПП, КП, КПП с  $p = 0,2$  и КПП с  $p = 0,8$ , с первой по четвертую строки соответственно; для КП и КПП взяты параметры  $(\alpha, \epsilon) \in \{(0, 0), (1, 0), (1, 1)\}$  (чёрные, тёмно-серые, серые точки). Усреднённые по 20 графам средний кластерный коэффициент  $\tilde{C}_t$  (см. (11)) против  $t$  и взвешенное число треугольников  $\tilde{W}_n$  (см. (16)) против числа узлов в начальном графе  $n$  для соответствующих методов присоединения (чёрные, тёмно-серые, серые линии, соответственно, в среднем и в правом столбцах). Эволюция происходит с удалением «старых» узлов с наименьшими пейджерангами и геометрическим графом с  $r = 0,25$  в качестве начального

В таблице 3 приведены оценки  $\gamma$  для  $\{d_{i,t}\}$  и  $\{\Delta_{i,t}\}$ . Для всех моделей эволюции  $\hat{\gamma} < 0$ , что соответствует распределениям с лёгкими хвостами. Лёгкие хвосты вызваны удалением «старых» маловлиятельных узлов, из-за чего «старые» влиятельные узлы теряют связи. Из теоретических результатов в [19], с. 5, лёгкие хвосты распределения числа связей могут возникнуть в ненаправленных графах при равномерном удалении узлов с их связями. Чем больше значения ИЭВ  $\gamma < 0$ , тем легче хвост распределения. Действительно, из теоремы Фишера – Типпета – Гнеденко следует, что отрицательные  $\gamma$  соответствуют классу Вейбулла  $\Psi_\gamma(x) = \exp(-(-x)^{-1/\gamma})$  области притяжения максимума последовательности случайных величин с общей функцией распределения  $F(x)$ , [8, 23]. Для  $\gamma_1 < \gamma_2$  легко можно получить, что  $\Psi_{\gamma_2}(x) < \Psi_{\gamma_1}(x)$ . Из таблицы 3 следует, что  $\{\Delta_{i,t}\}$  имеют более легкие хвосты, чем  $\{d_{i,t}\}$ . Число неповторяющихся наблюдений для  $\{\Delta_{i,t}\}$  больше, чем для  $\{d_{i,t}\}$ , см. таблицу 2, что составляет достаточную статистику для оценивания.

$dCor$  в таблице 3 указывает на сильную зависимость между  $\{\Delta_{i,t}\}$  и  $\{d_{i,t}\}$ , так как все значения близки к единице. Максимальное значение достигается при реализации модели эволюции КП с параметрами  $(\alpha, \epsilon) = (1, 1)$ . Все значения  $dCor$  не ниже показателя 0,979. Напомним, что  $dCor = 1$  соответствует совпадению двух линейных подпространств рассматриваемых случайных векторов, что показывает взаимосвязь между векторами.

### 6.3. СЕТЬ С УДАЛЕНИЕМ УЗЛОВ ПО ВЕРОЯТНОСТИ

Рассмотрим эволюцию с удалением одного узла с вероятностью (9), где  $\eta = 1$ , на каждом шаге эволюции. В качестве начального графа используем геометрический граф, см. раздел 5.2. Начальные графы, построенные по КП и ПП, ведут к аналогичным результатам и не рассматриваются. Данные  $\{\Delta_{i,t}\}$  содержат небольшое число различных значений, см. таблицу 2, что затрудняет оценивание распределения. Поэтому вместо  $\{\Delta_{i,t}\}$  рассмотрим  $\{W_{i,t}\}$ .

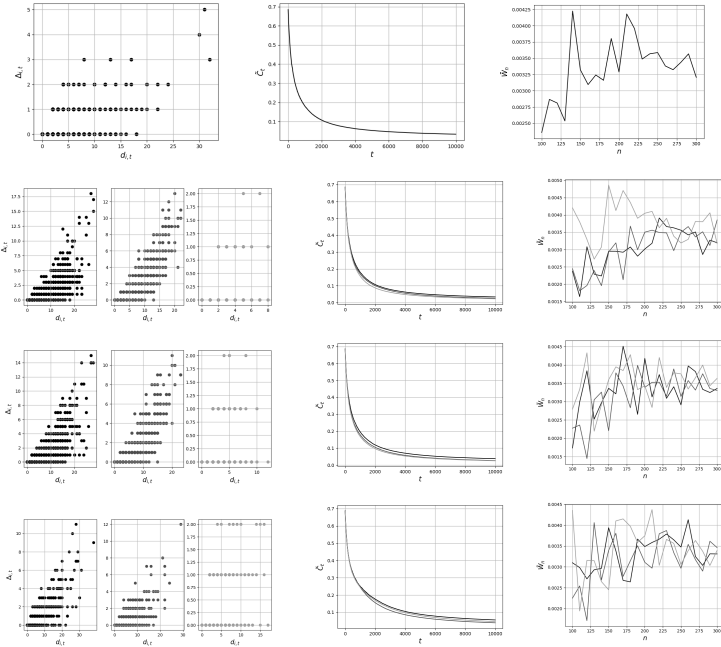


Рис. 5.  $\{\Delta_{i,t}\}$  против  $\{d_{i,t}\}$ ,  $i \in V_t$  для узлов графов после  $t = 10^4$  шагов эволюции по ПП, КП, КПП с  $p = 0,2$  и КПП с  $p = 0,8$ , с первой по четвёртую строки соответственно; для КП и КПП взяты параметры  $(\alpha, \epsilon) \in \{(0, 0), (1, 0), (1, 1)\}$  (чёрные, тёмно-серые, серые точки). Усреднённый по 20 графам средний кластерный коэффициент  $\tilde{C}_t$ , см. (11) против  $t$ , и взвешенное число треугольников  $\tilde{W}_t$ , см. (16) против числа узлов в начальном графе  $n$ , для соответствующих методов присоединения (чёрные, тёмно-серые, серые линии, соответственно, в среднем и правом столбцах). Эволюция происходит с удалением узлов с вероятностью (9) из начального геометрического графа

Рис. 5 (левый столбец) показывает положительную корреляцию между  $\Delta_{i,t}$  и  $d_{i,t}$  для КП и КПП с параметрами  $(\alpha, \epsilon) = \{(0, 0), (1, 0)\}$ . Удаление узлов по вероятности, в отличие

от удаления по «времени жизни» (см. раздел 3.1), приводит к существенно меньшему числу узлов, входящих в треугольники. КП и КПП с  $(\alpha, \epsilon) = (1, 1)$  приводят к наименьшему числу треугольников, приходящихся на каждый узел, т.е. к меньшей связности узлов. Это отражается и в наименьших значениях  $dCor$ .

Рис. 5 (средний столбец) показывает спадание  $\tilde{C}_t$  к нулю, более медленное, чем при неудаении узлов. Параметры моделей мало влияют на скорость убывания  $\tilde{C}_t$ .

Рис. 5 (правый столбец) демонстрирует стабилизацию  $\tilde{W}_n$  около значения, близкого к нулю, с ростом числа узлов выбранного начального графа  $n$  для всех моделей эволюции. Из этого следует отсутствие в графах геолокации узлов в результате эволюции.

Таблица 4 показывает, что  $d_{i,t}$  имеет легкие хвосты распределений при КП и КПП с  $(\alpha, \epsilon) = \{(0, 0), (1, 0)\}$ , и тяжелые при ПП и КПП с  $(\alpha, \epsilon) = (1, 1)$ . Оценивать распределения  $\{\Delta_{i,t}\}$  невозможно из-за недостаточности статистики и можно предположить, что их хвосты легче, чем у  $d_{i,t}$ , так как треугольников меньше, чем числа связей.  $\{W_{i,t}\}$  имеют тяжелые хвосты, кроме случая КП с  $(\alpha, \epsilon) = (1, 1)$ , когда хвосты легкие.

$dCor$  в таблице 4 указывает на зависимость в данных. Для всех моделей КП и КПП минимальное значение  $dCor$  достигается при значениях параметров  $(\alpha, \epsilon) = (1, 1)$ . Для модели ПП значение  $dCor$  говорит о слабой зависимости в данных. Напомним, что значение  $dCor = 0$  означает независимость между векторами.

## 7. Анализ реальных ненаправленных сетей

Рассмотрим реальные сети из регистров [30, 32]. В сети полётов (Flight) [30] узлами можно считать 13169 наиболее крупных аэропортов, принявших 3573482 рейсов за 121 день 2019 года. Тройки аэропортов, между которыми есть рейсы за день, принимаются за треугольники узлов. Число полётов между аэропортами – это число связей узлов. В [30] пренебрегается направлением связей (полётов). Более полная выборка по полётам представлена



Таблица 4. Средние по выборкам 20 графов оценки  $\tilde{\gamma}_t$  (12) для  $\{d_{i,t}\}$  и  $\{W_{i,t}\}$ ,  $i \in V_t$ ; значения dCor для  $\{d_{i,t}\}$  и  $\{\Delta_{i,t}\}$ ,  $i \in V_t$ , со среднеквадратическими отклонениями в скобках для начального геометрического графа (с числом узлов  $10^4$  и  $r = 0,25$ ), эволюционирующего с удалением узлов с вероятностью (9) до  $t = 10^4$

| Модель           | $(\alpha, \epsilon)$ | Оценка MM         |                   | Оценка M          |                   | Оценка УН         |                   | dCor             |
|------------------|----------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|------------------|
|                  |                      | $\{d_{i,t}\}$     | $\{W_{i,t}\}$     | $\{d_{i,t}\}$     | $\{W_{i,t}\}$     | $\{d_{i,t}\}$     | $\{W_{i,t}\}$     |                  |
| ПП               |                      | 0,153<br>(0,043)  | 0,16<br>(0,066)   | 0,148<br>(0,052)  | 0,121<br>(0,031)  | 0,123<br>(0,026)  | 0,179<br>(0,029)  | 0,324<br>(0,006) |
| КП               | (0, 0)               | -0,102<br>(0,042) | 0,133<br>(0,019)  | -0,092<br>(0,021) | 0,11<br>(0,013)   | -0,178<br>(0,083) | 0,086<br>(0,026)  | 0,847<br>(0,003) |
|                  | (1, 0)               | -0,204<br>(0,048) | 0,532<br>(0,019)  | -0,186<br>(0,012) | 0,262<br>(0,003)  | -0,159<br>(0,028) | 0,303<br>(0,028)  | 0,894<br>(0,001) |
|                  | (1, 1)               | -0,141<br>(0,091) | -0,395<br>(0,121) | -0,315<br>(0,003) | -0,365<br>(0,008) | -0,303<br>(0,114) | -0,104<br>(0,009) | 0,131<br>(0,009) |
| КПП<br>$p = 0,2$ | (0, 0)               | -0,059<br>(0,012) | 0,059<br>(0,046)  | -0,021<br>(0,015) | 0,059<br>(0,019)  | -0,064<br>(0,02)  | 0,047<br>(0,023)  | 0,845<br>(0,002) |
|                  | (1, 0)               | -0,181<br>(0,012) | 0,375<br>(0,021)  | -0,288<br>(0,021) | 0,21<br>(0,048)   | -0,201<br>(0,051) | 0,259<br>(0,044)  | 0,891<br>(0,001) |
|                  | (1, 1)               | 0,203<br>(0,002)  | 0,303<br>(<0,001) | 0,342<br>(0,002)  | 0,494<br>(0,007)  | 0,279<br>(0,012)  | 0,696<br>(0,028)  | 0,158<br>(0,012) |
| КПП<br>$p = 0,8$ | (0, 0)               | 0,1<br>(0,055)    | 0,19<br>(0,067)   | 0,054<br>(0,037)  | 0,366<br>(0,035)  | 0,058<br>(0,038)  | 0,335<br>(0,079)  | 0,658<br>(0,028) |
|                  | (1, 0)               | 0,115<br>(0,097)  | 0,170<br>(0,083)  | 0,202<br>(0,033)  | 0,314<br>(0,025)  | 0,061<br>(0,078)  | 0,125<br>(0,023)  | 0,606<br>(0,049) |
|                  | (1, 1)               | 0,125<br>(0,09)   | 0,721<br>(<0,001) | 0,183<br>(0,057)  | 0,488<br>(0,001)  | 0,103<br>(0,061)  | 0,772<br>(0,007)  | 0,227<br>(0,013) |

в [33]. Поскольку полетов за день может быть несколько, они могут считаться параллельными связями, которые засчитываются за одну связь при подсчете числа треугольников. Закрытие аэропорта (его терминала) с прекращением полетов интерпретируется, как удаление узла с его связями.

В сети обмена e-мейлами (ia-enron-email-dynamic) приведено общение 87273 сотрудников энергетической компании Enron [32]. Игнорируя направление передачи, e-мейлы (всего 1148072) фиксировались раз в 6,4 секунд в течение полумесяца. Более полная выборка e-мейлов представлена в [14]. Обмен e-мейлами между тремя индивидуумами в течение 6,4 секунд интерпретируется

как треугольник узлов; число e-мейлов – как число связей узлов; несколько сообщений между двумя адресатами – как параллельные связи, также рассматриваемые за одну связь при подсчете числа треугольников. Число связей  $\{d_{i,t}\}$  и треугольников  $\{\Delta_{i,t}\}$  с течением времени не уменьшается. Увольнение сотрудника, когда он прекращает общение, рассматривается как удаление узла с его связями.

Транспортные сети могут иметь лёгкие хвосты распределений [37], а социальные сети – тяжёлые для числа связей [2].

Рис. 6 (левый столбец) показывает положительную корреляцию между  $\{\Delta_{i,t}\}$  и  $\{d_{i,t}\}$  и присутствие аномально больших наблюдений (outliers), что свойственно распределениям с тяжелыми хвостами. Проверим это предположение оцениванием  $\gamma$ .

Как видно из таблицы 5, для сети Flight  $\{d_{i,t}\}$  имеют тяжелые хвосты, а  $\{\Delta_{i,t}\}$  – легкие. Для сети Enron обе характеристики имеют тяжелые хвосты распределений. Тяжелые хвосты распределений для  $\{d_{i,t}\}$ , причем более тяжелый для Enron, могут указывать на возможную модель эволюции ПП для Enron и КПП с преобладанием КП для Flight.

Таблица 5. Оценки ИЭВ  $\gamma$  и значения  $dCor$  для  $\{d_{i,t}\}$  и  $\{\Delta_{i,t}\}$ ,  $i \in V_t$ , для реальных сетей

| Сеть   | Оценка MM     |                    | Оценка M      |                    | Оценка UN     |                    | dCor  |
|--------|---------------|--------------------|---------------|--------------------|---------------|--------------------|-------|
|        | $\{d_{i,t}\}$ | $\{\Delta_{i,t}\}$ | $\{d_{i,t}\}$ | $\{\Delta_{i,t}\}$ | $\{d_{i,t}\}$ | $\{\Delta_{i,t}\}$ |       |
| Flight | 0,211         | -0,189             | 0,243         | -0,228             | 0,263         | -0,186             | 0,818 |
| Enron  | 0,461         | 0,205              | 0,491         | 0,205              | 0,47          | 0,21               | 0,846 |

Имеется сильная зависимость между  $\{d_{i,t}\}$  и  $\{\Delta_{i,t}\}$  ввиду  $dCor$ , близкого к 1, в таблице 5. Средний кластерный коэффициент  $\tilde{C}_t$  сетей стабилизируется с эволюцией, см. рис. 6, средний столбец, что говорит о наличии геолокации узлов.

Исследуем поведение последовательности взвешенного числа треугольников  $\{W_n\}$  от числа узлов  $n$ ,  $n \geq 1$ , подграфов сети. Каждый подграф содержит  $n$  узлов с наибольшими числами связей. Рис. 6, правый столбец, показывают рост  $W_n$ , что подтверждает наличие геометрии обеих сетей согласно [27].

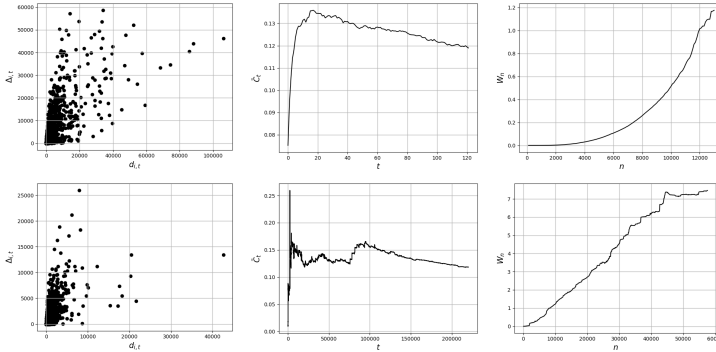


Рис. 6.  $\{\Delta_{i,t}\}$  против  $\{d_{i,t}\}$ ,  $i \in V_t$  для узлов сетей полётов (Flight) за  $t = 121$  день (верхний слева) и сети обмена e-мейлами (Enron) за  $t = 220500$  слотов по 6,4 сек (нижний слева); изменение среднего кластерного коэффициента  $\bar{C}_t$ , см. (7), от времени  $t$  (средний столбец) и взвешенного числа треугольников  $W_n$ , см. (15), от числа узлов в подграфе сети  $n$  (правый столбец) для соответствующих сетей

## 8. Заключение

Исследованы распределения и зависимости числа связей и треугольников узлов в графах, развивающихся по моделям эволюции ПП, КП, КПП при различных стратегиях удаления узлов. Рассмотрена детекция геометрии случайного графа по величинам суммарного по всем узлам графа взвешенного числа треугольников, к которым узлы принадлежат.

В предложенной смешанной модели КПП новый узел присоединяется к существующим узлам с вероятностью, определяемой моделью КП или ПП в заданной пропорции.

Рассмотрены стратегии без удаления узлов и связей, с удалением одного «старого» узла с наименьшим пейджрангом и с удалением одного узла с вероятностью, обратно пропорциональной числу его связей, на каждом шаге эволюции. Узел удаляется со всеми его связями. При этих стратегиях удаления узлов определяется тяжесть хвоста распределения числа связей и треугольников

ков узлов, для чего оценивается ИЭВ (величина, обратная к ХИ). Изучена динамика среднего кластерного коэффициента и взвешенного числа треугольников во времени и зависимости числа связей узлов и числа треугольников, к которым узлы принадлежат, при разных стратегиях удаления узлов.

Главные выводы: а) удаление узлов снижает тяжесть хвоста распределений числа связей и треугольников; б) при удалении маловлиятельного узла с вероятностью (9) и эволюции без удаления узлов зависимости между числом связей и треугольников мало отличаются, в отличие от удаления маловлиятельных «старых» узлов, когда зависимость сильная при любой модели эволюции; в) смешанная модель КПП определяется доминирующей компонентой смеси; г) геометрия сети исчезает в ходе эволюции, если начальный граф геометрический, и происходит удаление узлов.

Детальные выводы:

– Без удалений узлов модель КПП приводит к потере тяжелых хвостов распределений, свойственных ПП, для числа связей, если в смешанной модели преобладает доля КП. Число взвешенных треугольников имеет тяжелый хвост при всех моделях эволюции, причем КПП усиливает тяжесть хвоста вплоть до отсутствия первого момента распределения. Индекс корреляции расстояний числа связей узлов и числа треугольников близок к единице для моделей эволюции КП и КПП с параметрами  $(\alpha, \epsilon) \in \{(0, 0), (1, 0)\}$ , что означает сильную зависимость этих сл.в.. Индекс понижен для ПП и КП (КПП) с параметрами  $(1, 1)$  из-за слабой зависимости. Средний кластерный коэффициент убывает во времени для всех моделей, т.е. тенденция к созданию кластеров, тесно связанных групп узлов с относительно высокой плотностью связей, уменьшается.

– Удаление самых «старых» узлов с наименьшими пейджрангами ведет к легким хвостам распределений для числа связей и треугольников для всех моделей, включая ПП, ввиду роста числа изолированных узлов. Имеется сильная зависимость числа связей и числа треугольников для всех моделей эволюции. Гео-

метрия начального графа в ходе эволюции не сохраняется.

– С удалением узлов с вероятностью, обратно пропорциональной числу их связей, зависимость числа треугольников от числа связей ослабевает по сравнению с двумя предыдущими случаями. Геометрия начального графа не сохраняется. Распределение числа связей имеет легкие хвосты при КП и КПП с доминированием КП, и тяжелые – при ПП и КПП с доминированием ПП. Хвост распределения числа взвешенных треугольников тяжелый, кроме случаев КП с параметрами  $(\alpha, \epsilon) = \{(0, 0), (1, 1)\}$  и КПП с преобладанием КП и с  $(\alpha, \epsilon) = (0, 0)$ , когда ИЭВ принимает отрицательные или близкие к нулю значения, и хвосты легкие.

Анализ реальных сетей Flight и Enron показал, что распределения числа связей узлов и треугольников в социальной сети Enron имеют тяжелые хвосты; в сети полетов Flight число связей имеет тяжелый хвост, а число треугольников – легкий. Зависимость между числом связей и треугольников сильная. Способность образовывать треугольники (средний кластерный коэффициент) несколько выше для Enron, чем для Flight. Поведение числа взвешенных треугольников и средних кластерных коэффициентов подтверждает геометрию сети, вызванную геолокацией узлов.

### Литература

1. AIELLO W., BONATO A., COOPER C., JANSSEN J., PRALAT P. *A spatial web graph model with local influence regions* // Internet Mathematics. – 2009. – No. 5. – P. 175–196.
2. ALBERT R., BARABÁSI A.-L. *Statistical mechanics of complex networks* // Rev. Mod. Phys.. – 2002. – Vol. 74. – P. 47–97.
3. ALBERT R., BARABÁSI A.-L. *Emergence of scaling in random networks* // Science. – 2002. – No. 286. – P. 509–512.
4. ARNOLD N.A., MONDRAGON R.J., CLEGG R.G. *Likelihood-based approach to discriminate mixtures of network models that vary in time* // Sci. Rep. – 2021. – No. 11. – P. 5205.

5. AVRACHENKOV K., LEBEDEV D. *PageRank of scale-free growing networks* // Internet Mathematics. – 2006. – Vol. 3, No. 2. – P. 207–231.
6. BARABÁSI A.-L., ALBERT R. *Statistical mechanics of complex networks* // Rev. Modern Phys. – 1999. – No. 74. – P. 47–97.
7. BAGROW J., BROCKMANN D. *Natural Emergence of Clusters and Bursts in Network Evolution* // Physical Review X. – 2012. – Vol. 3., No. 2. – P. 021016.
8. BEIRLANT J., GOEGEBEUR Y., TEUGELS J., SEGERS J. *Statistics of Extremes: Theory and Applications*. – Chichester, West Sussex: Wiley, 2004. –504 p.
9. BRIN S., PAGE L. *The anatomy of a large-scale hypertextual Web search engine* // Computer Networks and ISDN Systems. – 1998. – Vol. 30, No. 1–7. – P. 107–117.
10. BRINGMANN K., KEUSCH R., LENGLER J. *Geometric inhomogeneous random graphs* // Theoretical Computer Science. – 2019. – No. 760. – P. 35–54.
11. BOLLOBÁS B., RIORDAN O.M. *Mathematical Results on Scale-Free Random Graphs*. – Weinheim: Wiley-WCH, 2002.
12. CIRKOVIC D., TIANDONG WANG, RESNICK S.I. *Preferential attachment with reciprocity: properties and estimation* // Journal of Complex Networks. – 2023. – No. 11, – Issue. 5. – P. cnad031.
13. CHEN N., LITVAK N., OLVERA-CRAVIOTO M. *PageRank in Scale-Free Random Graphs* // WAW 2014, LNCS 8882, ed. A. Bonato et al. Switzerland: Springer. – 2014. – P. 120–131.
14. COHEN W.W. // <http://www.cs.cmu.edu/~enron/> (дата обращения: 17.04.2024).
15. DE HAAN L., FERREIRA A. *Extreme Value Theory: An Introduction*. – Springer Science and Business Media, 2006. – 417 p.
16. DEKKERS A.L.M., EINMAHL J.H.J., DE HAAN L. *A Moment Estimator for the Index of an Extreme-Value Distribution* // Ann. Statist. – 1989. – No. 17. – P. 1833–1855.

17. ESTRADA E. *The Structure of Complex Networks: Theory and Applications*. – Oxford, 2011; online edn, Oxford Academic, 2013.
18. FRAGA ALVES M.I., GOMES M.I., DE HAAN L. *Mixed moment estimator and location invariant alternatives // Extremes*. – 2009. – No. 12. – P. 149–185.
19. GHOSHAL G., CHI L., BARABASI A.L. *Uncovering the role of elementary processes in network evolution // Sci. Rep.* – 2013. – No. 3. – P. 2920.
20. ISKHAKOV L., KAMINSKI B., MIRONOV M. et al. *Clustering Properties of Spatial Preferential Attachment Model // In: Bonato, A., Pralat, P., Raigorodskii, A. (eds.) Algorithms and Models for the Web Graph. WAW 2018. Lecture Notes in Computer Science*. – 2018. – Vol. 10836. – P. 30–43.
21. JACOB E., MORTERS P. *A Spatial Preferential Attachment Model with Local Clustering // In: Bonato, A., Mitzenmacher, M., Pralat, P. (eds) Algorithms and Models for the Web Graph. WAW 2013. Lecture Notes in Computer Science*. – 2013. – Vol. 8305. – P. 14–25.
22. LAI Z., XIAO W., LI M., ZHANG Z. *An Exponential Distribution Complex Network Model Constructed by Degree Sequence Length Iteration // IEEE Int. Conf. on Computational Science and Engineering (CSE) and IEEE Int. Conf. on Embedded and Ubiquitous Computing (EUC) – 2017*. – P. 267–271.
23. MARKOVICH N.M. *Nonparametric Analysis of Univariate Heavy-Tailed Data: Research and Practice*. – Chichester, West Sussex: Wiley, 2007. – 343 p.
24. MARKOVICH N.M., VAIČIULIS M. *Extreme Value Statistics for Evolving Random Networks // Mathematics*. – 2023. – Vol. 11, No. 9. – P. 2171.
25. MARKOVICH N.M., VAIČIULIS M. *Investigation of triangle counts in graphs evolved by uniform clustering attachment // arXiv: 2401.11548v1*. – 2024. – P. 1–16.
26. MARKOVICH N.M., RYZHOV M.S., VAIČIULIS M.

- Inferences for Random Graphs Evolved by Clustering Attachment* // arXiv: 2403.00551v1. – 2024. – P. 1–25.
27. MICHELAN R., LITVAK N., STEGEHUIS C. *Detecting hyperbolic geometry in networks: why triangles are not enough* // Phys. Rev. E. – 2022. – Vol. 106, No. 5. – P. 054303.
  28. NORROS I., REITTU H. *On a conditionally poissonian graph process* // Advances in Applied Probability. – 2006. – No. 38. – P. 59–75.
  29. PENROSE M. *Random Geometric Graphs*. – Oxford Studies in Probability: Oxford Academic, 2003.
  30. POURSAFAEI F., HUANG S., PELRINE K. et al. *Towards better evaluation for dynamic link prediction* // Advances in Neural Information Processing Systems. – 2022. – Vol. 35. – P. 32928–32941.
  31. RAMOS-CARRENO C., TORRECILLA J.L. *dcor: Distance correlation and energy statistics in Python* // SoftwareX. – 2023. – Vol. 22. – P. 101326.
  32. ROSSI R.A., AHMED N.K. *The Network Data Repository with Interactive Graph Analytics and Visualization* // Proc. of the AAAI Conf. on Artificial Intelligence. – 2015. – Vol. 29, No. 1. – P. 4292–4293.
  33. STROHMEIER M., OLIVE X., LUBBE J. et al. *Crowdsourced air traffic data from the OpenSky network 2019–20* // Earth System Science Data Discussions. – 2020. – No. 2020. – P. 1–15.
  34. WAN P., WANG T., DAVIS R. A., RESNICK S.I. *Are extreme value estimation methods useful for network data?* // Extremes. – 2020. – No. 23. – P. 171–195.
  35. WANG T., RESNICK S.I. *Consistency of Hill estimators in a linear preferential attachment model* // Extremes. – 2019. – No. 22. – P. 1–28.
  36. WANG T., RESNICK S.I. *2RV+HRV and Testing for Strong VS Full Dependence* // arXiv:2312.16332v1 [math.ST]. – 2023. – P. 1–46.
  37. WEI-BING D., GUO L., LI W. et al. *Worldwide*



*Marine Transportation Network: Efficiency and Container Throughput // Chinese Physics Letters. – 2009. – No. 26. – P. 118901.*

## **INVESTIGATION OF DEPENDENCIES AND DISTRIBUTIONS IN RANDOM NETWORKS EVOLVED BY MIXED MODELS OF THE EVOLUTION AND NODE DELETION**

**Natalia Markovich**, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Doctor of Science (nat.markovich@gmail.com).

**Maksim Ryzhov**, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, researcher (maksim.ryzhov@frtk.ru)

**Michail Kulik**, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, student (mishakulik2002@yandex.ru).

*Abstract: The evolution of a random network by models of preferential, clustering and mixed attachments to form links between newly appending nodes and existing nodes is studied. Strategies of node deletion at each step of network evolution are considered: 1) without node and edge deletion; 2) deletion the least influential node among the most 'old', where the node's PageRank is used as a measure of the node's influence; 3) deletion a node with a probability inversely proportional to the node degree. For these deletion strategies the dependence of two characteristics of random networks, namely, the node degrees and node triangle counts (that is, the triples of interconnected nodes in which the node is involved), and the behavior of clustering coefficients of nodes are compared by simulation. The heaviness of the distribution tails for the node degrees and the node triangle counts is estimated. The mixed clustering-preferential attachment is proposed here for the first time.*

Keywords: network evolution, preferential and clustering attachment, mixed clustering-preferential attachment, node degree, triangle counts, node deletion, clustering coefficient.

УДК 519.24; 519.25

ББК 22.172

DOI: 10.25728/ubs.2024.109.5

*Статья представлена к публикации  
членом редакционной коллегии А.А. Галяевым.*

*Поступила в редакцию 06.02.2024.*

*Дата опубликования 31.05.2024.*