

# О МОДЕЛЯХ ПОСТРОЕНИЯ ГРАФА ВЗАИМОДЕЙСТВУЮЩИХ ОБЪЕКТОВ В СЕТИ TELEGRAM-КАНАЛОВ

Попов В. А.<sup>1</sup>, Чеповский А. А.<sup>2</sup>

DOI: 10.21681/2311-3456-2024-3-105-112

**Цель исследования:** сравнение широкого набора различных моделей построения графов взаимодействующих объектов в сети публичных Telegram-каналов с целью выявления среди них наиболее подходящих, при которых полученный граф наиболее близок к безмасштабным сетям.

**Метод исследования:** для построенных взвешенных графов в рамках каждой из рассматриваемых моделей находятся степенные законы, наиболее приближающие эмпирические распределения полученных весов вершин, после чего оценивается качество полученного приближения.

**Полученный результат:** в статье представлены модели построения графов, характеризующих информационное воздействие в сети Telegram-каналов. В данной работе представлены результаты исследования 180 случаев – для 12 моделей проведены исследования на 15 наборах данных. В рамках этих исследований найдены параметры степенных законов, приближающих эмпирические данные. Показано, у каких из моделей эти параметры оказываются не свойственными для безмасштабных сетей. С помощью критерия Колмогорова проверены гипотезы о характере распределения у моделей. Приведены иллюстрации, наглядно показывающие результаты исследования. Показано, какая из моделей лучше всего подходит для формирования графов взаимодействующих объектов в сети Telegram-каналов. Такие графы могут быть впоследствии проанализированы с целью выделения ключевых вершин.

**Научная новизна:** предложены модели для представления имевшего место взаимодействия объектов сети Telegram-каналов в виде взвешенных графов. Исследовано распределение весов вершин у полученных графов взаимодействующих объектов. Изучение этого важного свойства для взвешенных графов, полученных при импорте данных из реальных сетей, дало важный теоретический и практический результат. Выявлено, что UMR-модель построения таких графов, обладает свойством, характерным для безмасштабных сетей.

**Ключевые слова:** безмасштабные сети, модель информационного воздействия, выделение сообществ, анализ социальных сетей, критерий согласия Колмогорова, степенной закон распределения, вес вершин.

## ABOUT MODELS TO CONSTRUCT A GRAPH OF INTERACTING OBJECTS IN A NETWORK OF TELEGRAM CHANNELS

Popov V. A.<sup>3</sup>, Chepovskiy A. A.<sup>4</sup>

**The purpose of the study:** comparison of a wide range of different models to construct graphs of interacting objects in a public Telegram channels network in order to identify among them the most suitable ones, in which the resulting graph is closest to scale-free networks.

**Method:** for the constructed weighted graphs, within the framework of each of the models under consideration, power laws are found that most closely approximate the empirical distributions of the obtained vertices weights, after which the quality of the resulting approximation is assessed.

- 1 Попов Владимир Александрович, аспирант Департамента прикладной математики МИЭМ НИУ ВШЭ, Москва, Россия. E-mail: vapopov@hse.ru
- 2 Чеповский Александр Андреевич, кандидат физико-математических наук., доцент, Департамент прикладной математики МИЭМ НИУ ВШЭ, Москва, Россия. E-mail: aachepovsky@hse.ru
- 3 Vladimir A. Popov, Ph.D. student, School of Applied Mathematics, HSE MIEM, Moscow, Russia. E-mail: vapopov@hse.ru
- 4 Alexander A. Chepovskiy, Ph.D. in Physics and Mathematics, Associate Professor, Department of Applied Mathematics, Moscow Institute of Economics, National Research University Higher School of Economics, Moscow, Russia. E-mail: aachepovsky@hse.ru

**Results:** the article presents models to construct graphs that characterize the information impact in the Telegram channels network. This paper presents the results of a study of 180 cases – studies were conducted for 12 models on 15 data sets. As part of these studies, parameters of power laws that approximate empirical data were found. It is shown which of the models have these parameters that are not characteristic of scale-free networks. Using the Kolmogorov criterion, hypotheses about the nature of the distribution of the models were tested. Illustrations are provided to clearly show the results of the study. It is shown which of the models is best suited to construct graphs of interacting objects in a network of Telegram channels. Such graphs can subsequently be analyzed to identify key vertices.

**Scientific novelty:** models are proposed to represent the interaction of objects in the Telegram channel network in the form of weighted graphs. The distribution of vertex weights in the resulting graphs of interacting objects has been researched. Studying this important property for weighted graphs obtained by importing data from real networks has yielded important theoretical and practical results. It was revealed that the UMR-model to construct such graphs has a property characteristic of scale-free networks.

**Keywords:** scale-free networks, model of information impact, community detection, analysis of social networks, Kolmogorov goodness-of-fit test, power law distribution, vertex weight.

## 1. Введение

В науке много лет известно понятие сложных сетей (complex networks) – биологических, технологических, телекоммуникационных, социальных сетей, содержащих большое число объектов и относительно малое число связей между ними. Такие сети, возникающие на практике, относятся к так называемым безмасштабным сетям (free-scale networks)<sup>5</sup>. Анализ графов, характеризующих безмасштабные сети является предметом подробных исследований многих авторов последних двух десятилетий [1, 2], построены различные алгоритмы, позволяющие выделять на графе подграфы, именуемые неявными сообществами, а также ключевые вершины графа, используя специализированные центральности [3, 4].

В контексте практических исследований для телекоммуникационных сетей важным является импорт реальных данных и построение на их основе графов взаимодействующих объектов – графов, характеризующих наличие коммуникации между объектами исходной сети или степень ее интенсивности, а также метаданные о содержании этого взаимодействия, включая текстовую информацию. Одной из актуальных для анализа данных является сеть Telegram-каналов. Помимо обмена личными сообщениями между пользователями в Telegram реализована функциональность для организации публичных Telegram-каналов, представляющих собой информационно-новостные ленты сообщений. Многие СМИ, информационные сообщества, блогеры имеют свои Telegram-каналы и регулярно публикуют в них контент, а пользователи Telegram могут подписаться на данные каналы и получать информацию в виде сообщений от имени канала.

При этом Telegram предоставляет широкий спектр инструментов для ведения каналов. Помимо уникального текстового и фото-видео контента, администратор канала может опубликовать пост другого канала в своем (сделать репост), процитировать или упомянуть другие Telegram-каналы, сделать ссылку на внешний адрес в сети интернет. Использование данных возможностей создает связи между Telegram-каналами, что позволяет рассматривать их сеть для импорта данных и построения взвешенного графа взаимодействующих объектов. Так, авторами в предыдущих работах была представлена  $(U, M, R)$ -модель [5] и алгоритм для последующего выделения на построенном с ее помощью графе неявных сообществ [6].

В практическом плане одной из актуальных задач является выбор модели при построении графа, ибо в зависимости от этого выбора дальнейший анализ, основанный в том числе на весе ребер и вершин, будет давать разный результат. Поэтому важно выбрать модель, при которой построенный граф обладает свойствами безмасштабных сетей.

В данной работе в первой части описывается общая модель построения графов взаимодействующих объектов и 12 моделей, при которых по-разному определяется вес на ребрах графа. После этого показано, как по набору весов вершин графа определить параметры степенного распределения, приближающего данный набор значений. На основе полученного приближения сравниваются рассматриваемые модели. Далее демонстрируются примеры приближения эмпирических данных степенными функциями, для них строятся и сравниваются функции распределения и функции плотности вероятности.

5 Fortunato S. Community Detection in Graphs // Physics Reports. – 2010. – 486(3). – P. 75–174.

Евин И. А. Введение в теорию сложных сетей // Компьютерные исследования и моделирование. – 2010. – 2(2). – С. 121–141.

## 2. Модели построения графа и наборы данных

В данной работе рассматриваются публичные каналы в мессенджере Telegram, для которых можно выделить следующие ключевые факторы взаимодействия: репосты между каналами, упоминания одного канала другим и наличие общих внешних URL в постах двух каналов. Для построения таких моделей изначально необходимо импортировать данные из мессенджера за выбранный период времени, а именно данные о постах, интересующих Telegram-каналов. С этой целью было разработано программное обеспечение, использующее официальный API Telegram, способное импортировать информацию о каналах в данные специального AVS-формата [6, 7], содержащие все необходимые составляющие для дальнейшего анализа и выявления взаимодействий между каналами. Для построения графа взаимодействующих объектов первично формируется полный граф  $G(V,E)$ , содержащий все импортированные вершины, соответствующие каналам, и все возможные ребра между ними. Далее осуществляется переход к взвешенному графу  $G(V,\tilde{E})$ , где на ребрах задан вес, определенный исходя из выявленного взаимодействия между соответствующими вершинами. В случае нулевого веса ребра удаляются. При этом исходный импорт данных и построение множества  $V$  устроены таким образом, что граф  $G(V,\tilde{E})$  представляет из себя одну компоненту связности.

Зададим на исходном множестве ребер  $E$  весовую функцию  $w$ , зависящую от выявленных факторов взаимодействия  $\delta_{e_{AB}}^U, \delta_{e_{AB}}^M, \delta_{e_{AB}}^R$ :

$$w(e_{AB}) = F(\delta_{e_{AB}}^U, \delta_{e_{AB}}^M, \delta_{e_{AB}}^R), \quad (1)$$

где  $\delta_{e_{AB}}^U$  – количество общих уникальных внешних ссылок (URL) в постах у каналов  $A$  и  $B$  за выбранный период;  $\delta_{e_{AB}}^M$  – количество постов, где в тексте канал  $A$  упомянул канал  $B$  плюс количество постов, где  $B$  упомянул  $A$  за выбранный период (для каждого поста смотрятся уникальные упоминания, то есть если в одном посте канал  $A$  упомянул несколько раз канал  $B$ , то это упоминание все равно учитывается единожды в данном коэффициенте;  $\delta_{e_{AB}}^R$  – количество репостов каналом  $A$  сообщений канала  $B$ , плюс количество репостов каналом  $B$  сообщений канала  $A$  за выбранный период;  $F$  – функция, зависящая от  $\delta_{e_{AB}}^U, \delta_{e_{AB}}^M, \delta_{e_{AB}}^R$ , возвращает неотрицательные значения.

Различные модели определяются тем, какую функцию  $F$  выбрать при определении веса ребра между вершинами. Далее рассмотрим 12 моделей, соответствующих разным функциям  $F$ . Первые две из них являются  $(U,M,R)$ -моделью [5] с различающимися значениями параметров  $U, M$  и  $R$  и представляют из себя линейные функции:

$$F_{1,1} = 1 \cdot \delta_{e_{AB}}^U + 2 \cdot \delta_{e_{AB}}^M + 3 \cdot \delta_{e_{AB}}^R \quad (2)$$

$$F_{1,2} = 1 \cdot \delta_{e_{AB}}^U + 1 \cdot \delta_{e_{AB}}^M + 1 \cdot \delta_{e_{AB}}^R \quad (3)$$

Формулы (2) и (3) означают, что для  $F_{1,1}$  в базовой  $(U,M,R)$ -модели взяты значения  $U = 1, M = 2, R = 3$ . А для  $F_{1,2}$  взяты значения  $U = M = R = 1$ . Помимо них рассмотрим более сложные конструкции, не подпадающие под  $(U,M,R)$ -модель, но основанные на тех же факторах взаимодействия как аргументах для функций  $F$ .

Для второй группы рассматриваемых функций используем логарифмическую составляющую:

$$F_{2,1} = 1 \cdot \ln(\delta_{e_{AB}}^U + 1) + 1 \cdot \ln(\delta_{e_{AB}}^M + 1) + 1 \cdot \ln(\delta_{e_{AB}}^R + 1) \quad (4)$$

$$F_{2,2} = 1 \cdot \ln(\delta_{e_{AB}}^U + 1) + 2 \cdot \ln(\delta_{e_{AB}}^M + 1) + 3 \cdot \ln(\delta_{e_{AB}}^R + 1) \quad (5)$$

Добавим в анализируемый набор варианты для  $F$ , у которых есть компоненты из показательных функций. В указанных ниже функциях под  $\max \delta_e^U, \max \delta_e^M, \max \delta_e^R$  мы будем понимать максимальные по всему графу  $G(V,E)$  значения для  $\delta_e^U, \delta_e^M, \delta_e^R$  соответственно:

$$F_{3,1} = 2 \frac{\delta_{e_{AB}}^U}{\max \delta_e^U} + 2 \frac{\delta_{e_{AB}}^M}{\max \delta_e^M} + 2 \frac{\delta_{e_{AB}}^R}{\max \delta_e^R} - 3 \quad (6)$$

$$F_{3,2} = 2 \frac{\delta_{e_{AB}}^U}{\max \delta_e^U} + 2^2 \frac{\delta_{e_{AB}}^M}{\max \delta_e^M} + 2^3 \frac{\delta_{e_{AB}}^R}{\max \delta_e^R} - 3 \quad (7)$$

$$F_{3,3} = 2 \frac{\delta_{e_{AB}}^U}{\max \delta_e^U} + 2 \cdot 2 \frac{\delta_{e_{AB}}^M}{\max \delta_e^M} + 3 \cdot 2 \frac{\delta_{e_{AB}}^R}{\max \delta_e^R} - 6 \quad (8)$$

И рассмотрим еще несколько вариантов для функции  $F$ , скомбинированных из предыдущих:

$$F_{4,1} = \delta_{e_{AB}}^U + \ln(\delta_{e_{AB}}^M + 1) \cdot \ln(\delta_{e_{AB}}^R + 1) \quad (9)$$

$$F_{4,2} = (\delta_{e_{AB}}^U + 1) \cdot \ln(\delta_{e_{AB}}^M + 1) + \ln(\delta_{e_{AB}}^R + 1) \quad (10)$$

$$F_{4,3} = (\delta_{e_{AB}}^U + 1) \cdot \ln(\delta_{e_{AB}}^R + 1) + \ln(\delta_{e_{AB}}^M + 1) \quad (11)$$

$$F_{4,4} = (\delta_{e_{AB}}^U + 1) \cdot (\ln(\delta_{e_{AB}}^U + 1) + \ln(\delta_{e_{AB}}^R + 1)) \quad (12)$$

$$F_{4,5} = (\delta_{e_{AB}}^U + 1) \cdot (2 \cdot \ln(\delta_{e_{AB}}^U + 1) + 3 \cdot \ln(\delta_{e_{AB}}^R + 1)) \quad (13)$$

Для сравнения моделей, полученных на основе описанных выше весовых функций, были скачаны данные по 15 сетям. Импорт данных для каждой из них проводился от некоторых стартовых вершин наперед заданную глубину по фиксированному для каждого из случаев временному периоду. Как и было описано в начале раздела, далее на основе этих импортированных данных были сформированы графы  $G_i(V,E)$ . До применения функции  $F$  и перехода к графам  $G_i(V,\tilde{E})$  можно посмотреть на характеристики скачанных подсетей с учетом выявленных факторов взаимодействия и ребер, вес на которых будет ненулевым. Они едины для всех весовых функций, применяемых к  $G_i(V,E)$  и зависят именно от отсутствия имевшего место взаимодействия. Так же можно подсчитать для таких графов диаметр и классический коэффициент кластеризации, они не будут зависеть от весов ребер, а только от наличия ненулевых ребер. Такие данные приведены в табл. 1.

Характеристики импортированных подсетей

Граф	$N$ – количество вершин в графе $G_i(V, \bar{E})$	$M$ – количество ненулевых ребер в графе $G_i(V, \bar{E})$	Диаметр графа	Коэффициент кластеризации графа
$G_1$	179	723	6	0,29
$G_2$	185	919	4	0,31
$G_3$	302	2641	4	0,31
$G_4$	437	4892	4	0,30
$G_5$	451	3580	6	0,27
$G_6$	464	1921	6	0,24
$G_7$	590	4352	9	0,22
$G_8$	600	18009	4	0,32
$G_9$	619	2973	7	0,22
$G_{10}$	625	6137	6	0,20
$G_{11}$	773	6611	6	0,18
$G_{12}$	1252	17841	4	0,28
$G_{13}$	1432	44595	4	0,24
$G_{14}$	1592	34877	4	0,19
$G_{15}$	3736	93215	6	0,21

Далее на ребрах этих графов применяются весовые функции  $F$ , описанные ранее. За счет чего получаем графы  $G_i(V, \bar{E})$ , с которыми и будут проведены дальнейшие действия. Фактически для каждой из 15 сетей мы получили по 12 различных взвешенных графов и общий набор данных составил 180 взвешенных графов. Вес вершины определяется как сумма весов инцидентных ей ребер.

### 3. Распределение весов вершин графов

В работах, посвященных безмасштабным сетям<sup>6</sup>, полученным на основе взаимодействия социальных объектов, показывается, что распределение весов вершин подчиняется степенному закону с параметром степени  $\alpha \in [2; 3]$ :

$$f(x) = P(X = x) = Cx^{-\alpha} \quad (14)$$

Это свойство используется в методах генерации случайных сетей при моделировании сложных сетей [8, 9].

На рассматриваемом наборе графов сравним эмпирические значения весов вершин и получаемые распределения для определения лучшей весовой функции. Под весом вершины понимаем сумму весов всех ребер, инцидентных этой вершине. Для каждого

6 Newman M. E. J. Networks: An Introduction. – Oxford University Press, 2010. – 784 p.  
G. Caldarelli, Scale-Free Networks. Oxford University Press, Oxford (2007).

Таблица 1.

из 180 рассматриваемых графов будем приближать эмпирические распределения весов вершин  $\{x_{ij}\}_{i=1}^N$  функциями со степенным законом, находить оптимальный параметр степени  $\alpha$  и считать соответствующую ошибку – насколько хорошо распределения приближают эмпирические данные. Для нахождения параметров степенных распределений, которые наиболее точно описывают полученные эмпирические распределения весов вершин, воспользуемся методом максимального правдоподобия и критерием согласия Колмогорова<sup>7</sup>.

Степенной закон, указанный в формуле (14), у безмасштабных сетей имеет место для «хвоста» распределения, начиная с некоторого значения  $x_{min}$ . Будем приближать не все эмпирическое распределение, а только часть, для которой выполняется условие:

$$x_i > x_{min} \quad (15)$$

Это означает, что реально рассматривается не вся выборка  $\{x_{ij}\}_{i=1}^N$ , а только ее часть  $\{x_{ij}\}_{i=1}^k$ , которая удовлетворяет условию (15). При таком подходе возможна ситуация, что эмпирические данные приближаются степенной функцией наилучшим образом при слишком высоком значении  $x_{min}$ . А это означает, что число вершин  $k$ , для которых будет рассматриваться выборка  $\{x_{ij}\}_{i=1}^k$  мало. Поэтому введем еще одно ограничение, которое позволяет рассматривать содержательно существенное число вершин. А именно, будем рассматривать такие  $x_{min}$ , что количество вершин  $k$  удовлетворяет условию:

$$k > 0,2 \cdot N \quad (16)$$

Теперь перейдем к рассмотрению  $\{x_{ij}\}_{i=1}^k$ . Для начала с помощью метода максимального правдоподобия находится показатель степени распределения  $\hat{\alpha}$ , считая, что значение  $x_{min}$  известно. Данный метод доказуемо дает точные оценки параметров моделей для выборок большого размера<sup>8</sup>. После этого, минимизируя расстояние Колмогорова-Смирнова,

7 Clauset A, Shalizi CR, Newman MEJ Power-law distributions in empirical data. SIAM Review 51: 661–703 (2009).

8 Wasserman, L., All of Statistics: A Concise Course in Statistical Inference (Springer-Verlag, Berlin), 2003.  
Barndorff-Nielsen, O. E., and D. R. Cox, Inference and Asymptotics (Chapman and Hall, London), 1995.

Таблица 2

Полученные значения  $\hat{\alpha}$  для  $x_{min}$  с учетом ограничения (16)

	$G_1$	$G_2$	$G_3$	$G_4$	$G_5$	$G_6$	$G_7$	$G_8$	$G_9$	$G_{10}$	$G_{11}$	$G_{12}$	$G_{13}$	$G_{14}$	$G_{15}$
$F_{1,1}$	2,15	2,07	2,36	2,03	2,00	2,06	2,25	2,31	2,22	2,31	2,09	2,35	2,01	1,95	2,00
$F_{1,2}$	1,81	2,05	2,48	2,20	2,08	2,16	2,19	2,34	2,22	2,35	1,83	2,26	2,06	1,99	2,13
$F_{2,1}$	2,38	2,11	2,96	2,82	2,26	2,26	2,83	2,46	2,03	2,25	2,52	1,96	2,17	2,45	2,41
$F_{2,2}$	2,92	2,32	2,87	2,65	2,47	2,43	2,74	2,75	2,64	2,48	2,47	2,64	2,24	2,41	2,33
$F_{3,1}$	1,82	2,37	2,69	1,72	1,94	2,31	2,21	2,25	2,20	2,11	2,12	2,04	2,05	2,02	2,26
$F_{3,2}$	1,87	2,11	2,33	1,84	2,06	2,14	2,24	2,22	1,72	2,30	2,08	2,12	2,20	2,09	2,02
$F_{3,3}$	2,15	2,20	2,59	1,87	2,07	2,14	2,29	1,88	1,70	2,27	2,13	2,13	2,21	1,99	2,03
$F_{4,1}$	2,16	2,62	2,79	1,79	1,97	2,40	2,68	2,53	2,44	2,29	2,12	2,10	2,20	2,14	2,24
$F_{4,2}$	1,85	2,07	2,25	2,11	2,09	2,14	2,16	2,34	2,31	2,21	2,25	2,26	2,03	2,13	2,02
$F_{4,3}$	1,96	2,07	2,31	2,21	2,13	1,95	2,09	2,51	2,48	2,38	2,25	2,42	2,06	2,24	1,98
$F_{4,4}$	1,73	1,91	2,04	2,02	1,98	1,99	2,70	2,30	2,27	2,20	2,16	2,22	2,02	2,21	1,92
$F_{4,5}$	2,06	1,92	1,96	2,04	2,02	1,91	2,68	2,36	2,25	2,41	2,18	2,23	2,01	2,22	1,90

выбирается значение  $x_{min}$ . С учетом условия (16) это будет не абсолютный минимум  $x_{min}$  для рассматриваемого графа, а локальный в этом интервале.

Возьмем графы, полученные для функций (2)–(13), и посмотрим на значение расстояния Колмогорова-Смирнова для оценки разницы между эмпирической функцией распределения вершин и найденной степенной функцией. Для каждого графа  $G_i(V, \hat{E})$  и каждой из функций подсчитана статистика Колмогорова [10]:

$$D_k = \sup_{x \geq x_{min}} |f(x) - \hat{f}_k(x)| \quad (17)$$

где  $f(x)$  – найденная степенная функция распределения, а  $\hat{f}_k(x)$  – эмпирическая функция распределения,  $k$  – размер соответствующей выборки (число рассматриваемых вершин).

Минимизируя  $D_k$  и находя так локальный минимум  $x_{min}$ , находится и значение  $\hat{\alpha}$ . Полученные значения  $\hat{\alpha}$  для разных весовых функций на исследуемом наборе графов с учетом описанных ранее условий представлены в таблице 2. Как видно, не во всех случаях выполнено  $\hat{\alpha} \in [2; 3]$ .

Используем критерий согласия Колмогорова для проверки на уровне значимости в 10% гипотезы о том, что эмпирически полученный для каждой из весовых функций набор весов вершин соответствует степенному закону распределения. Для этого проверим, что выполняется следующее условие:

$$\sqrt{k} D_k < z_{0,9}, \quad (18)$$

где  $z_{0,9} = 1,22$  – квантиль распределения Колмогорова для уровня значимости 10%.

Таблица 3

Результаты проверки критерия Колмогорова

	$G_1$	$G_2$	$G_3$	$G_4$	$G_5$	$G_6$	$G_7$	$G_8$	$G_9$	$G_{10}$	$G_{11}$	$G_{12}$	$G_{13}$	$G_{14}$	$G_{15}$
$F_{1,1}$	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ЛОЖЬ	ЛОЖЬ
$F_{1,2}$	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ЛОЖЬ	ИСТИНА	ЛОЖЬ	ИСТИНА	ЛОЖЬ	ЛОЖЬ	ЛОЖЬ	ИСТИНА	ЛОЖЬ
$F_{2,1}$	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ЛОЖЬ	ИСТИНА	ЛОЖЬ	ЛОЖЬ	ЛОЖЬ	ИСТИНА	ЛОЖЬ	ЛОЖЬ	ИСТИНА	ЛОЖЬ
$F_{2,2}$	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ЛОЖЬ	ЛОЖЬ	ИСТИНА	ЛОЖЬ
$F_{3,1}$	ИСТИНА	ИСТИНА	ИСТИНА	ЛОЖЬ	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ЛОЖЬ	ИСТИНА	ИСТИНА	ЛОЖЬ
$F_{3,2}$	ИСТИНА	ИСТИНА	ИСТИНА	ЛОЖЬ	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ЛОЖЬ	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ЛОЖЬ
$F_{3,3}$	ИСТИНА	ИСТИНА	ИСТИНА	ЛОЖЬ	ИСТИНА	ИСТИНА	ИСТИНА	ЛОЖЬ	ЛОЖЬ	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ЛОЖЬ	ЛОЖЬ
$F_{4,1}$	ИСТИНА	ИСТИНА	ИСТИНА	ЛОЖЬ	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ЛОЖЬ
$F_{4,2}$	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ЛОЖЬ	ИСТИНА	ИСТИНА	ЛОЖЬ	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ЛОЖЬ
$F_{4,3}$	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ЛОЖЬ	ЛОЖЬ	ИСТИНА	ЛОЖЬ	ИСТИНА	ИСТИНА	ИСТИНА	ЛОЖЬ	ИСТИНА	ЛОЖЬ
$F_{4,4}$	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ЛОЖЬ	ИСТИНА	ЛОЖЬ
$F_{4,5}$	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ИСТИНА	ЛОЖЬ	ИСТИНА	ЛОЖЬ

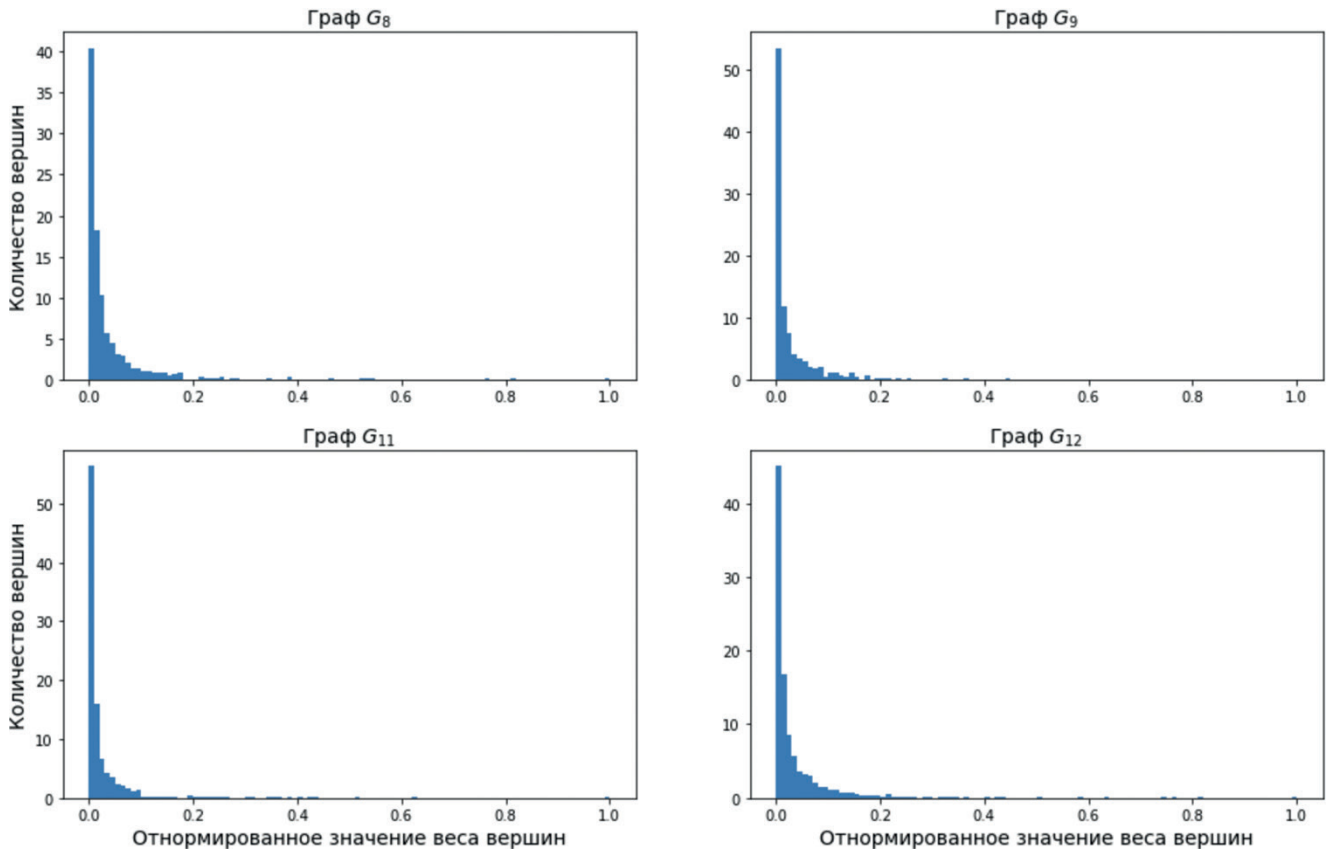


Рис 1. Гистограммы распределения весов вершин на графах  $G(V, E)$  для  $G_8, G_9, G_{11}, G_{12}$  при весовой функции  $F_{1,1}$ .

В таблице 3 представлены результаты проверки критерия. Из таблицы видно, что получаемый для весовой функции  $F_{1,1}$  набор весов вершин соответствует степенному распределению с найденным параметром  $\hat{\alpha}$ . Это подтверждает целесообразность использования  $F_{1,1}$  при построении взвешенных графов взаимодействующих объектов, обладающих свойствами безмасштабных сетей.

#### 4. Примеры приближения эмпирических данных степенной функцией

Убедимся еще и визуально, что полученные распределения весов вершин графов подчиняются степенному закону. Для этого на примере функции  $F_{1,1}$  и графов  $G_8, G_9, G_{11}, G_{12}$  из анализируемого набора построим гистограммы, представляющие собой частотные распределения (рис. 1). Визуально распределения весов вершин для функции  $F_{1,1}$  удовлетворяет ожиданиям.

Далее для рассматриваемых четырех эмпирических распределений весов вершин найдем функции плотности вероятности для выборки  $\{x_{ij}\}_{i=1}^k$ , удовлетворяющие условиям (15) и (16), и построим их на логарифмических осях (рис. 2). Также на этих графиках построим соответствующие плотности вероятности теоретических степенных распределений, параметры которых найдены в разделе 3 при помощи метода максимального правдоподобия:

- Для графа  $G_8$  – степенное распределение с параметром  $\alpha = 2.31$
- Для графа  $G_9$  – степенное распределение с параметром  $\alpha = 2.22$
- Для графа  $G_{11}$  – степенное распределение с параметром  $\alpha = 2.09$
- Для графа  $G_{12}$  – степенное распределение с параметром  $\alpha = 2.35$

Графиком плотности вероятности степенного распределения, построенном на логарифмических осях, является прямая. Как видно из рис. 2, значения плотности распределения весов вершин графов хорошо аппроксимируются прямыми – плотностями степенных распределений с найденными в разделе 3 параметрами. Поэтому можно сделать вывод, что полученные распределения эмпирических данных подчиняются степенным законам с указанными параметрами  $\alpha$ .

Также построим функции распределения для рассматриваемых распределений весов вершин графов  $G_8, G_9, G_{11}, G_{12}$  и найденных теоретических степенных распределений (рис. 3). Как видно из графиков, эмпирические функции распределения также хорошо приближаются найденными теоретическими. Именно максимумы разниц между этими эмпирическими функциями распределения вершин и найденными

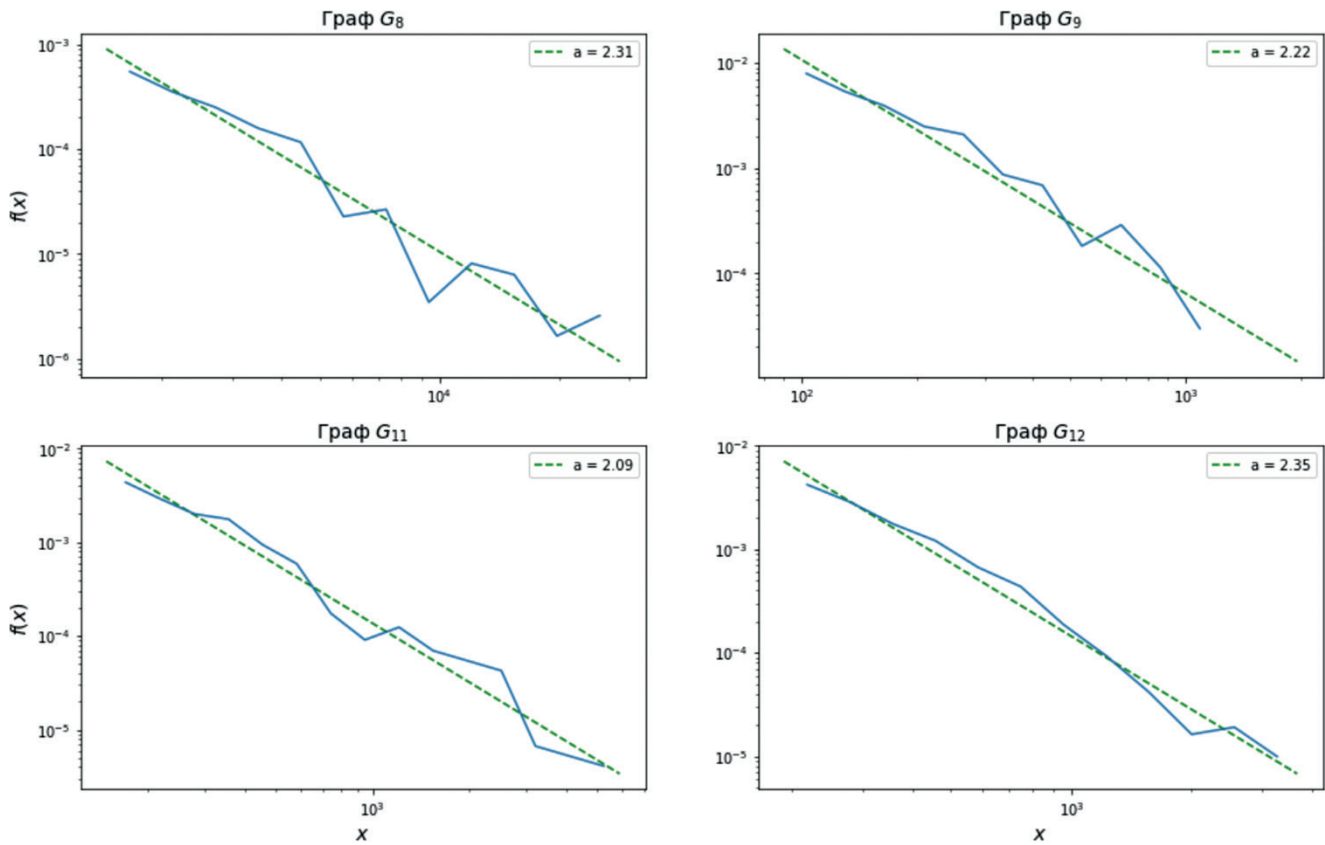


Рис. 2. Синие линии – графики плотности вероятности распределения весов вершин для графов  $G_8, G_9, G_{11}, G_{12}$  при весовой функции  $F_{1,1}$ ; Зеленые пунктирные линии – графики плотности вероятности степенных распределений с указанными параметрами  $\alpha$ .

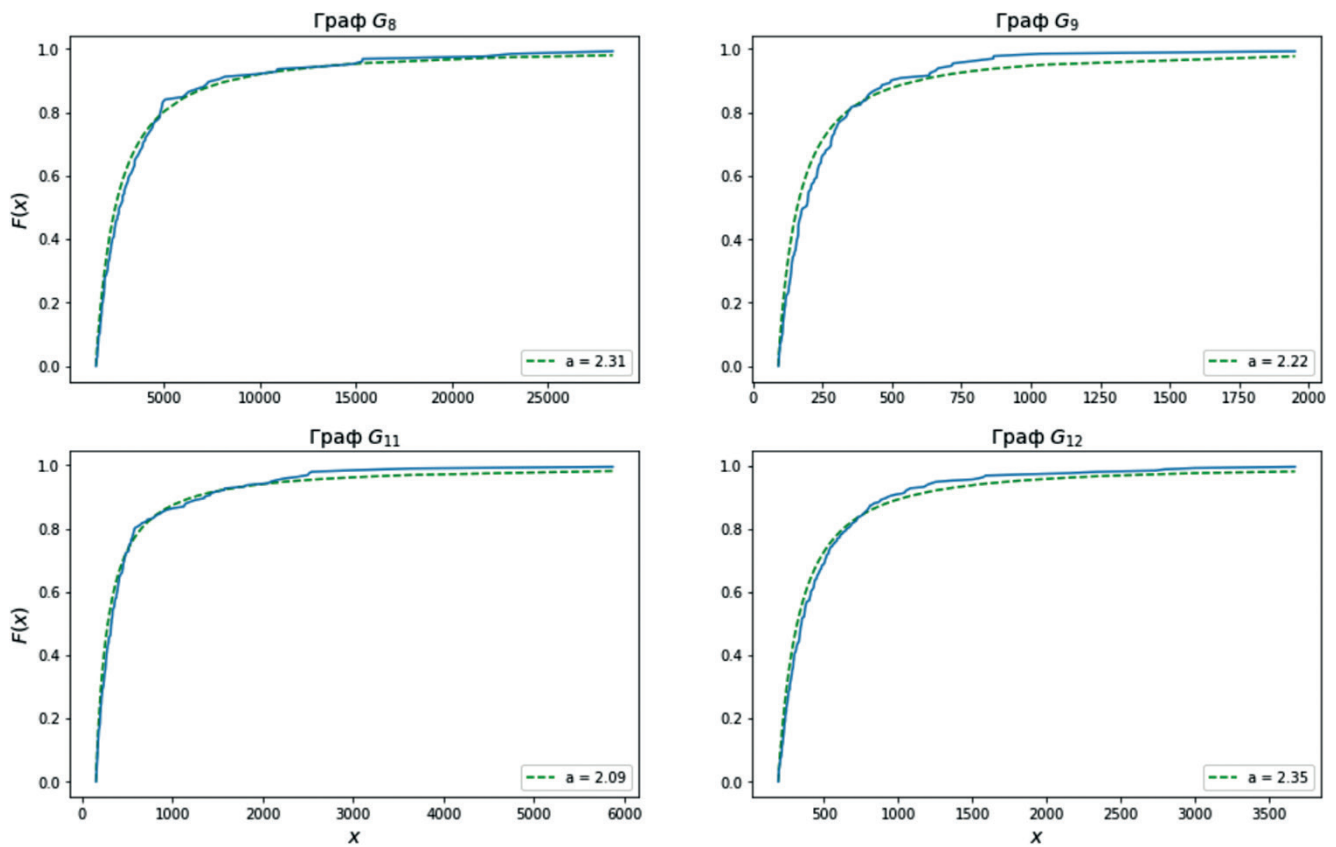


Рис. 3. Синие линии – графики функций распределений весов вершин на графах  $G(V, E)$  для  $G_8, G_9, G_{11}, G_{12}$  при весовой функции  $F_{1,1}$ ; Зеленые пунктирные линии – графики функций распределений степенных распределений с указанными параметрами  $\alpha$ .

степенными функциями являются значениями статистик Колмогорова (17), посчитанных в 3 разделе.

Таким образом, наглядно видно, что эмпирические распределения весов вершин для графов  $G_8$ ,  $G_9$ ,  $G_{11}$ ,  $G_{12}$  и функции  $F_{1,1}$  действительно являются степенными. А сравнение эмпирических и теоретических функций распределения и функций плотности вероятности демонстрирует подтверждение гипотезы о принадлежности выборок весов вершин степенным распределениям с зафиксированными параметрами  $\alpha$ .

## 5. Выводы

В данной статье приведено исследование, посвященное вопросу выбора модели для построения взвешенных графов взаимодействующих объектов сети публичных Telegram-каналов. Были исследованы 12 весовых функций, используемых

для построения таких графов. Тестирование проведено для импортированных 15 подсетей, на основе которых были построены 180 графов.

Результаты показывают, что графы, построенные с помощью весовой функции, которая соответствует  $(U, M, R)$ -модели в основном имеют распределение весов вершин, близкое к степенному закону. Более того, параметр степенного распределения, наиболее близкого к эмпирическим данным, находится для случая  $U = 1$ ,  $M = 2$ ,  $R = 3$  как правило на отрезке [2; 3]. Таким образом, для построенных так графов выполняется свойство безмасштабных сетей.

У графов, построенных при использовании других рассмотренных весовых функций, указанные свойства выполняются реже, что свидетельствует о целесообразности использования  $(U, M, R)$ -модели с параметрами  $U = 1$ ,  $M = 2$ ,  $R = 3$ .

## Литература

1. Fortunato, S., Newman, M. E. J. 20 years of network community detection // *Nat. Phys.* 2022. № 18. P. 848–850.
2. Чеповский А. А. О неявных сообществах на графе взаимодействующих объектов // *Успехи кибернетики.* 2023. Т.4. № 1. С. 56–64.
3. Blöcker, C., Nieves, J. C. & Rosvall, M. Map equation centrality: community-aware centrality based on the map equation. // *Appl Netw Sci.* 2022. № 7:56. – 24 p. DOI: 10.1007/s41109-022-00477-9
4. Rajeh, S., Savonnet, M., Leclercq, E. et al. Comparative evaluation of community-aware centrality measures // *Qual Quant.* 2023. 57. P. 1273–1302. DOI: 10.1007/s11135-022-01416-7.
5. Попов В. А., Чеповский А. А. Модели импорта данных из мессенджера Telegram // *Вестник Новосибирского государственного университета. Серия: Информационные технологии.* 2022. Т.20. №2. С. 60–71.
6. Попов В. А., Чеповский А. А. Выделение неявных пересекающихся сообществ на графе взаимодействия Telegram-каналов с помощью «метода Галактик» // *Труды ИСА РАН.* 2022. Т.72. №4. С. 39–50.
7. Чеповский А.А. Анализ графов взаимодействующих объектов. – М.: Национальный открытый университет «ИНТУИТ». 2022. – 270 с.
8. Щербаква, Н. Г. Модели сетей с предпочтительным присоединением // *Проблемы информатики.* 2019. № 3(44). С. 46–61.
9. Бадрызов, В. А., Юдина М. Н. Исследование процессов распространения информации в социальной сети методом имитационного моделирования / Десятая всероссийская научно-практическая конференция по имитационному моделированию и его применению в науке и промышленности «Имитационное моделирование // Теория и практика» (ИММОД-2021): Труды конференции, Санкт-Петербург, 20–22 октября 2021 года / Редакторы Плотников А. М., Долматов М. А., Смирнова Е. П. – Санкт-Петербург: АО «Центр технологии судостроения и судоремонта». 2021. С. 89–94.
10. Tereza Nečasová, Ninon Burgos, David Svoboda, Chapter 25 – Validation and evaluation metrics for medical and biomedical image synthesis / *Biomedical Image Synthesis and Simulation* // *Academic Press.* 2022. P. 573–600. DOI: 10.1016/B978-0-12-824349-7.00032-3.

