

МИНОБРНАУКИ РОССИИ

---

Федеральное государственное бюджетное  
образовательное учреждение высшего образования  
«Омский государственный технический университет»

---

Кафедра «Математические методы и информационные технологии в экономике»

# **СИСТЕМЫ УПРАВЛЕНИЯ, ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ**

Материалы  
II Всероссийской научно-практической конференции  
с международным участием  
(Омск, 19–20 мая 2020 года)

В двух томах

Том I

Омск  
Издательство ОмГТУ  
2020

УДК 004+519+65  
ББК 32.97+22.172+65.291.8  
С40

Ответственный редактор

*В. Н. Задорожный*, д-р техн. наук,  
профессор кафедры «ММиИТЭ» ОмГТУ

**Системы управления, информационные технологии и математическое моделирование** : материалы II Всерос. науч.-практ. конф. с международ. участием (Омск, 19–20 мая 2020 г.) : в 2 т. / Минобрнауки России, ОмГТУ, каф. «ММиИТЭ» ; [отв. ред. В. Н. Задорожный]. – Омск : Изд-во ОмГТУ, 2020.

ISBN 978-5-8149-3098-9

Т. I. – 140 с. : ил.

ISBN 978-5-8149-3099-6

Предлагаются новые методы и результаты математического моделирования, направленные на совершенствование современных систем управления и информационных технологий. Наряду с оригинальными научными результатами приводятся эффективные инженерные решения, каждое из которых можно рассматривать как значимое продвижение в области информационных технологий.

Аналитическими и численными методами исследуются модели, разработанные в терминах Науки о сетях (Network Science) и теории массового обслуживания, выводятся и решаются дифференциальные уравнения, описывающие технические системы, разрабатываются элементы технического и программного обеспечения систем автоматизированного управления.

УДК 004+519+65

ББК 32.97+22.172+65.291.8

*Утверждено к печати  
Программным комитетом конференции*

ISBN 978-5-8149-3099-6 (т. I)  
ISBN 978-5-8149-3098-9

© ОмГТУ, 2020

## Дорогие друзья, коллеги!

Мы проводим вторую Всероссийскую с международным участием научно-практическую конференцию «Системы управления, информационные технологии и математическое моделирование», участниками которой являются студенты, аспиранты, инженеры, преподаватели и ученые, представляющие отечественные и зарубежные вузы и организации.

Конференция проводится в непростых условиях борьбы с пандемией, начало которой ВОЗ объявила 11 марта. Несмотря на это, благодаря возможностям «сетевого» общения, состав участников нашей конференции расширяется. По сравнению с прошлым годом число авторов, работы которых представлены на конференции, возросло на 40%. Участниками конференции стали студенты, аспиранты и ученые – кандидаты и доктора наук – из вузов Омска, Челябинска, Волгограда и Калининграда. В конференции приняли участие выдающиеся ученые из Италии и Южной Кореи.

Тематика конференции охватывает актуальные вопросы, касающиеся современного состояния и перспектив развития информационных технологий и математического моделирования. Сюда входят задачи и методы компьютерного моделирования организационно-технических и природных систем, методы оптимизации и прогнозирования, вопросы развития и использования новых образовательных технологий, технологий создания адаптивных и интеллектуальных систем управления, проблемы информационной безопасности, транспортной и складской логистики и многое другое.

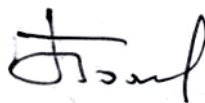
Конференция предоставляет всем участникам возможность поделиться новыми идеями и достижениями в этой широкой области исследований, оценить перспективы практического применения полученных результатов в промышленности, экономике, образовании, в любых сферах жизни.

Надеемся, что новые научные контакты, издание сборника материалов конференции, их размещение на сайте <https://imsait.github.io> и в Elibrary позволят и после конференции продолжить нашу совместную работу, как в непосредственном очном общении, так и в дистанционном режиме.

Пусть это участие в работе конференции будет для вас памятным, приятным и полезным, и станет источником новых идей в вашей научной и образовательной деятельности.

Благодарим всех нашедших возможность участия в конференции за этот праздник общения, за радость обмена пытливой исследовательской мыслью и инженерной смекалкой. Творческих успехов всем участникам!

Председатель  
программного комитета конференции  
профессор, д.э.н., к.ф.-м.н.



Г. Н. Бояркин

Заместитель председателя  
доцент, д.т.н.



В. Н. Задорожный

## ПРОГРАММНЫЙ КОМИТЕТ

### **Председатель:**

**Бояркин Г.Н.**, д.э.н., профессор кафедры ММиИТЭ ОмГТУ

### **Заместитель председателя:**

**Задорожный В.Н.**, д.т.н., профессор кафедры ММиИТЭ ОмГТУ

### **Члены программного комитета:**

**Ко Х. (Ko Hoon)**, профессор IT Research Institute Chosun University, South Korea

**Пагано М. (Pagano Michele)**, профессор University of Pisa, Italy

**Альтман Е.А.**, к.т.н., доцент каф. АиСУ ОмГУПС

**Батенькина О.В.**, к.т.н., доцент каф. ДиТМ ОмГУПС

**Бахмутский Ю.А.**, заведующий кафедрой ММиИТЭ ОмГТУ

**Гегечкори Е.Т.**, к.т.н., доцент каф. ММиИТЭ ОмГТУ

**Калиберда Е.А.**, к.т.н., доцент каф. ММиИТЭ ОмГТУ

**Малютин А.Г.**, к.т.н., доцент, зав. кафедрой АиСУ ОмГУПС

**Филимонов В.А.**, д.т.н., профессор, с.н.с. ОФ ИМ СО РАН

**Шевелева О.Г.**, ст. преподаватель каф. ММиИТЭ ОмГТУ

**Юдин Е.Б.**, к.т.н., доцент каф. ММиИТЭ ОмГТУ

### **Оргкомитет и секретариат:**

**Юдин Е.Б.**, к.т.н., доцент каф. ММиИТЭ ОмГТУ – председатель

**Задорожный В.Н.**, д.т.н., профессор каф. ММиИТЭ – зам. председателя

**Батенькина О.В.**, к.т.н., доцент каф. ММиИТЭ ОмГТУ

**Калиберда Е.А.**, к.т.н., доцент каф. ММиИТЭ ОмГТУ

**Кравченко К.В.**, ст. преподаватель каф. ММиИТЭ ОмГТУ

**Стариков В.И.**, к.т.н., доцент каф. ММиИТЭ ОмГТУ

**Шевелева О.Г.**, ст. преподаватель каф. ММиИТЭ ОмГТУ

*Материалы публикуются в авторской редакции*

# СОДЕРЖАНИЕ

## I. ПЛЕНАРНЫЕ ДОКЛАДЫ

**Задорожный В.Н., Пагано М., Захаренкова Т.Р.**

О реализации метода абсолютных приоритетов с бесконечными оптимальными разметками в сетях с коммутацией пакетов.....7

**Задорожный В.Н., Юдин Е.Б., Юдина М.Н.**

Методы комплексной калибровки графовых моделей растущих сетей.....26

**Наумов В.А., Землянов А.А.**

Гидравлическая модель нестационарного ламинарного течения по трубопроводу .....49

## II. СЕКЦИЯ 1

**Абдрахманов А.Н.**

Создание программы для построения гистограммы на основе графематического анализа данных из открытого интернет-источника .....55

**Бадрызлов В.А.**

Исследование случайных графов методом имитационного моделирования .....60

**Баркова И.М., Галдина Ю.П., Кравченко К.В., Бахмутский Ю.А.**

Оптимизация и автоматизация бизнес-процесса назначения повышенной стипендии в ОмГТУ .....70

**Берестинов А.А., Омельченко С.Д., Кулганатов А.З.**

Разработка беспроводного датчика температуры с совместным использованием микрокомпьютера ESP32 .....75

**Будзинская Е.О., Калиберда А.А., Новиков А.А.**

Анализ систем электронного документооборота для автоматизации процесса «внутренний аудит» .....79

**Галдина Ю.П., Кравченко К.В., Баркова И.М.**

Выбор программного продукта для автоматизации процесса управления вузом ..83

**Доценко Е.В., Гегечкори Е.Т.**

Моделирование бизнес-процессов для стандартизации деятельности компании .....87

**Задорожный В.Н., Галич Ю.Г.**

Оценка точности приближенных формул для средней длины очереди.....93

***Задорожный В.Н., Захаренкова Т.Р., Маркова М.П.***

Расчет и моделирование одноканальных систем с абсолютными приоритетами ..... **106**

***Задорожный В.Н., Юдина М.Н.***

Обобщение метода равновероятного выбора остовных деревьев ..... **111**

***Ольгина И.Г., Пронин И.В., Абдрахманов А. Н.***

Построение графовых моделей сети цитирования научных публикаций ..... **118**

***Сай Ван Квонг, Щербаков М.В.***

Система предсказательного технического обслуживания на основе глубоких нейронных сетей ..... **126**

***Юдина М.Н.***

Показатель статистической значимости подграфов  $R'$  в задаче выявления сетевых мотивов..... **130**

**Авторский указатель ..... 138**

УДК 519.245:004.421.5

## О РЕАЛИЗАЦИИ МЕТОДА АБСОЛЮТНЫХ ПРИОРИТЕТОВ С БЕСКОНЕЧНЫМИ ОПТИМАЛЬНЫМИ РАЗМЕТКАМИ В СЕТЯХ С КОММУТАЦИЕЙ ПАКЕТОВ

В. Н. Задорожный<sup>1</sup>, Микеле Пагано<sup>2</sup>, Т. Р. Захаренкова<sup>1</sup>

<sup>1</sup>Омский государственный технический университет, Омск, Россия

<sup>2</sup>Пизанский университет, Пиза, Италия

**Аннотация.** *Выполняется исследование возможности практического применения метода, предложенного авторами для кардинального снижения средних длин очередей и вероятностей потерь в системах с бесконечной дисперсией времени обслуживания заявок. Метод основан на введении дисциплины абсолютных приоритетов с дообслуживанием и назначении входящим заявкам приоритетов, оптимальным образом зависящих от времени их обслуживания. Формально число приоритетных уровней при этом получается бесконечным. При бесконечной дисперсии времени обслуживания (имеющей место, например, в современных сетях передачи данных с фрактальным трафиком), среднее время ожидания заявок благодаря использованию рассматриваемого метода становится конечным, а в случае очень больших дисперсий времени обслуживания среднее время ожидания резко сокращается. Однако возможность непосредственного применения данного метода в сетях с коммутацией пакетов наталкивается на препятствия: передача заявок – пакетов, поступающих в узлы сети передачи данных, не прерывается, поэтому применять к ним дисциплину абсолютных приоритетов не представляется возможным. Кроме того, почти все пакеты, на которые разбиваются передаваемые файлы, имеют одинаковые относительно небольшие размеры, и поэтому вводить для них приоритеты с бесконечным числом приоритетных уровней нецелесообразно. Тем не менее, рассматриваемый метод абсолютных приоритетов можно применять, если в качестве заявок рассматриваются не пакеты сообщений, а файлы, разбиваемые на пакеты. Эта возможность обосновывается соответствующей математической моделью и исследуется посредством имитационного и аналитического моделирования.*

**Ключевые слова:** *системы с очередями, фрактальный трафик, приоритеты, среднее время ожидания, бесконечные разметки диапазона значений времени обслуживания, оптимальное назначение приоритетов.*

## ON THE IMPLEMENTATION OF THE PREEMPTIVE PRIORITIES METHOD WITH INFINITE OPTIMAL MARKINGS FOR PACKET SWITCHED NETWORK

V. N. Zadorozhnyi<sup>1</sup>, M. Pagano<sup>2</sup>, T. R. Zakharenkova<sup>1</sup>

<sup>1</sup>Omsk State Technical University, Omsk, Russia,

<sup>2</sup>University of Pisa, Pisa, Italy

**Abstract.** *Analysis of practical implementation of the method proposed by the authors is made. This method is for drastic reduction of an average queue length and loss probabilities in queueing systems with an infinite variance of requests service time. The method is based on the introduction of discipline of preemptive priority with afterservice and on the priority assignment for incoming requests which depends on requests service time in an optimal way. Formally, the number of priority levels in such case is infinite. When variance of requests service time is infinite (this takes place, for*

*example, in modern networks with fractal traffic) average waiting time becomes finite when using the proposed method. In the case of very large variances of service time the average waiting time decrease drastically. However, the possibility of direct applicability the method in packet-switched networks faces obstacles: transmission of requests – packets, which are arriving in nodes of networks, is incessant, therefore, the applying of preemptive priority discipline does not seem possible. In addition, almost all packages, which make up transmitted files, have the same relatively small sizes, therefore, introduction of preemptive priorities with infinite number of priority levels is impractical. Nevertheless, the method of preemptive priorities could be applied if the divided into packets files are considered as requests, but not the packages. This possibility is justified by the corresponding mathematical model and investigated by means of simulation and analytical modeling.*

**Keywords:** *queueing systems, fractal traffic, priorities, average waiting time, infinite markup of service time range, optimal priorities assignment.*

## 1. Введение

Теория массового обслуживания (ТМО) – *Queueing Theory* – является одним из наиболее актуальных математических инструментов исследования сетей передачи данных (СПД) на системном уровне [1–5]. В конце XX века путем обработки большого объема статистических данных установлено [6] и в последующем многократно подтверждено независимыми исследователями, что современный сетевой трафик приобрел свойства статистического самоподобия, т.е. стал фрактальным. Такому трафику присущи посылки, распределения объемов которых имеют тяжелые хвосты, как правило – степенные; причем такие распределения обуславливают появление долговременных зависимостей (ДВЗ) между объемами данных, передаваемых в следующих друг за другом непересекающихся интервалах времени [7]. Как правило, степенные распределения объемов посылок имеют конечное математическое ожидание (м.о.) и бесконечную дисперсию, вследствие каковой среднее время ожидания в очередях и средние длины очередей становятся бесконечными [8–10].

Указанные особенности фрактального трафика приводят не только к проблемам передачи данных, структурирующимся вокруг больших очередей и соответствующих избыточных аппаратных затрат, но и к значительным затруднениям в исследовании методов и алгоритмов обслуживания трафика. Традиционные аналитические методы ТМО, используемые в классической теории телетрафика, как правило, не позволяют рассчитывать и оптимизировать системы с бесконечной дисперсией времени обслуживания. Непосредственное имитационное моделирование (ИМ) таких систем также сталкивается со специфическими труднопреодолимыми проблемами. К их числу относятся [11–15]:

- систематические смещения моментов случайных величин (с.в.), реализуемых с помощью стандартных генераторов случайных чисел (ГСЧ), если эти с.в. имеют распределения с тяжелыми хвостами (РТХ),
- чрезвычайно медленная сходимость имитационных оценок, обусловленная ДВЗ,
- необходимость применения ГСЧ с практически неограниченной длиной периода для выполнения длинных прогонов, обеспечивающих получение в ИМ оценок с приемлемой точностью.

Таким образом, серьезные проблемы при исследовании рассматриваемого



класса систем массового обслуживания (СМО) и, в особенности, при решении задач оптимизации таких СМО, возникают как при использовании аналитического подхода, так и в непосредственном ИМ.

Эти проблемы могут преодолеваются разработкой таких *аналитико-имитационных* подходов, в которых удастся органически синтезировать преимущества аналитического и имитационного подходов. Для этого решаемую задачу следует представить в виде совокупности взаимосвязанных подзадач, одна часть которых эффективно решается аналитическими методами, другая – имитационными. Как правило, реализация этого пути требует выполнения обширных предварительных исследований, не гарантирующих успеха, но в случае успеха приводящих к нетривиальным и действительно эффективным методам и алгоритмам решения сложных задач.

Подобный аналитико-имитационный подход позволил в работах [8–10] создать и развить метод, позволяющий кардинально снижать среднюю длину очереди (и вероятности потерь в случае конечных буферов) в системах с бесконечной дисперсией времени обслуживания заявок. Метод основан на введении дисциплины абсолютных приоритетов с дообслуживанием и назначении входящим заявкам приоритетов, определяемых бесконечными оптимальными разметками полубесконечного диапазона возможных значений времени обслуживания. При бесконечной дисперсии времени обслуживания среднее время ожидания заявок благодаря использованию этого метода становится конечным, а в случае очень больших дисперсий времени обслуживания среднее время ожидания (и, соответственно, средняя длина очереди) резко сокращаются. Далее этот метод будем называть методом абсолютных приоритетов с бесконечными оптимальными разметками (методом АП БОР).

Однако возможность непосредственного применения данного метода в сетях с коммутацией пакетов наталкивается на два препятствия. Первое препятствие заключается в том, что передача заявок – пакетов, поступающих в узлы СПД, не прерывается, поэтому применять к ним дисциплину абсолютных приоритетов с дообслуживанием в рамках концепции коммутации пакетов не представляется возможным. Второе препятствие состоит в том, что почти все пакеты, на которые разбиваются передаваемые файлы, имеют одинаковые относительно небольшие размеры (одинаковое время передачи по каналу), и поэтому вводить для них приоритеты с бесконечным числом приоритетных уровней, определяемых временем обслуживания, нецелесообразно. Тем не менее, бесконечные оптимальные разметки можно применять без разрушения концепции коммутации пакетов и при использовании относительных приоритетов, если в качестве заявок рассматриваются не пакеты сообщений, а файлы, разбиваемые на пакеты. Основная гипотеза, приводящая к выявлению такой возможности, состоит в том, что относительные приоритеты передаваемых файлов в условиях фрактального трафика и разбиения файлов на пакеты, должны функционировать практически так же, как абсолютные приоритеты файлов. Действительно, технико-экономические показатели СПД определяются в основном передачей файлов больших

размеров, встречающихся при степенном распределении размеров настолько часто, что при конечном среднем размере дисперсия размеров может быть бесконечной. На Рис. 1 показано, как происходит временное прекращение передачи пакетов некоторого файла (изображенных не закрашенными прямоугольниками, расположенными в очереди к каналу) при поступлении пакетов более приоритетного файла (черные прямоугольники).

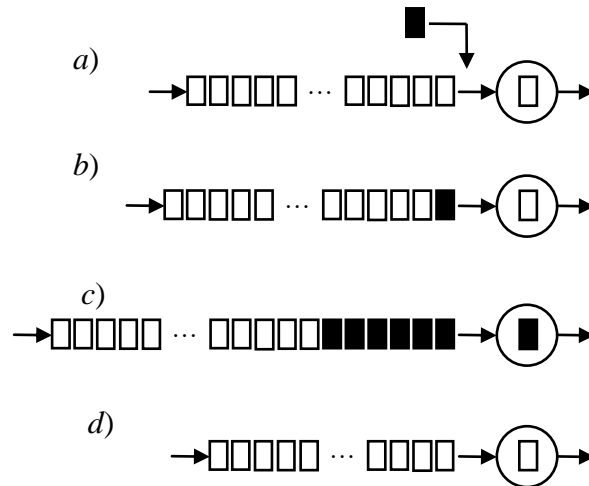


Рис. 1. Передача приоритетного файла, разбитого на пакеты (черные прямоугольники), при дисциплине относительных приоритетов

Рис. 1 *a)*, *b)* иллюстрирует поступление приоритетного «черного» пакета и ожидание им окончания передачи «белого» пакета, уже занимающего канал. Сразу после окончания передачи «белого» пакета начинает передаваться «черный», и в это время за ним, в начало очереди, поступают остальные «черные» пакеты приоритетного файла. Когда передача приоритетного файла заканчивается, возобновляется передача «белых» пакетов менее приоритетного файла. И, поскольку передаваемые файлы состоят из нескольких пакетов, то задержка начала обслуживания приоритетного файла, составляющая долю времени передачи одного пакета, относительно невелика, что приводит к небольшому отличию процесса передачи файлов с относительными приоритетами от процесса передачи файлов с абсолютными приоритетами.

Возможность применения метода бесконечных оптимальных разметок к передаче файлов с относительными приоритетами в условиях фрактального трафика далее обосновывается соответствующей принципиальной математической моделью, исследуемой посредством имитационного моделирования.

## 2. Постановка задачи

Решаемая далее задача заключается в проверке гипотезы об эффективности метода АП БОР в условиях передачи файлов, разбиваемых на пакеты, и применения в очередях пакетов дисциплины относительных приоритетов.

Рассматриваемая версия метода АП БОР основана на оптимальных экспо-

ненциальных разметках [8, 9] диапазона возможных значений времени  $x$  обслуживания, распределенного по закону Парето (Pa), в СМО M/Pa/1/ $\infty$  (Рис. 2). Математическое ожидание (м.о.)  $\bar{\tau}$  интервала поступления заявок входящего пуассоновского потока и среднее время  $\bar{x}$  обслуживания заявки удовлетворяют условию существования стационарного режима  $\rho = \bar{x} / \bar{\tau} < 1$ , где  $\rho$  – коэффициент загрузки системы.

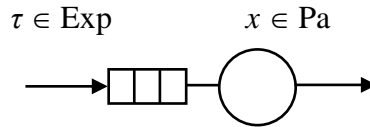


Рис. 2. Символическое изображение исследуемой СМО

В отличие от Рис.1 на Рис. 2 заявки-прямоугольники изображают передаваемые файлы, а не пакеты, на которые файлы разбиваются. Каждый прямоугольник на Рис. 2 – это файл, состоящий из группы пакетов, на которые этот файл разбит, или же его фрагмент, образованный из-за перерывов процесса передачи файла вследствие поступления в систему пакетов более приоритетного файла (см. Рис. 1). Если на Рис. 2 изображать каждый прямоугольник, соответствующий определенному файлу, своим цветом, то очередь будет состоять из произвольного числа разноцветных прямоугольников. Несколько прямоугольников одного цвета будут тогда соответствовать фрагментам одного и того же файла. При этом время обслуживания каждого прямоугольника будет определяться числом и размером входящих в него пакетов.

Таким образом, Рис. 2 представляет исследуемый процесс на макроуровне – уровне файлов, а Рис. 1 – на мезоуровне, уровне пакетов. Решаемая задача состоит в проверке эффективности метода оптимальных разметок, рассчитываемых на макроуровне. Требуется оценить влияние этого метода на среднее время передачи файлов. Этот показатель эффективности связи более других интересует конечных пользователей СПД. Естественно, следует оценить и изменения, происходящие на мезоуровне – на уровне характеристик очереди пакетов, чтобы можно было сформировать комплексное представление о влиянии исследуемого метода на характеристики СПД.

### 3. Теория

При анализе и моделировании фрактального трафика наиболее часто используется распределение Парето. Функция распределения вероятностей (ф.р.) Парето  $F(t)$  задается формулой

$$F(t) = 1 - \left( \frac{K}{t} \right)^\alpha, \quad \alpha > 0, \quad t \geq K, \quad (1)$$

где  $\alpha$  – параметр формы, безразмерный;  $K > 0$  – минимальное значение с.в., в

нашем случае – времени  $x$  обслуживания,  $t$  – вещественная переменная (возможные значения  $x$ ). С.в.  $x$  и переменная  $t$  выражаются в одних тех же единицах времени, поэтому параметр  $K$ , как видно из (1), является масштабным параметром распределения Парето.

Из (1) нетрудно определить м.о.  $M(x)$  и дисперсию  $D(x)$  времени  $x$ :

$$M(x) = \bar{x} = \int_{-\infty}^{\infty} t dF(t) dt = \begin{cases} \frac{\alpha K}{\alpha - 1} & \text{при } \alpha > 1, \\ \infty, & \text{если } \alpha \leq 1, \end{cases} \quad (2)$$

$$D(x) = \int_{-\infty}^{\infty} t^2 dF(t) dt - \bar{x}^2 = \begin{cases} \frac{\alpha K^2}{(\alpha - 1)^2 (\alpha - 2)} & \text{при } \alpha > 2, \\ \infty, & \text{если } \alpha \leq 2. \end{cases} \quad (3)$$

В соответствии с методом бесконечных оптимальных разметок [8–10] разобьем диапазон  $K \leq t < \infty$  возможных значений трудоемкости  $x$ , используя разметку (последовательность точек)

$$\{t_k\} = t_0, t_1, \dots, t_k, \dots \quad (4)$$

на промежутки

$$[t_0, t_1), [t_1, t_2), \dots, [t_{k-1}, t_k), \dots$$

где  $t_0 = K$ . В системе, изображенной на Рис. 2, если поступившая заявка-файл имеет время  $x$  передачи по каналу, принадлежащее  $k$ -му промежутку разметки (4), т.е. если  $x \in [t_{k-1}, t_k)$ , то отнесем эту заявку к  $k$ -му приоритетному классу. При этом будем считать приоритет заявки тем меньшим, чем выше номер  $k$  ее приоритетного класса. Заявке приоритетного класса 1 можно присвоить произвольный приоритет, заданный целым числом, заявке класса 2 – приоритет, на единицу меньший, и т.д. Тем самым мы допускаем отрицательные приоритеты, поскольку число приоритетных классов в общем случае бесконечно. Если число промежутков  $[t_{k-1}, t_k)$  конечно, то в последнем из них  $[t_{N-1}, t_N)$  имеем  $t_N = \infty$ . Метод бесконечных оптимальных разметок разработан для дисциплины абсолютных приоритетов с дообслуживанием. В используемой принципиальной модели системы передачи файлов (см. рис. 2) применяются относительные приоритеты (которые, согласно проверяемой гипотезе, учитывающей разбиение файлов на пакеты, могут действовать почти так же, как абсолютные).

Для рассматриваемой системы с абсолютными приоритетами и пуассоновскими входящими потоками известно следующее выражение среднего времени  $U_k$  пребывания в системе заявок  $k$ -го приоритетного класса [1, 8]:

$$U_k = \frac{b_k}{1 - \sigma_{k-1}} + \frac{\sum_{i=1}^k \lambda_i b_i^{(2)}}{2(1 - \sigma_k)(1 - \sigma_{k-1})}, \quad k = 1, 2, \dots, \quad (5)$$

используя которое, можно определить среднее время  $U$  пребывания заявки

$$U = \sum_k p_k U_k \quad (6)$$

и среднее время  $W$  ожидания заявки в очереди

$$W = U - b, \quad (7)$$

где  $b_k$  – среднее время обслуживания заявок  $k$ -го приоритетного класса;

$\lambda_i$  – интенсивность поступления заявок  $i$ -го приоритетного класса:  $\lambda_i = \lambda p_i$ ;

$\lambda = 1/\bar{\tau}$ ;  $p_i$  – то же, что и  $p_k$  (см. ниже), изменено лишь обозначение индекса;

$b_i^{(2)}$  – второй момент времени обслуживания заявок  $i$ -го приоритетного класса;

$\sigma_k = \sum_{i=1}^k \rho_i$  – сумма коэффициентов загрузки системы заявками приоритетных классов с 1-го по  $k$ -й;

$\rho_i = \lambda_i b_i$ ;  $\sigma_{k-1} = \sigma_k - \rho_k$ ;  $p_k = P(t_{k-1} \leq x < t_k)$ ;

$b$  – среднее время обслуживания заявки: обозначение  $b$  для ТМО традиционное, здесь  $b = M(x) = \bar{x}$ .

Таким образом, в (7) среднее время ожидания  $W$  зависит от разметки (4). В соответствии с рекомендациями [8] для снижения времени  $W$  будем использовать бесконечную экспоненциальную разметку  $\{t_k\} = t_0, t_1, \dots, t_k, \dots$ , в которой

$$t_0 = K, \quad t_k = K + ce^{ak}, \quad (k = 1, 2, \dots), \quad (8)$$

$c, a$  – коэффициенты, которые можно оптимизировать с целью минимизации среднего времени ожидания  $W$  при заданных параметрах  $\lambda, K, \alpha$  рассматриваемой СМО М/Па/1/∞. Следующие формулы, выведенные в [8], выражают используемые в (5–7) параметры  $b_k, p_k, b_k^{(2)}$  через координаты  $\{t_k\}$  разметки (8):

$$b_k = \frac{\alpha}{\alpha - 1} \cdot \frac{t_{k-1}^{1-\alpha} - t_k^{1-\alpha}}{t_{k-1}^{-\alpha} - t_k^{-\alpha}}, \quad (9)$$

$$p_k = \left( \frac{K}{t_{k-1}} \right)^\alpha - \left( \frac{K}{t_k} \right)^\alpha, \quad (10)$$

$$b_k^{(2)} = \begin{cases} \frac{\alpha}{\alpha - 2} \cdot \frac{t_{k-1}^{2-\alpha} - t_k^{2-\alpha}}{t_{k-1}^{-\alpha} - t_k^{-\alpha}}, & \alpha \neq 2, \\ \frac{2}{t_{k-1}^{-2} - t_k^{-2}} \ln \left( \frac{t_k}{t_{k-1}} \right), & \alpha = 2, \end{cases} \quad (11)$$

(неопределенность типа 0/0, возникающая в (11) в верхнем выражении при  $\alpha = 2$ , в нижнем выражении раскрыта по правилу Лопиталя). Тем самым формулами (9)–(11) параметры приоритетных классов выражаются через два коэффициента ( $a, c$ ) разметки (8), определяющих значения  $\{t_k\}$ .

Фиксируя параметры  $\lambda = 1/6, K = 1, \alpha = 1,5$  системы М/Па/1/∞ (при которых

$\bar{x} = 3$ ,  $D(x) = \infty$  и  $b_k^{(2)}$  конечны), реализуем в Excel по формулам (7–5) и (8–11) вычисление  $W$  через параметры  $(a, c)$ , заданные произвольно. Затем находим с помощью «Поиска решения» значения  $(a, c)$ , оптимальные по критерию  $W \rightarrow \min$ :

$$c \approx 0,5054, \quad a \approx 1,0550. \quad (12)$$

Для такой оптимизации бесконечной разметки, как это становится очевидным в ходе вычислений, достаточно учесть первые 68 (или меньше) приоритетных классов. При  $k = 68$  имеем  $t_k = K + ce^{ak} = 1 + 0,5054e^{1,055 \cdot 68} = 1,207 \cdot 10^{30}$ . Вероятность того, что время  $x$  обслуживания поступающей заявки превысит величину  $t_k$ , согласно (1), составляет  $(K/t_k)^\alpha = (1/(1,207 \cdot 10^{30}))^{1,5} \approx 5,1 \cdot 10^{-47}$ . Поэтому ограничение числа классов первыми 68-ю классами на практике не изменит процесса обслуживания заявок по сравнению с бесконечным числом классов. Кроме того, столь малые вероятности классов с номерами  $k > 68$  приводят к тому, что оптимизация параметров  $(a, c)$  разметки по первым 68 классам доставляет тем же оптимальные значения этих параметров (с точностью до многих значащих десятичных цифр), что и при учете любого большего 68 числа классов. Этот пример расчета числа учитываемых промежутков экспоненциальной разметки показывает, что на практике при ее использовании нет необходимости реализовать бесконечное (или очень большое) число приоритетных уровней. Можно обойтись использованием нескольких десятков приоритетов, и это практически не снизит эффективности метода бесконечных оптимальных разметок. При этом оптимизация параметров  $(a, c)$  выполняется за 1–5 секунд компьютерного времени.

Применяя экспоненциальную разметку с параметрами (12), оптимальными для абсолютных приоритетов файлов, проверим с помощью ИМ гипотезу о том, что в условиях разбиения файлов на пакеты действие относительных приоритетов файлов, назначаемых по такой разметке, близко к действию абсолютных приоритетов. Тогда относительные приоритеты файлов при экспоненциальной разметке с параметрами (12) обеспечат существенное снижение среднего времени  $W$  ожидания файлов (с бесконечной до небольшой конечной величины).

#### 4. Имитационная модель системы

Текст имитационной модели рассматриваемой системы представлен на Рис. 3 и Рис. 4. Модель написана на языке GPSS World (студенческая версия).

Включенные в модель три процедуры (см. Рис. 4) библиотеки [16] написаны на языке PLUS, входящем в состав системы имитационного моделирования GPSS World, и представляют собой высококачественные генераторы псевдослучайных величин. Библиотека этих процедур и примеры их применения описаны в [15]. Процедуры основаны на методе «Вихрь Мерсенна» (Mersenne Twister, MT), реализованном авторами в виде программы mtrand.dll, и на методе ARAND [11, 13] генерации несмещенных распределений с.в. с РТХ, в том числе – распределения Парето. Поэтому для выполнения программы, представленной на Рис. 3 и Рис. 4, следует в папку GPSS World с исполняемым файлом GPSS World Student.exe поместить файл mtrand.dll (его могут выслать авторы статьи).

A_	EQU	1.054995	; коэффициент а показателя экспоненциальной разметки
C_	EQU	0.505462	; масштабный множитель с экспоненциальной разметки
K_	EQU	1	; параметр K распределения Парето
ALPHA	EQU	1.5	; параметр "альфа" распределения Парето
RO_	EQU	(1/6#K_#ALPHA/(ALPHA-1))	; коэффициент загрузки при "лямбда" = 1/6
* Формирование потока заявок (передаваемых файлов)			
	GENERATE	(Mteponential(0,6))	; Поток файлов
	ASSIGN	TOBSL, (MteArandPareto(K_,ALPHA))	; Время передачи файла с учетом разбивки его на пакеты - вместе со служебной информацией и межадр. интервалами
* Определение числа пакетов, на которые разбивается файл			
	ASSIGN	NX, (Int(P\$TOBSL))	; Число пакетов единичной длит-сти. Ед. времени = K
	ASSIGN	Class, (INT(Log((P\$TOBSL-K_)/C_)/A_))	; Определение номера k приоритетного класса файла
	TEST L	P\$Class,1,NEXT	; Если получился номер k < 1, заменяем его единицей
	ASSIGN	Class,1	; Задаем приоритет заявки
NEXT	PRIORITY	(1000-P\$Class)	
	QUEUE	FILE_TIME	
	SPLIT	1,OSTATOK	; Копия транзакта для формирования последнего, короткого пакета
	TRANSFER	,PACKET	; Переход оригинала транзакта к формированию стандартных пакетов
OSTATOK	ASSIGN	TOBSL, (P\$TOBSL-INT(P\$TOBSL))	; Формирование короткого пакета
	TRANSFER	,OBSL	; Переход короткого пакета на обслуживание
PACKET	SPLIT	(P\$NX -1)	
	ASSIGN	TOBSL,1	; Файл разбивается на P\$NX_ пакетов стандартной длительности
* Обслуживание заявки (передача пакета)			
OBSL	QUEUE	BUFF	; Поступление пакета в буферную память узла
	SEIZE	1	; Начало передачи пакета по каналу связи
	DEPART	BUFF	; Выход пакета из буферной памяти
	ADVANCE	P\$TOBSL	; Передача пакета по каналу
	RELEASE	1	; Освобождение канала
	ASSEMBLE	(P\$NX +1)	; Сборка всех пакетов файла в один файл
	DEPART	FILE_TIME	; Завершение передачи файла
	TERMINATE		
*** TIMER			
	GENERATE	60000000	; Моделируемое время = времени передачи в среднем 10 млн. файлов
	TERMINATE	1	

Рис. 3. Начало модели исследуемой системы M/Pa/1/∞

```

*Процедура MtExponential

PROCEDURE MtExponential(Arg1,Arg2) BEGIN
TEMPORARY Z, XI;
STEP1: Z = Call("mtrand.dll", "?nextNum@@YAHXZ");
IF (Z = 0) THEN GOTO STEP1;
Z = Z/2147483647;
XI = -Arg2 # log(Z) + Arg1;
RETURN (XI);
END;

*Процедура MtArandPareto

PROCEDURE MtArandPareto(Arg1,Arg2) BEGIN
TEMPORARY Z, XI;
Z = MtArand();
XI = Arg1#Z^(-1/Arg2);
RETURN (XI);
END;

*Процедура MtArand

PROCEDURE MtArand() BEGIN
TEMPORARY CCC,zzz;
CCC = 1;
STEP1: zzz = Call("mtrand.dll", "?nextNum@@YAHXZ");
zzz = zzz/2147483647;
STEP2: IF (zzz < 0.1) THEN BEGIN
CCC = 10#CCC;
GOTO STEP1;
END;
STEP3: zzz = zzz/CCC;
RETURN (zzz);
END;

```

Рис. 4. Завершение модели исследуемой системы М/Па/1/∞

Реализованная в mtrand.dll версия «Вихря Мерсенна» генерирует последовательность целых псевдослучайных чисел в промежутке от нуля до  $2^{31} - 1$  и имеет практически неисчерпаемую, равную  $2^{19937} - 1 \approx 10^{6000}$  длину периода. (Периоды ГСЧ, встроенных в GPSS, представляют собой последовательности чисел, которые начинаются в разных местах одного и того же кольца чисел длиной  $2\,047\,999\,999 = 8\,863 \times 743 \times 311$  чисел [17]. Это при использовании в модели нескольких ГСЧ и при большой длине прогона приводит к значительной неконтролируемой корреляции псевдослучайных чисел и, следовательно, к некорректности результатов ИМ).

Модель запускается на выполнение командой START 1. Чтобы повысить точность сравнения прогонов, выполняемых при разных значениях параметров модели, рекомендуется использовать метод общих случайных чисел. Для этого достаточно непосредственно перед каждым стартом модели ретранслировать программу MTINIT+.gps с фиксированным начальным значением последовательности псевдослучайных чисел (Рис. 5). Программа MTINIT+.gps должна быть открыта в том же окне GPSS World, в котором находится модель исследуемой системы. Приводимые далее результаты экспериментов получены с помощью модели Рис. 3, Рис. 4 при начальном значении последовательности псевдослучайных чисел, задаваемом программой MTINIT+.gps, равном единице (как на Рис. 5).



```

GENERATE      1
ASSIGN        1, (MtInit(1))
TERMINATE     1
START         1, NP

PROCEDURE MtInit(Arg) BEGIN
Call_Integer("mtrand.dll", "?init@@YAHH@Z", Arg);
END;
```

Рис. 5. Программа MTINIT+.gps инициализации начального значения процедуры mtrand.dll

Если не требуется стартовать имитацию с одного и того же начального значения псевдослучайной последовательности, то программу инициализации использовать не нужно.

Другой вариант выполнения модели состоит в том, чтобы не использовать процедуры, приведенные на Рис. 4. Тогда их можно удалить из модели, а обращения к ним (во фрагменте модели, представленном на Рис. 3) заменить обращениями к соответствующим генераторам, встроенным в GPSS World. При этом следует учитывать, что результаты ИМ станут менее точными из-за появления систематических погрешностей, обусловливаемых смещением моментов распределения Парето в GPSS [11, 13] и из-за ограниченной длины периода ГСЧ, встроенных в GPSS [14, 17]. Кроме того, при любом варианте – с предложенными на Рис. 4 процедурами или без них – необходимо учитывать обнаруженные в GPSS World дефекты механизма модельного времени [14], которые, как правило, не позволяют корректно имитировать СМО при использовании длин прогонов, определяемых прохождением через модель СМО более 100 млн. транзактов.

## 5. Эксперименты

При выполнении модели, представленной на Рис. 3 и Рис. 4, получены результаты, показанные на Рис. 6 (ненужные в решаемой задаче данные удалены).

FACILITY	ENTRIES	UTIL.	AVE. TIME				
1	36128750	0.499996	0.830357				
QUEUE	MAX	CONT.	ENTRY	ENTRY (0)	AVE.CONT.	AVE.TIME	AVE. (-0)
FILE_TIME	11	1	10007819	0	0.749137	4.491308	4.491308
BUFF	199663	2	36128752	5002053	777.397333	1291.044871	1498.515470

Рис. 6. Результаты выполнения модели, представленной на Рис. 3, Рис. 4

Таблица FACILITY на Рис. 6 содержит данные об устройстве 1 (т.е. канале), полученные в результате ИМ. Эти данные, как видно из текста модели на Рис. 3, получены на мезоуровне, т.е. это данные о передаче пакетов через канал. Число пакетов (ENTRIES), вошедших в канал за время моделирования (т.е. за 60 млн. единиц времени), составило 36 128 750. Статистическая оценка  $\rho_{\text{им}}$  коэффициента загрузки (0.499996) получилась близкой к точному значению  $\rho = \lambda \bar{x} = (1/6) \times 3 = 0,5$ . Оценка  $b_{\text{п им}}$  среднего времени обслуживания пакета, равная 0,830357, показывает, что наряду со стандартными пакетами, обслуживаемыми за одну единицу времени, пришло достаточно много укороченных пакетов

(которые появляются по одному в каждом передаваемом файле).

Таблица QUEUE содержит данные о двух очередях: об очереди FILE\_TIME файлов, определенной на макроуровне, и об очереди BUFF пакетов, определенной на мезоуровне рассмотрения системы. Под заголовками MAX, CONT., ENTRY, ENTRY(0), AVE.CONT., AVE.TIME и AVE.(-0) приводятся, соответственно, максимальные длины очередей на моделируемом интервале времени, длины очередей в момент завершения прогона, число входов в каждую очередь, число входов без учета «нулевых» (с мгновенным занятием свободного канала) входов, средняя длина каждой очереди, среднее время ожидания в каждой очереди и среднее время ожидания без учета «нулевых» входов.

Обращают на себя внимание «неожиданно» хорошие показатели очереди файлов FILE\_TIME, согласующиеся с проверяемой гипотезой. Действительно, несмотря на то, что в среднем один файл состоит лишь из трех-четырех пакетов, средняя длина очереди файлов меньше средней длины очереди пакетов не в три-четыре раза, а приблизительно в  $777/0,749 \approx 1037,38 \approx 1000$  раз (см. Рис. 6). При этом среднее время передачи файла не в три-четыре раза больше среднего времени передачи пакета, а приблизительно в  $1291/4,49 \approx 287$  раз меньше. Применяя к этому результату теорию, разрабатываемую в [8–10], его следует объяснить тем, что характеристики очереди файлов, передаваемых в режиме «почти» абсолютных приоритетов, стационарны, несмотря на бесконечную дисперсию размеров файлов. А средняя длина очереди пакетов, передаваемых в режиме строго относительных приоритетов (приоритеты пакетов равны приоритетам соответствующих файлов, разбиваемых на эти пакеты) не сходятся к конечному стационарному значению. Стационарная средняя длина очереди пакетов существует (так как коэффициент загрузки СМО меньше единицы), но равна бесконечности.

Для верификации этого объяснения выполнены три прогона модели системы: наряду с рассмотренным прогоном, определяемым временем прохождения в среднем  $N = 10$  млн. заявок, выполнены еще два прогона, при  $N = 1$  млн. и  $N = 100$  тыс. Результаты всех трех прогонов сведены в Табл. 1.

**Таблица 1.** Сходимость имитационных оценок с ростом длины прогона

$N$	$\rho_{\text{им}}$	$b_{\text{п им}}$	$Q_{\text{п макс}}$	$Q_{\text{ф макс}}$	$L_{\text{п им}}$	$L_{\text{ф им}}$	$W_{\text{п им}}$	$W_{\text{ф им}}$
$10^5$	0,4795	0,8253	1486	8	18,24	0,700	31,4	4,22
$10^6$	0,5052	0,8319	43485	10	387,6	0,761	638,3	4,56
$10^7$	0,499996	0,8304	199663	11	777,4	0,749	1291,0	4,49

Результаты экспериментов, приведенные в Табл. 1, подтверждают предложенное выше теоретическое объяснение различий между характеристиками очереди файлов (макроуровень) и характеристиками очереди пакетов (мезоуровень). Действительно, имитационные оценки  $L_{\text{ф им}}$  средней длины  $L_{\text{ф}}$  очереди файлов и  $W_{\text{ф им}}$  среднего времени  $W_{\text{ф}}$  ожидания файлов с ростом  $N$  сходятся, судя по всему, к некоторым близким к 0,75 и, соответственно, к 4,5 значениям. Но оценки  $L_{\text{п им}}$  и  $W_{\text{п им}}$  соответствующих характеристик очереди пакетов с ростом  $N$  продолжают

расти так, как растут оценки, расходящиеся к бесконечности. Растут не пропорционально  $N$ , как при отсутствии стационарного режима, а с замедлением, подобным замедлению роста логарифмической кривой [15].

Таким образом, результаты эксперимента приводят к заключению, что стационарные характеристики  $L_{\pi}$  и  $W_{\pi}$  бесконечны, в то время как стационарные характеристики  $L_{\phi}$  и  $W_{\phi}$  конечны и весьма невелики. Тем самым выполненные эксперименты подтверждают объяснение, основанное на теории [8–10] и, одновременно, подтверждают проверяемую гипотезу. На уровне файлов относительные приоритеты в исследуемой системе действуют почти так же, как абсолютные приоритеты.

Сравним применение дисциплины относительных приоритетов, определяемых бесконечной оптимизированной разметкой, с беспriorитетным режимом и с режимом абсолютных приоритетов файлов, передаваемых без пакетирования. Результаты экспериментов, выполненных с оптимизированными относительными приоритетами и в беспriorитетном режиме при длине прогона  $N = 10$  млн., а также рассчитанные по формулам (4–11) характеристики системы при режиме абсолютных приоритетов с дообслуживанием сведены в Табл. 2.

**Таблица 2.** Характеристики исследуемой СМО при различных коэффициентах загрузки

№	$\lambda$	1/30	3/30	5/30	7/30	9/30
	$\rho$	0,1	0,3	0,5	0,7	0,9
1	$(a; c)_{\text{опт}}$	(1,172; 0,5203)	(1,122; 0,5086)	(1,055; 0,5054)	(0,9512; 0,5174)	(0,8132; 0,5663)
2	$W_{\phi \text{ а/п расч}}$	0,1777	0,637	1,33	2,55	5,81
3	$W_{\phi \text{ им}}$	<b>3,20</b>	<b>3,72</b>	<b>4,49</b>	<b>5,78</b>	<b>9,00</b>
4	$L_{\phi \text{ им}}$	<b>0,107</b>	<b>0,373</b>	<b>0,749</b>	<b>1,35</b>	<b>2,70</b>
5	$W_{\pi \text{ им}}$	729,3	936,3	1291	1910	2815
6	$L_{\pi \text{ им}}$	87,83	338,3	777,4	1610	2815
7	$W_{\phi \text{ им б/п}}$	90,91	340,9	779,5	1532	2377
8	$W_{\pi \text{ им б/п}}$	744,2	993,6	1423	2169	3000
9	$L_{\phi \text{ им б/п}}$	3,033	34,11	130,0	357,8	713,6
10	$L_{\pi \text{ им б/п}}$	89,63	358,0	856,6	1828	3251

При неизменном паретовском распределении времени обслуживания в имитационной модели СМО (см. Рис. 3, Рис. 4) изменялись значения интенсивности  $\lambda$  входящего потока, вследствие чего изменялся коэффициент  $\rho = \lambda \bar{x}$  системы (см. две строки верхнего заголовка Табл. 2). При различных коэффициентах загрузки рассчитаны значения десяти характеристик СМО (указанных во втором столбце бокового заголовка).

Оптимальные значения коэффициентов  $(a, c)$ , рассчитанных в Excel по формулам (4–11), определены с помощью сервиса «Поиск решения» (строка № 1 в Табл. 2). При этом определены и соответствующие минимизированные значения среднего времени ожидания файлов ( $W_{\phi \text{ а/п расч}}$ , строка 2) для режима абсолютных приоритетов файлов (который в СПД не реализуется).

В строке 3 приведены полученные имитационные оценки  $W_{ф\text{ им}}$  времени ожидания файлов в режиме относительных приоритетов, определяемых экспоненциальными разметками с оптимизированными ( $a, c$ ), указанными в соответствующих столбцах. В строке 4 – оценки  $L_{ф\text{ им}}$  средних длин очередей файлов в этом режиме. В строках 5 и 6 – соответствующие характеристики очередей пакетов в этом же режиме. Все оценки, значения которых набраны курсивом, получены усреднением времен ожидания или длин очередей по времени прохождения через СМО 10 млн. файлов. С ростом времени моделирования они растут, поскольку стационарные значения оцениваемых характеристик бесконечны [8–10]. В строках 7–10 приведены оценки среднего времени ожидания и средней длины очереди файлов и пакетов в бесприоритетном режиме.

Модель бесприоритетного режима получается из модели, представленной на Рис. 3, если в сегменте, следующем за комментарием

\* Определение числа пакетов, на которые разбивается файл

«закомментировать» строки

```

      ASSIGN   Class, (INT(Log((P$TOBSL-K_)/C_)/A_)) ;Определение
      TEST L   P$Class,1,NEXT
      ASSIGN   Class,1
NEXT    PRIORITY (10000-P$Class)

```

(или удалить их).

Приведенные на Рис. 6 и в Табл. 1, Табл. 2 результаты экспериментов надежно подтверждают проверяемую гипотезу и в комплексе с теорией, изложенной в [8–10], обосновывают целесообразность применения метода бесконечных экспоненциальных разметок для назначения относительных приоритетов файлам в СПД с фрактальным трафиком и коммутацией пакетов.

## 6. Обсуждение результатов

**О контроле корректности ИМ.** Результаты выполненных экспериментов целесообразно проверить с помощью известных аналитических соотношений. Например, имитационные оценки  $W_{ф\text{ им}}$  и  $L_{ф\text{ им}}$  в строках 3 и 4 Табл. 2 должны при любом  $\lambda$ , т.е. в каждом из пяти столбцов таблицы, удовлетворять формуле Литтла (с небольшими погрешностями, поскольку оценки приближенные). Проверяя выполнения равенства  $L_{ф\text{ им}}(\lambda) = \lambda \cdot W_{ф\text{ им}}(\lambda)$ , нетрудно убедиться, что оно выполняется с точностью до 2–3 значащих цифр. Тем самым подтверждается корректность этих оценок и объясняется их округление в таблице до трех значащих цифр.

Аналогично с помощью формулы Литтла проверяется и подтверждается корректность оценок, полученных для уровня пакетов. Заметим, что, применяя формулы Литтла к оценкам  $W_{п\text{ им}}$  и  $L_{п\text{ им}}$  (строки 5, 6), а также к оценкам  $W_{п\text{ им } б/п}$  и  $L_{п\text{ им } б/п}$  (строки 8, 10), необходимо знать интенсивность  $\lambda_{п}$  поступления в СМО пакетов, на которые разбиваются файлы. Ее можно найти из соотношения  $\lambda_{п} \cdot b_{п} = \rho$ , где  $b_{п}$  – средняя длительность обслуживания пакета:  $b_{п} \approx b_{п\text{ им}} =$

0,8304·(см. Табл. 1). Отсюда  $\lambda_{\pi} = \rho/b_{\pi} \approx \rho/0,8304$ . Формула Литтла, применяемая, например, для сопоставления оценок  $W_{\pi \text{ им}}$  и  $L_{\pi \text{ им}}$  (строки 5, 6) при  $\rho = 0,1$  (первый столбец) имеет вид равенства  $L_{\pi \text{ им}}(\rho) = \lambda_{\pi}(\rho) \cdot W_{\pi \text{ им}}(\rho) = (\rho/0,8304) \cdot 729,3 = (0,1/0,8304) \cdot 729,3 = 87,825$ . Это расчетное значение  $L_{\pi \text{ им}}$  достаточно близко к соответствующей имитационной оценке 87,83 (строка 6, столбец 1). Подобным же образом проверяется и подтверждается корректность оценок, приведенных в Табл. 2 для показателей обслуживания пакетов при других  $\rho$ , а также для другого – бесприоритетного (строки 8, 10) режима функционирования системы.

**Незавершенные переходные процессы.** Тем не менее, оценки, набранные в Табл. 2 курсивом, не являются оценками соответствующих стационарных значений показателей, поскольку последние бесконечны. Стационарные характеристики системы на макроуровне (строки 3, 4 Табл. 2) конечны, и их оценки, как можно судить по результатам ее моделирования при длинах прогонов  $N \in \{100 \text{ тыс.}, 1 \text{ млн.}, 10 \text{ млн.}\}$ , близки к стационарным значениям.

Что касается оценок, приведенных курсивом, то они позволяют проверить, не достигается ли высокая эффективность относительных приоритетов, проявляющаяся на макроуровне, за счет ухудшения характеристик мезоуровня. Но, как видно из Табл. 2, характеристики мезоуровня, приведенные в строках 5, 6 для режима относительных приоритетов, на самом деле *несколько лучше* соответствующих характеристик бесприоритетного режима (строки 8, 10).

Что же касается характеристик макроуровня при относительных приоритетах (строки 3, 4), то они *в бесконечное число раз лучше* соответствующих характеристик бесприоритетного режима (строки 7, 9). В этом утверждении учитывается, что оценки, представленные в строках 7, 9 курсивом, соответствуют лишь некоторому моменту переходного процесса, и с ростом длины выборки приближаются к своим стационарным – бесконечным – пределам.

**Физический смысл показателей  $W_{\text{ф}}$ ,  $L_{\text{ф}}$  и дальнейшее развитие метода.** Необходимо отметить следующую особенность интерпретации показателя  $L_{\text{ф}}$  – средней длины очереди файлов в режиме относительных приоритетов. Этот показатель не выражает физического среднего объема памяти, занимаемой очередью файлов. Физическому объему памяти соответствует средняя длина  $L_{\pi}$  очереди пакетов (строка 6 Табл. 2). В стационарном режиме  $L_{\pi} = \infty$ . Следовательно, в реальной технической среде пакеты будут с некоторой вероятностью теряться, что приведет к определенной вероятности ошибок при передаче файлов. Снижение  $L_{\pi}$ , достигаемое за счет оптимизации относительных приоритетов, ничтожно; в действительности кардинально снижается только среднее время  $W_{\text{ф}}$  передачи файлов. Поэтому возникает задача дальнейшего развития и исследования рассмотренной математической модели с целью поиска методов более значительного снижения затрат памяти СПД.

Более значительного снижения  $L_{\pi}$  можно ожидать при комбинировании ме-

тогда оптимизации относительных приоритетов с методом наращивания числа параллельно функционирующих каналов в многолинейных СМО [18]. Имеет смысл, в частности, разработать и исследовать более полную модель СПД в виде сети с очередями, в которой пакеты одного и того же файла могут перемещаться между корреспондирующими пунктами разными маршрутами. Выбор для пакета одного из нескольких возможных направлений передачи в некоторой мере реализует свойства многолинейных СМО, и это позволяет ожидать более значительного сокращения длин  $L_{\text{п}}$  физических очередей к узлам.

Кроме того, предложенный в статье способ приближенной оптимизации относительных приоритетов, несмотря на высокую эффективность снижения времени  $W_{\text{ф}}$  передачи файлов, оставляет пока открытым вопрос о целесообразности разработки точного метода оптимизации (или хотя бы исследования погрешностей предложенного приближенного метода).

**Случай конечных, но больших дисперсий.** При  $\alpha > 2$  дисперсия  $D(x)$  (3) распределения Парето конечна. При этом если параметр  $\alpha$  близок к двум, то дисперсия с.в.  $x$  велика (по сравнению с квадратом м.о. этой с.в., т.е.  $C^2(x) = D(x)/M^2(x) > 1$ , причем  $C(x) \rightarrow \infty$  при  $\alpha \downarrow 2$ ). В этом случае все показатели исследуемой системы, поименованные в Табл. 2, конечны. Однако их расчет с помощью ИМ связан при таких  $\alpha$  со значительными трудностями. Для их преодоления необходимо разрабатывать соответствующие ускоренные аналитико-имитационные методы.

Рассмотрим, например, вариант исследуемой системы, в котором распределение Парето  $\text{Pa}(K; \alpha)$  времени  $x$  обслуживания (передачи) файла имеет параметр  $\alpha = 2,05$ . Тогда по формулам (2), (3) при  $K = 1$  получаем  $M(x) = \bar{x} = 1,9524$ ,  $D(x) = 37,1822$ ,  $b^{(2)} = D(x) + M^2(x) = 41,0000$ . Коэффициент загрузки  $\rho = 0,5$  достигается при  $\lambda = 0,256098$ . Оценка  $W_{\text{ф им б/п}}$  среднего времени ожидания файлов в беспriorитетном режиме при длине прогона  $N = 10$  млн. получилась равной 5,98, и при  $N = 50$  млн. – равной 5,47. Возникает впечатление, что оценка близка к стационарному среднему времени  $W_{\text{ф б/п}}$  ожидания файлов в беспriorитетном режиме, т.е. что  $W_{\text{ф б/п}}$  мало отличается от 5,47. Однако это далеко не так. Действительное значение  $W_{\text{ф б/п}}$ , которое в данном случае можно рассчитать по формуле Полачека-Хинчина, составляет

$$W_{\text{ф б/п}} = \frac{\lambda b^{(2)}}{2(1-\rho)} = \frac{0,256098 \cdot 41}{2(1-0,5)} = 10,5, \quad (13)$$

что намного превышает имитационную оценку  $W_{\text{ф им б/п}} = 5,47$ . Таким образом, здесь имеет место просто очень медленная сходимости имитационной оценки к точному значению оцениваемого показателя. Причиной столь медленной сходимости является то, что реализованная в ИМ при  $N = 10$  млн. выборка значений  $x$  еще недостаточно велика для того, чтобы в полной мере «воплотить» дисперсию  $D(x)$ , определяющую (наряду с первыми моментами), согласно формуле (13),

среднее время  $W_{\text{ф б/п}}$ . Каким же должен быть объем выборки  $N$ , чтобы оценка  $W_{\text{ф им б/п}}$  стала стабильно близка к точному значению  $W_{\text{ф б/п}}$ ?

Заметим, что при анализе сходимости оценок вторых моментов с.в.  $x$ , т.е. оценок для  $M(x^2)$ , центральная предельная теорема в данном случае неприменима, поскольку второй момент усредняемой с.в.  $(x^2)$ , т.е.  $M((x^2)^2) = M(x^4)$ , как можно легко установить по ф.р. (1), бесконечен. Поэтому оценим скорость сходимости оценки для дисперсии рассматриваемой с.в.  $x$  с распределением  $\text{Pa}(1; 2,05)$  с помощью непосредственного ИМ. Чтобы на эксперименте не сказывались дефекты модельного времени GPSS, проявляющиеся при больших объемах выборки, модельное время использовать не будем. Выполним разной длины прогоны программы, основная часть которой представлена на Рис. 7. Эта основная часть должна быть продолжена процедурами MtArandPareto и MtArand, приведенными на Рис. 4.

N_	EQU	1000000000	;определение объема выборки
K_	EQU	1	;параметр K распределения Парето
ALPHA	EQU	2.05	;параметр альфа распределения Парето
	GENERATE	1	
	ASSIGN	1,N_	;задание параметра цикла
BEG	ASSIGN	X_, (MtArandPareto(K_,ALPHA))	;начало цикла
	SAVEVALUE	1+, P\$X_	;суммирование элементов выборки
	SAVEVALUE	2+, (P\$X_ ^2)	;суммирование квадратов элементов
	LOOP	1,BEG	;возврат к началу / выход из цикла
	SAVEVALUE	MEAN, (X1/N_)	;расчет оценки для м.о.
	SAVEVALUE	B2, (X2/N_)	;расчет оценки для 2-го момента
	TERMINATE	1	;завершение прогона

Рис. 7. Программа генерации выборок разного объема из распределения  $\text{Pa}(1; 2,05)$

При длинах  $N$  прогонов 10 млн., 100 млн. и 1 млрд. оценки м.о. составили 1,9500, 1,9523 и 1,9522. Точное значение  $M(x) = 1,9524$  (здесь все хорошо). Однако для второго момента  $b^{(2)}$  получены оценки, соответственно, 13,516, 21,334 и 25,193, в то время как точное значение  $b^{(2)} = 41$ . При этом затраты времени персонального компьютера на последний прогон составили около 1 ч 3 мин. Погрешности полученных оценок второго момента – их отклонения от точного значения 41, при увеличении длины прогона на порядок сокращаются слишком медленно. Эти погрешности приблизительно пропорциональны  $(1/N^{0,1})$ . Следовательно, для «воплощения» в выборке второго момента  $b^{(2)} = 41$  с погрешностью  $\pm 1$  потребовалась бы выборка объема  $N \approx 10^{19}$  (и, следовательно, такой же длины прогон нужен для оценки стационарного значения среднего времени  $W_{\text{ф б/п}}$  ожидания файлов в беспriorитетном режиме исследуемой СМО). И такого же порядка длины прогонов нужны для оценки стационарных характеристик этой СМО в режиме относительных приоритетов. Что практически не реализуемо на современных компьютерах.

Этим примером наглядно подтверждается целесообразность разработки для дальнейших исследований соответствующих ускоренных аналитико-имитационных методов.

## 7. Заключение

В статье предложен и исследован метод кардинального снижения среднего времени передачи файлов в условиях фрактального трафика в сетях с коммутацией пакетов. Метод состоит в назначении файлам относительных приоритетов на основе бесконечной экспоненциальной разметки полубесконечного диапазона возможных значений времени обслуживания файлов. Экспоненциальные разметки, задаваемые двумя параметрами, оптимизируются для режима абсолютных приоритетов, так как такая оптимизация выполняется относительно легко и достаточно быстро, за секунды компьютерного времени. И эта разметка используется для относительных приоритетов как приближенно оптимальная. Относительные приоритеты, присвоенные файлам, наследуются пакетами, на которые файлы разбиваются для передачи в режиме коммутации пакетов.

В ходе исследований, выполненных в статье, получены следующие основные результаты.

1) Показано, что предлагаемый метод оптимизации относительных приоритетов позволяет эффективно снижать среднее время передачи файлов. Если дисперсия размеров передаваемых файлов при фрактальном трафике бесконечна, то среднее время ожидания файлов в беспriorитетном режиме также бесконечно. Назначение относительных приоритетов методом, предлагаемым в статье, делает среднее время ожидания файлов конечным и сравнимым со средним временем обслуживания файла (передачи файла по каналу).

2) Поставлены задачи развития и уточнения предложенного метода, к числу которых относятся:

- задача комбинирования предложенного метода с методом наращивания числа параллельно функционирующих каналов,
- задача более полного исследования эффективности метода на многоузловых сетевых моделях,
- задача оценки погрешностей предложенного для оптимизации относительных приоритетов приближенного метода,
- задача разработки точного метода оптимизации относительных приоритетов в условиях фрактального трафика в сетях с коммутацией пакетов.

3) Продемонстрирована медленная сходимость имитационных оценок среднего времени ожидания файлов при конечных, но больших дисперсиях степенного распределения времени обслуживания файла. Обоснована необходимость разработки соответствующих ускоренных аналитико-имитационных методов моделирования сетей передачи данных, функционирующих в условиях фрактального трафика.

## ЛИТЕРАТУРА

1. Клейнрок Л. Теория массового обслуживания. Пер. с англ. / Л. Клейнрок; пер. И. И. Грушко; ред. В. И. Нейман. – М. : Машиностроение, 1979. – 432 с.
2. Вишневский В. М. Теоретические основы проектирования компьютерных сетей / В. М. Вишневский. – М. : Техносфера, 2003. – 512 с.



3. Олифер В., Олифер Н. Компьютерные сети. Принципы, технологии, протоколы : учеб. для вузов, 5-е изд., – СПб. : Питер, 2016. – 992 с.
4. *Giambene G.* Queuing Theory and Telecommunications: Networks and Applications. 2nd ed. – New York : Springer, 2014 – 516 p.
5. *Zwart A. P.* Queueing Systems with Heavy Tails. Eindhoven: Eindhoven University of Technology, 2001. – 227 p.
6. On the Self-Similar Nature of Ethernet Traffic / W. E. Leland [et al.] // ACM SIGCOMM'93. San Fransisco, 1993. – P. 183-193.
7. *Задорожный В. Н.* Очереди и степенные распределения: монография / В. Н. Задорожный. – Новосибирск : Изд-во СО РАН, 2016. – 161 с.
8. *Задорожный В. Н., Захаренкова Т. Р.* Методы снижения вероятности потерь в системах с бесконечной дисперсией времени обслуживания // Системы управления, информационные технологии и математическое моделирование: материалы I Всерос. научн.-практ. конф. с междунар. участием. – Омск : ОмГТУ, 2019. – С. 7-26.
9. *Zadorozhnyi V. N., Zakharenkova T. R.* Methods to Reduce Loss Probability in Systems with Infinite Service Time Dispersion. Communications in Computer and Information Science // Springer, 2019. – P. 296-311.
10. *Zadorozhnyi V. N., Zakharenkova T. R., Pagano M.* Queue normalization methods in systems GI/GI/1/m with infinite variance of service time // Journal of Physics: Conference Series, 2020. – P. 1-11.
11. *Задорожный В. Н.* Особенности моделирования систем массового обслуживания с тяжелыми хвостами распределений на GPSS WORLD. Метод ARAND / Омский научный вестник, 2015. – № 3(143). – С. 307-311.
12. *Zadorozhnyi V. N.* Fractal Queues Simulation Peculiarities // Communications in Computer and Information Science // Springer, 2015. – V. 564. – P. 415-432.
13. *Задорожный В. Н.* Метод ARAND // Седьмая всерос. научн.-практ. конф. "Имитационное моделирование. Теория и практика" (ИММОД-2015), 21-23 октября 2015 года. – М : Институт проблем управления им. В.А. Трапезникова РАН, 2015. – С. 239-244.
14. *Задорожный В. Н.* Реализация больших выборок при моделировании систем массового обслуживания на GPSS WORLD // Седьмая всерос. научн.-практ. конф. "Имитационное моделирование. Теория и практика" (ИММОД-2015), 21-23 октября 2015 года. – М. : Институт проблем управления им. В.А. Трапезникова РАН, 2015. – С. 225-230.
15. *Zadorozhnyi V. N., Zakharenkova T. R.* Methods of Simulation Queueing Systems with Heavy Tails // Communications in Computer and Information Science, 2016. – V. 638. – P. 382-396.
16. *Задорожный В. Н., Захаренкова Т. Р., Мясцев С. В.* Комплекс GPSS-процедур для генерации случайных величин на основе генератора «Вихрь Мерсенна». Свидетельство о регистрации электронного ресурса № 22124 от 01 сентября 2016 года / Министерство образования и науки РФ, Объединенный фонд электронных ресурсов «Наука и образование».
17. *Задорожный В. Н.* Повышение точности GPSS-моделей путем применения генератора случайных чисел «Вихрь Мерсенна» // Омский научный вестник, 2016. – № 1 (145). – С. 90-94.
18. *Zadorozhnyi V. N., Zakharenkova T. R.* Minimization of Packet Loss Probability in Network with Fractal Traffic // Communications in Computer and Information Science // Springer, 2017. – V. 800. – P. 168-183.

---

**Задорожный** Владимир Николаевич, профессор каф. ММИИТЭ ОмГТУ, zwn2015@yandex.ru

**Pagano** Michele, профессор Пизанского университета, Италия, michele.pagano@iet.unipi.it

**Захаренкова** Татьяна Романовна, ассистент каф. ММИИТЭ ОмГТУ, zakharenkovatiana@gmail.com

УДК 519.175.4

## МЕТОДЫ КОМПЛЕКСНОЙ КАЛИБРОВКИ ГРАФОВЫХ МОДЕЛЕЙ РАСТУЩИХ СЕТЕЙ

В. Н. Задорожный, Е. Б. Юдин, М. Н. Юдина

Омский государственный технический университет, Омск, Россия

**Аннотация.** Решается актуальная задача калибровки графовых моделей больших сетей, заключающаяся в подборе таких параметров растущих случайных графов, которые обеспечивают получение требуемых вероятностных характеристик графа. Разрабатывается метод, позволяющий калибровать графы с нелинейным правилом предпочтительного связывания (НППС) одновременно по трем характеристикам: по распределению степеней связности вершин, по двумерному распределению степеней связности ребер (дуг) и по коэффициенту кластеризации графа. Для такой комплексной калибровки графов выводятся соответствующие аналитические выражения. Задача комплексной калибровки растущего графа с НППС по распределению степеней связности ребер сводится к экстремальной задаче сближения распределения степеней ребер графа с соответствующим распределением степеней связей в моделируемой сети. Предлагается метод решения этой задачи, реализуемый с использованием электронных таблиц. Приводится пример комплексной калибровки графа с НППС со сложными стохастическими приращениями. Граф калибруется по эмпирическим характеристикам моделируемой сети автономных систем интернета. Калиброванный разработанным методом граф с НППС сравнивается с калиброванным графом Барабаши-Альберт и калиброванным графом Уотса-Строгатца путем имитационного моделирования процесса распространения вируса по сети и процесса случайных потерь узлов и связей. Устанавливается, что комплексная калибровка графов с НППС обеспечивает существенное их преимущество по сравнению с другими классами графов по точности моделирования сетевых процессов.

**Ключевые слова:** графы с нелинейным правилом предпочтительного связывания, коэффициент кластеризации, распределение степеней связности вершин, распределение степеней связности ребер.

## METHODS FOR CALIBRATION OF MODELS OF GROWING NETWORKS

V. N. Zadorozhnyi, E. B. Yudin, M. N. Yudina

Omsk State Technical University, Omsk, Russia

**Abstract.** The adjusting parameters problem for generating nonlinear preferential attachment random graphs with complex stochastic increments is solved. The goal is to implement the required probabilistic characteristics of the random graph. The following characteristics of the random graph are realized: 1) the vertex degree distribution, the two-dimensional distribution of the degrees of end vertices of a randomly selected edge/arc (also known as degree-degree correlations), and the clustering coefficient. Corresponding analytical expressions for the implementation of the required characteristics derived in the article. The process of selecting graph generation parameters (calibration) is exposed as an extreme problem of convergence of the characteristics of the random graph and the corresponding characteristics of the graph of the analyzed network. This process can be implemented using spreadsheets. An example of adjusting the generation parameters of a nonlinear preferential attachment graph with complex stochastic increments to implement the corresponding probabilistic

*characteristics of the network under study is given. The empirical data were obtained by analysis of the Internet autonomous systems net-work. The adequacy of preferential attachment random graphs for network processes simulation is explored. The spread of the virus in a network is considered (the multi-agent epidemiological model “susceptible” - “infected” - “susceptible”). The processes of random loss of nodes and links are also considered. An increase in the number of probability characteristics implemented in a random graph leads to an increase in network processes simulation accuracy.*

**Keywords:** *nonlinear preferential attachment graph, clustering coefficient, vertex degree distribution, degree-degree correlation.*

## 1. Введение

С конца XX века по настоящее время в связи с актуализацией задач исследования больших сетей интенсивно разрабатываются разнообразные математические модели сетей [1–31], представляющие собой различные виды случайных графов, в том числе конфигурационные [17, 21, 25, 26, 27] и эволюционные модели [16, 19, 22, 24, 29]. Случайные графы позволяют воспроизводить важные структурные характеристики сетей: «правило шести рукопожатий» (свойство «тесного мира» [9, 27]), степенное распределение степени связности вершин («безмасштабность» сети [12, 18, 24]), наличие сообществ (высокий коэффициент кластеризации) [15, 17, 21, 27] и т.д. Широкую популярность получили растущие случайные графы [12, 18, 22, 29], которые адекватно воспроизводят механизмы генезиса моделируемых сетей и позволяют прогнозировать динамику их развития. Самым известным классом растущих графов являются графы предпочтительного связывания, предложенные американскими учеными А. Барабаши и Р. Альберт (графы БА [7]). Позже стали появляться модификации и обобщения графов БА, относящиеся к семейству графов предпочтительного связывания [12, 18, 22, 29].

Графы БА обладают свойствами «тесного мира» и характеризуются следующим финальным (определяемым при числе вершин  $N \rightarrow \infty$ ) асимптотически степенным распределением степени связности (РСС) вершин:  $Q_k \propto k^{-3}$  при  $k \rightarrow \infty$ , где  $Q_k$  – вероятность того, что случайно выбранная в графе вершина будет иметь степень связности  $k$ . Подобные асимптотически степенные РСС описывают степени связности узлов в большом числе известных растущих сетей и дают основание называть такие сети безмасштабными (scale-free) сетями. Анализ графов БА помог объяснить ряд ранее установленных фактов [9], например, таких как устойчивость интернета и ряда других безмасштабных сетей к случайным отказам их элементов – узлов и связей, восприимчивость этих сетей к вирусам и чувствительность к атакам на так называемые «хабы» – узлы с высокой степенью связности.

Однако графы БА оказались достаточно «грубой» моделью сетей. Выяснилось, что в исследуемых больших сетях асимптотически степенное по  $k$  распределение  $Q_k \propto k^{-\alpha}$  имеет параметр  $\alpha < 3$  [6, 24], в то время как в графе БА  $\alpha = 3$ . В последние годы также ведется дискуссия о «безмасштабности» больших сетей [10, 8]. Поэтому большое значение имеют графы предпочтительного связывания, РСС вершин которых можно «регулировать» [12, 22], т.е. которые можно калибровать для реализации любого заданного РСС вершин. Исследования в области

разработки новых графов предпочтительного связывания ведутся еще и по причине того, что самый известный граф предпочтительного связывания – граф БА – отличается от графов больших сетей и по другим важным вероятностным характеристикам. Самой известной из них является двумерное РСС концевых вершин дуг/ребер (РСС дуг/ребер) графа [11, 13, 20, 23]. Причем для некоторых графов предпочтительного связывания [3, 28, 30] удастся получить формулы, выражающие РСС дуг/ребер непосредственно через параметры графа. Для графов с нелинейным правилом предпочтительного связывания (НППС), предложенных в [2], такие формулы удастся использовать для одновременной реализации известных РСС вершин и РСС дуг/ребер моделируемой сети [31].

Другой важной вероятностной характеристикой графовых моделей сетей является коэффициент кластеризации (доля «треугольников» среди всех связных подграфов на трех вершинах). Выяснилось, что графы БА имеют существенно заниженный по сравнению с моделируемыми сетями коэффициент кластеризации [15]. Поэтому актуальной задачей исследования больших сетей является разработка графов предпочтительного связывания, которые могут калиброваться одновременно не только по РСС вершин и дуг/ребер, но и по коэффициенту кластеризации.

Наиболее гибкие возможности калибровки достигаются при моделировании больших сетей случайными графами с НППС. В [4] предложены графы с НППС со сложными стохастическими приращениями (НППС-СП) и выведены выражения, обеспечивающие возможность калибровки графа с НППС-СП одновременно по РСС вершин и по коэффициенту кластеризации.

В настоящей статье предлагается обобщение графа с НППС-СП и выводятся соотношения, позволяющие калибровать графы с НППС-СП сразу по трем структурным характеристикам: по РСС вершин, по РСС ребер и по коэффициенту кластеризации. Демонстрируются преимущества использования калиброванных графов с НППС-СП при исследовании сетевых процессов.

## 2. Графы с НППС со сложными стохастическими приращениями

### 2.1. Ориентированный граф Барабаши-Альберт

Выращивание ориентированного графа БА выполняется итерационно, начиная с некоторого графа-затравки. На каждом шаге выращивания графа к нему добавляется приращение – новая вершина с  $m$  исходящими из нее дугами. Концы дуг приращения связываются с вершинами графа случайно, причем вероятность  $p_i$  выбора дугой вершины  $i$  для связывания с ней зависит от степени  $k_i$  связности этой вершины:

$$p_i = k_i / \sum_{j=1}^N k_j, \quad i = 1, \dots, N,$$

где  $N$  – число вершин в графе перед добавлением приращения.

Такой граф реализует принцип «богатый становится богаче»: чем больше степень вершины, тем больше вероятность выбора этой вершины для связывания (т.е. увеличения ее степени).

Заметим, что ориентированный граф БА можно модифицировать, определяя число дуг в приращении не как константу  $m$ , а как целочисленную положительную случайную величину  $x$  со средним значением  $m$  (не обязательно целым). Такой подход реализуется, например, в библиотеке `igraph` [14].

Заметим также, что если в ориентированном графе БА дуги заменить ребрами, то получится неориентированный граф БА, предложенный в известной работе [7]. Как правило, удобнее исследовать ориентированный граф БА. При необходимости несложно от найденных его свойств перейти к соответствующим свойствам неориентированной версии графа.

### 2.1. Графы с нелинейным правилом предпочтительного связывания

В ходе генерации ориентированного графа с НППС при добавлении очередной новой вершины каждая из  $x$  инцидентных ей дуг связывается с одной из  $N$  вершин графа, которая выбирается с вероятностью

$$p_i = f(k_i) / \sum_{j=1}^N f(k_j), \quad i = 1, \dots, N, \quad (1)$$

где  $f(k) \equiv f_k \geq 0$  – некоторая произвольно заданная функция от степени  $k$  вершины (функция предпочтения),  $f(k_i)$  – предпочтительность (вес) вершины  $i$ ,  $x \geq g$  – целочисленная положительная случайная величина с некоторым произвольно заданным распределением  $r_k = P(x = k)$ ;  $g \geq 1$  – целое число. Очевидно, что граф БА является частным случаем графа с НППС, получаемым при  $f(k) = k$ ,  $x \equiv m = \text{const}$ . Процесс выращивания графа с НППС иллюстрируется на Рис. 1.

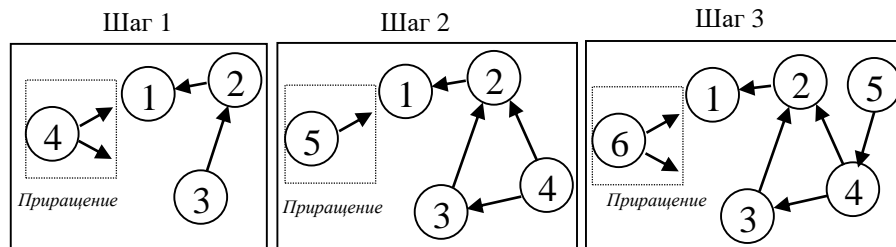


Рис.1. Пример первых шагов выращивания графа с НППС. Число дуг в приращениях разыгрывается по распределению  $r_k$

Если используется функция  $f(k) = \ln(k)$ , то можно говорить, что моделируется закон Вебера-Фехнера, в соответствии с которым интенсивность ощущения пропорциональна логарифму интенсивности раздражителя [2]. Если используется функция  $f(k) = k^\gamma$ , то реализуются так называемые графы со степенным предпочтением [18]. Частными случаями графов с НППС являются многие известные графы предпочтительного связывания, в том числе графы, предложенные в [12, 22] (о графах [22] как частном случае графов с НППС см. [30]).

В работе [2] установлено, что РСС вершин графа с НППС рассчитывается по рекуррентной формуле

$$Q_k = \frac{r_k \langle f \rangle + m f_{k-1} Q_{k-1}}{\langle f \rangle + m f_k}, \text{ при } k = g, g+1, \dots, M, \quad (2)$$

где  $Q_k$  – финальная вероятность того, что случайно выбранная вершина будет иметь степень связности  $k$ ;  $g$  – минимальная степень вершин;  $M \leq \infty$  – максимальная степень вершин (если функция предпочтения  $f(k) = f_k$  положительна в конечном промежутке  $g \leq k \leq M-1$  и равна нулю при других  $k$ , то максимальная степень вершин конечна и равна  $M$ );  $\langle f \rangle$  – финальный средний вес вершин графа;  $m = M(x) = \sum_k k \cdot r_k$  – среднее число дуг в приращении графа. Так как вес  $f_{g-1} = 0$ , то при вычислении  $Q_g$  второе слагаемое в числителе формулы (2) определено и равно нулю.

Финальный средний вес вершин графа с НППС определяется формулой

$$\langle f \rangle = \sum_{k=g}^M f_k Q_k \quad (3)$$

и рассчитывается описанным в [2] численным методом.

Задача калибровки ориентированных графов с НППС по РСС вершин решается путем элементарного «обращения» рекуррентной формулы (2). При этом можно положить, что средний вес  $\langle f \rangle = m$  или любой другой константе, поскольку от среднего веса результат применения формулы (1) не зависит (все веса в ней можно умножить на любую общую для них константу, при этом  $p_i$  не изменятся). Задача калибровки неориентированных графов решается в [29, 31] путем нахождения распределения  $r_k$  и функции предпочтения  $f_k$ , реализующих требуемое финальное РСС  $Q_k$ . В результате задача калибровки по РСС решается точно и в общем случае имеет бесконечное множество точных решений, чем обеспечивается достаточное число степеней свободы для одновременной калибровки графа и по другим его характеристикам.

Задача калибровки ориентированных графов с НППС по финальному РСС ребер решается в [31]. Решение основано на следующей ранее выведенной формуле для РСС  $Q_{l,k}$  дуг графа с НППС: при заданных параметрах графа (распределении  $r_k$  и функции предпочтения  $f_k$ ) имеем

$$Q_{l,k} = \frac{f_{k-1}(l r_l Q_{k-1} + m^2 Q_{l,k-1}) + f_{l-1} m^2 Q_{l-1,k}}{m(\langle f \rangle + m f_k + m f_l)}, \quad l, k \geq g, \quad (4)$$

где  $Q_{l,k}$  – вероятность того, что случайно выбранная в графе дуга исходит из вершины со степенью  $l$  и заходит в вершину со степенью  $k$ . При  $l = g$  и/или  $k = g$  соответствующие слагаемые в числителе формулы (4) с множителями  $f_{l-1} = f_{g-1}$  и/или  $f_{k-1} = f_{g-1}$ , равными нулю, определены и равны нулю.

Для частного случая – для графов БА в [3] получена формула, представляющая собой соответствующий частный случай формулы (4).

Заменяя дуги графа ребрами, получаем неориентированный граф с НППС. Если обозначить через  $\mathbf{Q}$  матрицу вероятностей  $\mathbf{Q} = \|Q_{lk}\|$ , задающую РСС дуг, то матрицу  $\mathbf{\Theta} = \|\Theta_{lk}\|$ , задающую РСС ребер, можно вычислить по формуле

$\Theta = (\mathbf{Q} + \mathbf{Q}^T)/2$ , где  $T$  – символ транспонирования. Вероятность  $\Theta_{lk}$  интерпретируется как вероятность того, что случайно выбранное ребро графа, проходящее в случайно выбранном направлении, ведет из вершины со степенью  $l$  в вершину со степенью  $k$ .

Задача калибровки графа по РСС  $Q_k$  вершин и РСС  $\Theta_{lk}$  ребер заключается в нахождении распределения  $r_k$  и функции предпочтения  $f_k$ , которые порождают граф с заданным финальным РСС вершин  $Q_k$  и обеспечивают минимум отклонения РСС  $\Theta_{lk}$  ребер графа от заданного РСС  $\tilde{\Theta}_{l,k}$  связей в сети:

$$\rho(\Theta, \tilde{\Theta}) = \left[ \sum_{l,k=g}^u (\Theta_{l,k} - \tilde{\Theta}_{l,k})^2 \right]^{\frac{1}{2}} \rightarrow \min, \quad (5)$$

где  $u \leq M$  – степень, выбираемая для ограничения сверху характерного промежутка степеней, для которого целесообразно получить максимальную точность реализации заданного РСС ребер. Таким образом, калибровка графа с НППС по РСС ребер сводится к решению задачи (5). При расчете отклонения  $\rho(\Theta, \tilde{\Theta})$  по формуле (5) ее правая часть выражается через параметры графа с помощью формул (2)–(4) и формулы  $\Theta = (\mathbf{Q} + \mathbf{Q}^T)/2$ .

### 2.1. Графы со сложными стохастическими приращениями

Для решения проблемы заниженного коэффициента кластеризации в графах предпочтительного связывания в работе [4] предложена и исследована версия графов с НППС, выращиваемых путем добавления сложных приращений (графы с НППС-СП). На каждом шаге генерации такого графа к нему добавляется приращение, которое с вероятностью  $(1 - \gamma)$  является обычным (простым) приращением, с вероятностью  $\gamma$  – сложным.

Сложное приращение состоит из двух вершин – вершины А (ведущей) и вершины В (ведомой). Случайное число дуг, исходящих из ведущей вершины А, разыгрывается по заданному распределению вероятностей  $r_k''$ , а число дуг, исходящих из вершины В, равно числу дуг, исходящих из вершины А.

Одна из дуг вершины В всегда заходит в вершину А, поэтому степень связности вершины А всегда на единицу больше степени связности вершины В (см. шаг 2 на Рис. 2).

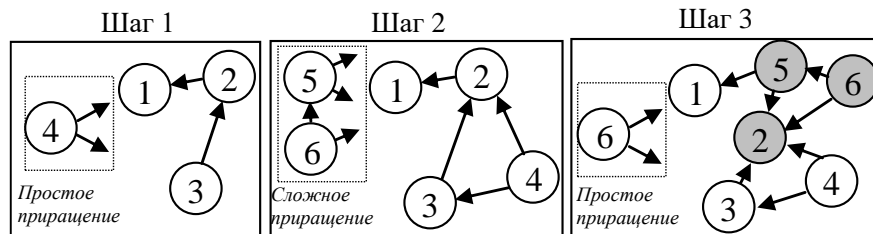


Рис. 2. Пример первых шагов генерации графа с НППС-СП. На шаге 1 добавляется простое приращение, на шаге 2 – сложное. Его вершина 5 ведущая, вершина 6 – ведомая. Его дуги (5→2) и (6→2) – сопряженные (см. шаг 3)

Свободные концы дуг сложного приращения присоединяются к вершинам графа следующим образом. Сначала одна из свободных дуг вершины В выбирает по правилу (1) вершину  $i$  для связывания и присоединяется к ней. Далее с вероятностью  $P$  принимается решение о внедрении «треугольника». Если выпадает решение внедрить треугольник, то к выбранной вершине  $i$  присоединяется также одна из дуг, исходящих из вершины А. Эти действия повторяются, пока не свяжутся все дуги вершины В. Остающиеся свободными дуги вершины А присоединяются к вершинам графа по правилу (1) независимо.

Число дуг простого приращения описывается распределением  $r'_k$ . Граф, выращиваемый посредством смешанных приращений – простых и сложных, будем называть гомогенным, если распределения  $r'_k$  и  $r''_k$  совпадают. В противном случае будем называть граф гетерогенным. Выражения для РСС вершин и РСС ребер гомогенных графов выведены в [4]. В следующем разделе выводятся выражения для характеристик гетерогенных графов.

### 3. Анализ гетерогенных графов с НППС-СП

#### 3.1. Элементарные свойства смешанных приращений гетерогенного графа

Среднее число вершин, добавляемых за один шаг выращивания графа, равно  $1(1 - \gamma) + 2\gamma = 1 + \gamma$ .

Среднее число дуг простого приращения составляет  $m' = \sum k r'_k$ , среднее число дуг, исходящих из одной вершины сложного приращения, равно  $m'' = \sum k r''_k$ . За один шаг выращивания графа к нему добавляется в среднем  $(1 - \gamma)m' + 2\gamma m''$  дуг. На одну вершину в выращенном бесконечном графе приходится в среднем  $m = [(1 - \gamma)m' + 2\gamma m''] / (1 + \gamma)$  дуг. Наименьшее число дуг, добавляемых за один шаг, равно  $g$ ,  $g \geq 1$ . Для простоты будем полагать, что параметр  $g$  у простых и сложных приращений имеет одно и то же значение.

Если вероятность  $P$  сопряжения свободной дуги ведомой вершины с дугой ведущей вершины равна единице, то среднее число внедряемых сложным приращением треугольников составляет  $\mu = (m'' - 1)$ . В общем случае среднее число внедряемых сложным приращением треугольников  $\mu = (m'' - 1)P$ . Среднее число треугольников, внедряемых за один шаг, равно  $\mu\gamma$ .

Из того, что на одну вершину графа приходится в среднем  $m$  дуг, вытекает, что финальная средняя степень вершин графа  $\langle k \rangle = 2m$ .

#### 3.2. РСС вершин гетерогенного графа с НППС-СП

Задачу вывода РСС вершин для гетерогенного графа с НППС-СП решим путем составления уравнения баланса вероятностей для финального РСС вершин графа [2]. В растущем графе в стационарном режиме, т.е. при  $N \rightarrow \infty$  средняя доля вершин со



степенью  $k$  на любом очередном шаге выращивания графа не изменяется. Для гетерогенного графа с НППС-СП имеем следующее уравнение баланса:

$$\frac{|A_k|}{N} = \frac{|A_k| + \Delta}{N + (1 + \gamma)}, \quad k \geq g, \quad (6)$$

где  $A_k$  – множество вершин графа, имеющих степень  $k$ ,  $|A_k|$  – число вершин в множестве  $A_k$ ,  $\Delta$  – среднее изменение числа вершин в множестве  $A_k$  после связывания с приращением,  $(1 + \gamma)$  – среднее число вершин, добавляемых приращением. Перепишывая (6) в виде  $|A_k|N + |A_k|(1 + \gamma) = |A_k|N + \Delta \cdot N$  или  $\frac{|A_k|}{N}(1 + \gamma) = \Delta$  и, замечая, что  $|A_k|/N \sim Q_k$ , получаем

$$Q_k(1 + \gamma) = \Delta, \quad k \geq g. \quad (7)$$

Остается выразить здесь  $\Delta$  через  $Q_k$  и известные параметры графа, чтобы из получаемого уравнения найти все  $Q_k$ .

При условии поступления обычного приращения число вершин  $|A_k|$  в множестве  $A_k$  за один шаг выращивания графа с вероятностью  $r'_k$  увеличивается на единицу, поскольку с этой вероятностью вершина приращения имеет степень  $k$ , т.е. попадает в множество  $A_k$ . Поэтому  $|A_k|$  увеличивается за один шаг в среднем на  $1 \cdot r'_k = r'_k$ . При связывании добавляемых дуг с вершинами графа в среднем  $m'P_k$  вершин будет выбрано для связывания из множества  $A_k$ , где  $P_k$  – финальная при  $N \rightarrow \infty$  вероятность связывания дуги с вершиной множества  $A_k$ . При этом после присоединения новых дуг к вершинам множества  $A_k$  степени выбранных вершин увеличатся на единицу и вершины «перейдут» в множество  $A_{k+1}$ . Таким образом, в среднем число вершин в множестве  $A_k$  уменьшится на  $m'P_k$ . В то же время в среднем  $m'P_{k-1}$  вершин будут выбраны для связывания из множества  $A_{k-1}$ , и после присоединения к ним новых дуг эти вершины «перейдут» в множество  $A_k$ , за счет чего оно увеличится на  $m'P_{k-1}$  вершин. Суммируя перечисленные изменения (иллюстрируемые на Рис. 3 слева), найдем среднее приращение  $\Delta_1$  величины  $|A_k|$  при условии поступления простого приращения графа:

$$\Delta_1 = r'_k + m'P_{k-1} - m'P_k. \quad (8)$$

Здесь вероятность  $P_k$  связывания дуги приращения с одной из вершин  $v_i$  множества  $A_k$  найдем при  $N \rightarrow \infty$  с учетом формулы (1) следующим образом:

$$P_k = \sum_{v_i \in A_k} p_i = \frac{1}{N} \sum_{j=1} f(k_j) \sum_{v_i \in A_k} f(k_i) = \frac{|A_k| f(k)}{N \langle f \rangle} \sim \frac{Q_k f_k}{\langle f \rangle}. \quad (9)$$

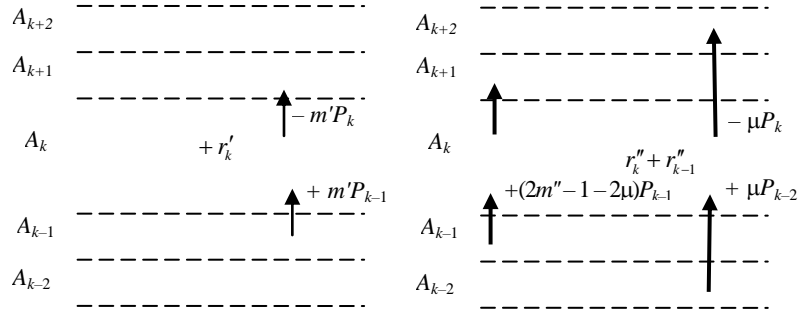


Рис. 3. Схема изменения числа вершин во множестве  $A_k$  при добавлении простого (слева) и сложного (справа) приращения

Аналогично определяется среднее приращение числа вершин в множестве  $A_k$ , происходящее в результате поступления сложного приращения графа:

$$\Delta_2 = r''_k + r''_{k-1} + P(m'' - 1)P_{k-2} + \\ + [2(1 - P)(m'' - 1) + 1]P_{k-1} - (2m'' - Pm'' + P - 1)P_k. \quad (10)$$

Здесь  $m''$  – среднее число дуг вершины сложного приращения,

$P$  – вероятность сопряжения дуги вершины В с дугой вершины А,

$P(m'' - 1)P_{k-2}$  – среднее число вершин, выбранных парами сопряженных дуг во множестве  $A_{k-2}$ , переведенных в  $A_k$ ,

$[2(1 - P)(m'' - 1) + 1]P_{k-1}$  – среднее число вершин, выбранных одиночными дугами в множестве  $A_{k-1}$  и переведенных в  $A_k$ ,

$(2m'' - Pm'' + P - 1)P_k$  – среднее число выборов вершин в множестве  $A_k$  одиночными (не сопряженными) и сопряженными дугами.

Вероятности  $r''_k$  и  $r''_{k-1}$  суммируются потому, что вершины А и В одновременно не могут попасть в множество  $A_k$ , т.е. попадания их в это множество – несовместные события.

Поскольку среднее приращение  $\Delta$  в (7) является суммой  $\Delta = (1 - \gamma) \Delta_1 + \gamma \Delta_2$ , то, учитывая (7) – (10), запишем:

$$Q_k(1 + \gamma) = \Delta = (1 - \gamma)(r'_k + m'P_{k-1} - m'P_k) + \gamma r''_k + \gamma r''_{k-1} + \gamma P(m'' - 1)P_{k-2} + \\ + \gamma [2(1 - P)(m'' - 1) + 1]P_{k-1} - \gamma (2m'' - Pm'' + P - 1)P_k = \\ = (1 - \gamma)r'_k + (1 - \gamma)m' \frac{Q_{k-1}f_{k-1}}{\langle f \rangle} - (1 - \gamma)m' \frac{Q_k f_k}{\langle f \rangle} + \\ + \gamma r''_k + \gamma r''_{k-1} + \gamma P(m'' - 1) \frac{Q_{k-2}f_{k-2}}{\langle f \rangle} + \\ + \gamma [2(1 - P)(m'' - 1) + 1] \frac{Q_{k-1}f_{k-1}}{\langle f \rangle} - \gamma (2m'' - Pm'' + P - 1) \frac{Q_k f_k}{\langle f \rangle}.$$

Собирая все члены с  $Q_k$  в левой части полученного уравнения, выражаем эту вероятность через остальные параметры и получаем следующую рекуррентную формулу:

$$Q_k = \frac{A_k + B_{k-1}Q_{k-1} + C_{k-2}Q_{k-2}}{(1 + \gamma)\langle f \rangle + [(1 - \gamma)m' + \gamma(2m'' - Pm'' + P - 1)]f_k}, \quad k \leq M, \quad (11)$$

где

$$A_k = (1 - \gamma)r'_k\langle f \rangle + \gamma r''_k\langle f \rangle + \gamma r''_{k-1}\langle f \rangle,$$

$$B_{k-1} = [(1 - \gamma)m' + 2\gamma(1 - P)(m'' - 1) + \gamma]f_{k-1},$$

$$C_{k-2} = \gamma P(m'' - 1)f_{k-2}.$$

В формулу (11) входит неизвестный параметр  $\langle f \rangle = \sum_{k \geq g} Q_k f_k$ , который при

заданной последовательности значений  $f_k$  невозможно рассчитать до определения всех  $Q_k$ , а их невозможно рассчитать без определения значения  $\langle f \rangle$ . В действительности же расчет всех  $Q_k$  по формуле (11) легко выполняется следующим образом. Вначале все  $Q_k$  рассчитываются через  $\langle f \rangle$ , которому присвоено некоторое произвольное значение, а затем подбирается такое  $\langle f \rangle$ , при котором выполняется равенство  $\langle f \rangle = \sum_{k \geq g} Q_k f_k$ . Это удобно делать, например, в среде электронных таблиц.

Для контроля правильности подбора  $\langle f \rangle$  и соответствующих ему значений  $Q_k$  используется проверочное равенство  $\langle k \rangle = \sum_k k Q_k = 2m$ , где, согласно разделу 3.1 статьи,  $m = [(1 - \gamma)m' + 2\gamma m'']/(1 + \gamma)$ .

Заметим, что при  $\{r'_k\} = \{r''_k\} = \{r_k\}$  получаем  $m' = m'' = m$ , и формула (11) для РСС  $Q_k$  гетерогенного графа с НППС-СП превращается, как и следовало ожидать, в уравнение для финального РСС вершин гомогенного графа с НППС-СП [4].

Заметим также, что при  $k < g$  числитель в формуле (11) превращается в нуль, знаменатель – в  $(1 + \gamma)\langle f \rangle$ , т.е. при  $k < g$  значения  $Q_k$  этой формулой определены и равны нулю.

### 3.3. Калибровка графа с НППС-СП по РСС вершин

Задача калибровки графа по РСС вершин заключается в том, чтобы по заданному РСС  $Q_k$  вершин найти параметры генерации графа:  $\gamma$ ,  $P$ , распределения  $r'_k, r''_k$  и веса  $f_k$  – такие, при которых выращенный граф будет иметь РСС вершин  $Q_k$ .

Эта задача решается обращением рекуррентной последовательности (11): в этой последовательности требуется параметры графа выразить через заданные вероятности  $Q_k$ . Для этого формулу (11) можно несколько упростить, так как в отличие от задачи расчета  $Q_k$ , решаемой формулой (11), в задаче калибровки мы не связаны с заранее заданной последовательностью  $f_k$ .

Результаты вычислений по формуле (1) и, соответственно, по формуле (11) не зависят от мультипликативной константы, на которую можно умножить все  $f_k$ . Действительно, если в формуле (11) все  $f_k$  умножить на одно и то же число (при этом

автоматически в это же число раз изменится и среднее  $\langle f \rangle$ ), то определяемые полученными  $f_k$  значения  $Q_k$  не изменятся. Используя это свойство, примем в формуле (11)  $\langle f \rangle / a = 1$  (и определим тем самым равенство  $\langle f \rangle = a$ ), где

$$a = (1 - \gamma)m' + \gamma(2m'' - Pm'' + P - 1). \quad (12)$$

В результате формула (11) примет такой вид:

$$Q_k = \frac{A_k + B_{k-1}Q_{k-1} + C_{k-2}Q_{k-2}}{(1 + \gamma)a + af_k}, \quad (13)$$

где

$$\begin{aligned} A_k &= [(1 - \gamma)r'_k + \gamma r''_k + \gamma r''_{k-1}]a, & B_{k-1} &= [(1 - \gamma)m' + 2\gamma(1 - P)(m'' - 1) + \gamma]f_{k-1}, \\ C_{k-2} &= \gamma P(m'' - 1)f_{k-2}. \end{aligned} \quad (14)$$

Теперь заметим, что значения  $\gamma$  и  $P$  в формуле (13) можно считать известными (их можно будет варьировать для калибровки графа по коэффициенту кластеризации). Поэтому остается найти распределения вероятностей  $r'_k$ ,  $r''_k$  и значения весов  $f_k$ , отвечающие формулам (13), (14) и условию  $a = \langle f \rangle$ . Выражая из (13)  $f_k$  через  $f_{k-1}$  и  $f_{k-2}$ , получаем рекуррентную формулу, пригодную для расчета всей последовательности весов:

$$\begin{aligned} f_k &= \frac{A_k - Q_k(1 + \gamma)a + B_{k-1}Q_{k-1} + C_{k-2}Q_{k-2}}{aQ_k} = [(1 - \gamma)r'_k + \gamma r''_k + \gamma r''_{k-1}] \frac{1}{Q_k} - (1 + \gamma) + \\ &+ [a - \gamma P(m'' - 1)] \frac{f_{k-1}Q_{k-1}}{aQ_k} + \gamma P(m'' - 1) \frac{f_{k-2}Q_{k-2}}{aQ_k}, \end{aligned} \quad (15)$$

где выражения для  $A_k$ ,  $B_{k-1}$  и  $C_{k-2}$  приведены в (14).

Расчет последовательности весов  $f_k$  по этой формуле легко и естественно выполняется в электронных таблицах. При этом распределения  $r'_k$ ,  $r''_k$ , в принципе, могут быть заданы произвольно; однако для сохранения физического смысла рекомендуется задавать их так, чтобы были реализованы такие средние  $m'$  и  $m''$ , которые формируют требуемую известную для моделируемой сети среднюю степень  $\langle k \rangle = 2m$  (используемые для этого формулы см. в разделе 3.1 статьи). Это тоже может быть обеспечено множеством способов; из них необходимо выбирать такие, при которых веса, рассчитываемые по формуле (15), все неотрицательны. Все эти условия легко обеспечиваются в электронных таблицах, в том числе с помощью сервиса «Поиск решения».

Пример результатов калибровки, выполненной в электронной таблице, приведен на Рис. 4. По известному РСС  $Q_k$  узлов моделируемой сети автономных систем (сети АС) интернета получены параметры генерации графа (в том числе последовательность весов  $f_k$ ), при которых выращенный граф имеет это же РСС у вершин. В последнем столбце выполнен расчет РСС вершин графа, выращиваемого при полученных параметрах генерации. Для расчета использована формула (11). При этом расчет значительно упрощается, поскольку известно, что

$\langle f \rangle = 1$ . Диаграмма на листе электронной таблицы позволяет видеть, что заданное РСС  $Q_k$  (маркеры) совпадает с РСС вершин выращиваемого графа (непрерывная линия). Это можно видеть и путем непосредственного сравнения чисел в столбцах  $Q_k$  задан. и  $Q_k$  расч.

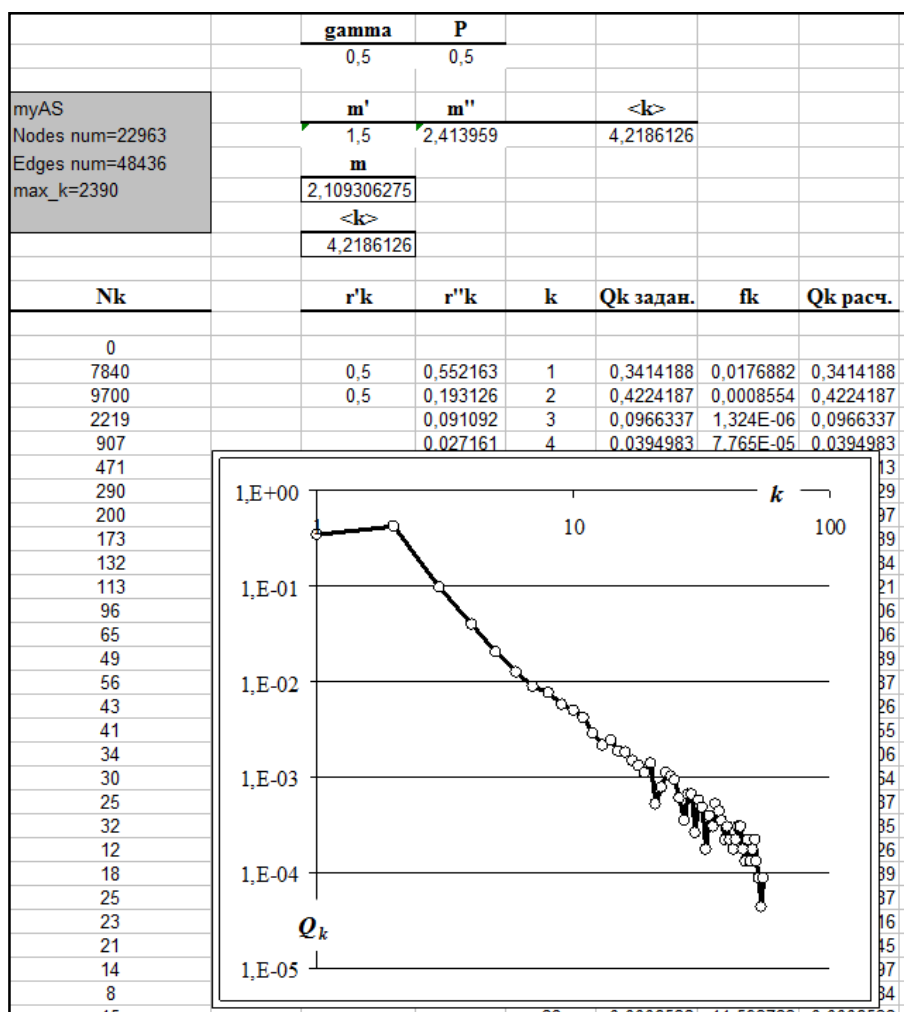


Рис. 4. Фрагмент таблицы с калибровкой графа по РСС вершин

### 3.4. РСС дуг и ребер гетерогенного графа с НППС-СП

Выведем формулы для РСС дуг и РСС ребер гетерогенного графа с НППС-СП путем составления уравнения баланса финальных РСС дуг. Обозначим через  $B(l,k)$  множество дуг графа, начальная вершина которых имеет степень  $l$ , а конечная – степень  $k$ . Через  $A_k$  по-прежнему будем обозначать множество вершин, имеющих степень  $k$ .

1. Рассмотрим изменение числа дуг в множестве  $B(l,k)$  при добавлении простого приращения.

1.1. Число  $|B(l,k)|$  дуг в множестве  $B(l,k)$  при добавлении простого приращения уменьшается, когда дуга этого приращения связывается с вершиной множества  $A_k$ : эта вершина переходит в множество  $A_{k+1}$ , и те дуги множества  $B(l,k)$ , которые в нее заходили, «уходят» в множество  $B(l,k+1)$ . В среднем на одну вершину таких дуг приходится

$$\frac{|B_{l,k}|}{|A_k|} \sim \frac{mNQ_{l,k}}{NQ_k} = \frac{mQ_{l,k}}{Q_k}, \quad (16)$$

где  $mN$  – среднее число дуг в графе. С учетом вероятности  $P_k$  того, что новая дуга выберет вершину из  $A_k$ , и среднего числа  $m'$  дуг в простом приращении находим, что среднее число дуг, выводимых из  $B(l,k)$  простым приращением, составляет

$$\frac{|B_{l,k}|}{|A_k|} P_k m' \sim \frac{mQ_{l,k}}{Q_k} P_k m' = mm' \frac{Q_{l,k}}{Q_k} \frac{Q_k f_k}{\langle f \rangle} = mm' \frac{Q_{l,k} f_k}{\langle f \rangle}.$$

1.2. Число дуг в множестве  $B(l,k)$  уменьшается и в случае, когда очередная дуга приращения связывается с вершиной со степенью связности  $l$ . По аналогии с пунктом 1.1 находим, что среднее число выводимых за счет этого из  $B(l,k)$  дуг составляет

$$\frac{|B_{l,k}|}{|A_l|} \sim \frac{mQ_{l,k}}{Q_l} P_l m' = mm' \frac{Q_{l,k}}{Q_l} \frac{Q_l f_l}{\langle f \rangle} = mm' \frac{Q_{l,k} f_l}{\langle f \rangle}.$$

1.3. С другой стороны в множество  $B(l,k)$  добавляются дуги при связывании дуг простого приращения с вершинами множеств  $A_{k-1}$  и  $A_{l-1}$ . За счет этого в состав множества  $B(l,k)$  добавляется в среднем  $mm' \frac{Q_{l,k-1} f_{k-1}}{\langle f \rangle} + mm' \frac{Q_{l-1,k} f_{l-1}}{\langle f \rangle}$  дуг.

1.4. Кроме того, каждая дуга простого приращения добавляется к множеству  $B(l,k)$ , если степень вершины этого приращения равна  $l$ , а дуга связывается с вершиной, имеющей степень  $(k-1)$ . Среднее число таких дуг составляет  $lr'_l \frac{Q_{k-1} f_{k-1}}{\langle f \rangle}$ .

2. Рассмотрим изменение числа дуг в множестве  $B(l,k)$  при добавлении сложного приращения.

2.1. За счет присоединения в среднем  $2m'' - Pm'' + P - 1$  одиночных дуг и пар сопряженных дуг (см. (10)) сложного приращения к вершинам множества  $A_k$  из  $B(l,k)$  выводится в среднем  $m(2m'' - Pm'' + P - 1) \frac{Q_{l,k} f_k}{\langle f \rangle}$  дуг.

2.2. Аналогично за счет присоединения одиночных и сопряженных дуг сложного приращения к вершинам множества  $A_l$  из  $B(l,k)$  выводится в среднем  $m(2m'' - Pm'' + P - 1) \frac{Q_{l,k} f_l}{\langle f \rangle}$  дуг.

2.3. При добавлении свободных одиночных дуг сложного приращения к вершинам множеств  $A_{k-1}$  и  $A_{l-1}$  в состав множества  $B(l,k)$  добавляется в среднем  $m(2m'' - 2Pm'' + 2P - 1) \frac{Q_{l,k-1} f_{k-1}}{\langle f \rangle}$  и  $m(2m'' - 2Pm'' + 2P - 1) \frac{Q_{l-1,k} f_{l-1}}{\langle f \rangle}$  дуг, соответственно.

2.4. Пары сопряженных дуг добавляют в  $B(l,k)$  в среднем  $mP(m''-1) \frac{Q_{l-2,k}f_{l-2}}{\langle f \rangle}$  и  $mP(m''-1) \frac{Q_{l,k-2}f_{k-2}}{\langle f \rangle}$  дуг.

2.5. Кроме того, сами дуги сложного приращения могут добавляться в  $B(l,k)$ .

2.5.1. Если вершины А, В приращения имеют по  $l$  исходящих дуг (что случается с вероятностью  $r_l''$ , Рис. 5 а), то в среднем  $(1-P)(l-1)$  дуг вершины В, не сопряженных с дугами вершины А, с вероятностью  $P_{k-1}$  каждая связываются с вершинами, принадлежащими множеству  $A_{k-1}$ , которые в результате переходят в множество  $A_k$ , и сами дуги переходят в множество  $B(l,k)$ . В среднем таким путем в  $B(l,k)$  добавляется  $(1-P)(l-1)P_{k-1}$  новых дуг. При этом если  $k = l + 1$ , то дуга, ведущая из В в А, также добавляется в  $B(l,k)$ , т.е. при любом  $k$  в множество  $B(l,k)$  добавляется ровно  $\delta_{l+1,k}$  дуг, связывающих А с В ( $\delta_{l+1,k} = 1$  при  $l + 1 = k$  и  $\delta_{l+1,k} = 0$  в противном случае). Учитывая вероятность  $r_l''$  того, что В имеет  $l$  исходящих дуг, и, выражая  $P_{k-1}$  через известные величины, находим, что при поступлении сложного приращения множество  $B(l,k)$  увеличивается в среднем на  $r_l''(1-P)(l-1) \frac{Q_{k-1}f_{k-1}}{\langle f \rangle} + r_l''\delta_{l+1,k}$  одиночных (несопряженных) дуг, исходящих из вершины В сложного приращения.

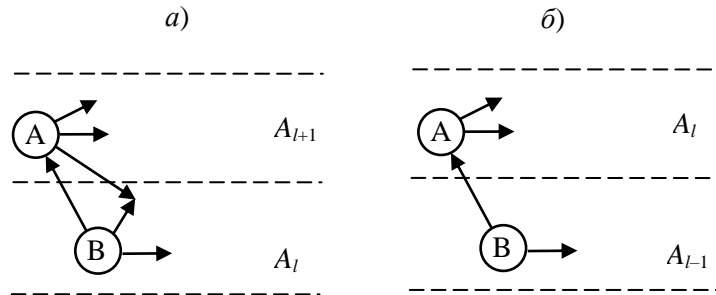


Рис. 5. Вершины сложного приращения попадают в соседние множества вершин: вершина А с вероятностью  $r_l''$  попадает в множество  $A_{l+1}$ , с вероятностью  $r_{l-1}''$  – в множество  $A_l$

Исходящие из вершины В дуги, сопряженные с дугами вершины А, – это в среднем  $P(l-1)$  дуг вершины В, с вероятностью  $P_{k-2}$  каждая связываются с вершинами множества  $A_{k-2}$ , переводят их в множество  $A_k$ , и сами становятся дугами множества  $B(l,k)$ , увеличивающегося при этом в среднем на  $P(l-1)P_{k-2}$  дуг. С учетом вероятности  $r_l''$  эта средняя добавка к  $B(l,k)$  составляет  $r_l''P(l-1) \frac{Q_{k-2}f_{k-2}}{\langle f \rangle}$  дуг.

Суммируя, находим, что при поступлении сложного приращения число дуг в множестве  $B(l,k)$  за счет добавления в него дуг ведомой вершины В возрастает в среднем на величину

$$r_l''(1-P)(l-1)\frac{Q_{k-1}f_{k-1}}{\langle f \rangle} + r_l''\delta_{l+1,k} + r_l''P(l-1)\frac{Q_{k-2}f_{k-2}}{\langle f \rangle}. \quad (17)$$

2.5.2. Если вершины А, В приращения имеют по  $l-1$  исходящих дуг (что случается с вероятностью  $r_{l-1}''$ , см. Рис. 5 б), то в множество  $B(l,k)$  могут добавляться дуги, исходящие из вершины А. Здесь исходящих из А сопряженных дуг в среднем  $P(l-2)$ , – столько же, как и у вершины В, а исходящих из А одиночных дуг ровно на одну больше, чем у вершины В, т.е. в среднем  $(1-P)(l-2) + 1 = l - Pl + 2P - 1$  дуг. Рассуждая далее по аналогии с п. 2.5.1, находим, что среднее число дуг вершины А сложного приращения, добавляемых в множество  $B(l,k)$ , составляет величину

$$r_{l-1}''(l - Pl + 2P - 1)\frac{Q_{k-1}f_{k-1}}{\langle f \rangle} + r_{l-1}''P(l-2)\frac{Q_{k-2}f_{k-2}}{\langle f \rangle}. \quad (18)$$

Итого средний прирост числа  $|B(l,k)|$ , получаемый при условии поступления сложного приращения, представляет сумму выражений (17) и (18).

3. Построим уравнение баланса вероятностей дуг в графе с НППС-СП. В стационарном режиме при добавлении произвольного приращения доля дуг, входящих в множество  $B(l,k)$  не изменяется:

$$\frac{|B(l,k)|}{mN} = \frac{|B(l,k)| + \Delta}{mN + (1-\gamma)m' + 2\gamma m''},$$

где  $m = [(1-\gamma)m' + 2\gamma m'']/(1+\gamma)$  – среднее число дуг, приходящихся на одну вершину графа,  $(1-\gamma)m' + 2\gamma m'' = m(\gamma+1)$  – среднее число дуг, добавляемых за один шаг выращивания графа,  $\Delta$  – среднее число дуг, добавляемых в множество  $B(l,k)$  за шаг выращивания графа.

Перепишем записанное уравнение баланса следующим образом:

$$\begin{aligned} |B(l,k)|mN + |B(l,k)|m(\gamma+1) &= |B(l,k)|mN + \Delta mN, \\ |B(l,k)|m(\gamma+1) &= \Delta mN, \\ \frac{|B(l,k)|}{mN}m(\gamma+1) &= \Delta, \\ Q_{l,k}m(\gamma+1) &= \Delta. \end{aligned} \quad (19)$$

Теперь нам остается выразить  $\Delta$  через другие параметры графа, чтобы найти  $Q_{l,k}$ .

С вероятностью  $(1-\gamma)$  на шаге выращивания графа поступает простое приращение, с вероятностью  $\gamma$  – сложное. Учитывая это, выразим в (19)  $\Delta$  через суммарные приращения числа дуг, получаемые при поступлении простого (см. п. 1) и сложного (см. п. 2) приращения графа. В результате получаем:



$$\begin{aligned}
Q_{l,k}m(\gamma+1) = & \\
& - (1-\gamma)mm' \frac{Q_{l,k}f_k}{\langle f \rangle} - (1-\gamma)mm' \frac{Q_{l,k}f_l}{\langle f \rangle} + (1-\gamma)mm' \frac{Q_{l,k-1}f_{k-1}}{\langle f \rangle} + \\
& + (1-\gamma)mm' \frac{Q_{l-1,k}f_{l-1}}{\langle f \rangle} + (1-\gamma)lr'_l \frac{Q_{k-1}f_{k-1}}{\langle f \rangle} - \\
& - \gamma m(2m'' - Pm'' + P - 1) \frac{Q_{l,k}f_k}{\langle f \rangle} - \gamma m(2m'' - Pm'' + P - 1) \frac{Q_{l,k}f_l}{\langle f \rangle} + \\
& + \gamma m(2m'' - 2Pm'' + 2P - 1) \frac{Q_{l,k-1}f_{k-1}}{\langle f \rangle} + \gamma m(2m'' - 2Pm'' + 2P - 1) \frac{Q_{l-1,k}f_{l-1}}{\langle f \rangle} + \\
& + \gamma mP(m'' - 1) \frac{Q_{l-2,k}f_{l-2}}{\langle f \rangle} + \gamma mP(m'' - 1) \frac{Q_{l,k-2}f_{k-2}}{\langle f \rangle} + \\
& + \gamma r_l''(1-P)(l-1) \frac{Q_{k-1}f_{k-1}}{\langle f \rangle} + \gamma r_l''\delta_{l+1,k} + \gamma r_l''P(l-1) \frac{Q_{k-2}f_{k-2}}{\langle f \rangle} + \\
& + \gamma r_{l-1}''(l-Pl+2P-1) \frac{Q_{k-1}f_{k-1}}{\langle f \rangle} + \gamma r_{l-1}''P(l-2) \frac{Q_{k-2}f_{k-2}}{\langle f \rangle}.
\end{aligned}$$

Перенося в этом уравнении члены с  $Q_{l,k}$  в левую часть, вынося  $Q_{l,k}$  за скобки и выражая эту вероятность в явном виде, получаем:

$$\begin{aligned}
Q_{l,k}\{m(\gamma+1)\langle f \rangle + (1-\gamma)mm'(f_k + f_l) + \gamma m(2m'' - Pm'' + P - 1)(f_k + f_l)\} = & \\
(1-\gamma)mm'Q_{l,k-1}f_{k-1} + (1-\gamma)mm'Q_{l-1,k}f_{l-1} + (1-\gamma)lr'_lQ_{k-1}f_{k-1} + & \\
+ \gamma m(2m'' - 2Pm'' + 2P - 1)Q_{l,k-1}f_{k-1} + \gamma m(2m'' - 2Pm'' + 2P - 1)Q_{l-1,k}f_{l-1} + & \\
+ \gamma mP(m'' - 1)Q_{l-2,k}f_{l-2} + \gamma mP(m'' - 1)Q_{l,k-2}f_{k-2} + & \\
+ \gamma r_l''(1-P)(l-1)Q_{k-1}f_{k-1} + \gamma r_l''\delta_{l+1,k}\langle f \rangle + \gamma r_l''P(l-1)Q_{k-2}f_{k-2} + & \\
+ \gamma r_{l-1}''(l-Pl+2P-1)Q_{k-1}f_{k-1} + \gamma r_{l-1}''P(l-2)Q_{k-2}f_{k-2}. &
\end{aligned}$$

Отсюда

$$Q_{l,k} = \frac{A + BQ_{l,k-1} + CQ_{l-1,k} + DQ_{l,k-2} + EQ_{l-2,k}}{(\gamma+1)\langle f \rangle + [(1-\gamma)m' + \gamma(2m'' - Pm'' + P - 1)](f_k + f_l)}, \quad (20)$$

где

$$\begin{aligned}
A = \{ & [(1-\gamma)lr'_l + \gamma r_l''(1-P)(l-1) + \gamma r_{l-1}''(l-Pl+2P-1)]Q_{k-1}f_{k-1} + \\
& + \gamma r_l''\delta_{l+1,k}\langle f \rangle + \gamma[r_l''(l-1) + r_{l-1}''(l-2)]PQ_{k-2}f_{k-2} \} / m,
\end{aligned}$$

$$B = [(1-\gamma)m' + \gamma(2m'' - 2Pm'' + 2P - 1)]f_{k-1},$$

$$C = [(1-\gamma)m' + \gamma(2m'' - 2Pm'' + 2P - 1)]f_{l-1},$$

$$D = \gamma P(m'' - 1)f_{k-2}, \quad E = \gamma P(m'' - 1)f_{l-2}.$$

Формула (20) выражает двумерную рекурсию (аналог дифференциального уравнения в частных производных), в соответствии с которой матрица  $\mathbf{Q} = \|\mathbf{Q}_{l,k}\|$  может рассчитываться построчно слева направо сверху вниз. Поскольку компоненты формулы (20), расположенные в числителе и содержащие ссылки на предыдущие элементы матрицы, содержат весовые множители  $f_{k-1}$  и  $f_{k-2}$ , обращающиеся в нуль на границах матрицы, значения  $\mathbf{Q}_{l,k}$  на соответствующих границах определены.

Если положим, что  $\{r'_k\} = \{r''_k\} = \{r_k\}$  и  $\gamma = 0$ , то из (20) получим формулу (4), выведенную для графа с НППС.

### 3.5. Калибровка графа с НППС-СП по РСС ребер

Задача калибровки графа по РСС ребер заключается в том, чтобы по заданному РСС  $\mathbf{Q}_{l,k}$  найти параметры генерации графа, при которых он будет иметь это заданное РСС ребер.

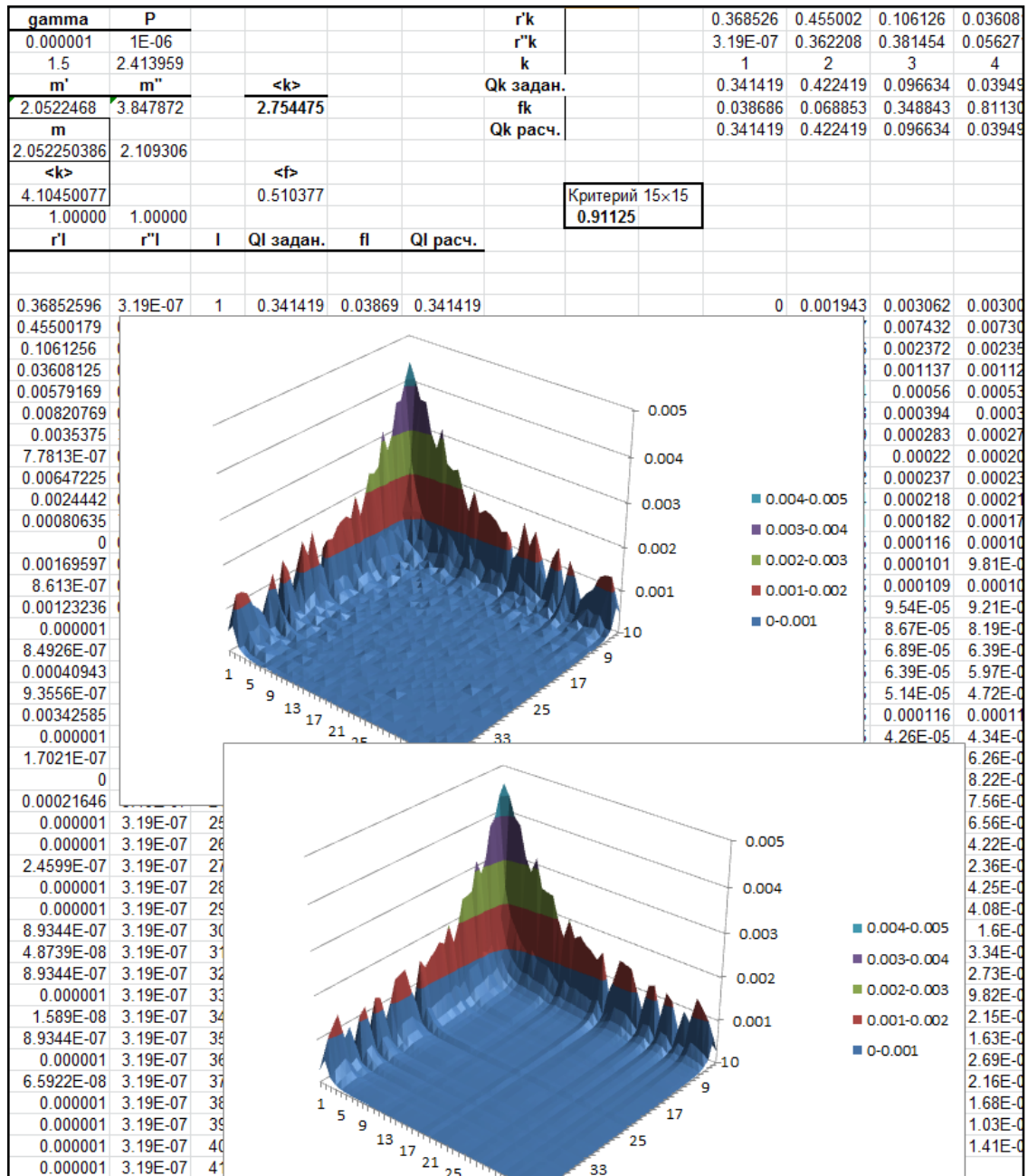
Эта задача сводится к оптимизационной задаче (5) и достаточно эффективно решается в среде электронных таблиц. Пример результатов калибровки графа по РСС ребер приведен на Рис. 6. В верхнем левом углу таблицы заданы параметры генерации, такие же, как в таблице на Рис. 4. Шесть столбцов слева являются копиями соответствующих столбцов этой же таблицы, индексируемых как степень  $l$ , а шесть строк сверху получены путем транспонирования этих столбцов и соответствуют индексу  $k$ . Указанные столбцы и строки представляют собой левый и, соответственно, верхний входы в таблицу (матрицу)  $\mathbf{Q} = \|\mathbf{Q}_{l,k}\|$  РСС дуг графа, рассчитываемую по формуле (20).

Ниже располагается матрица РСС ребер  $\mathbf{\Theta} = \frac{\mathbf{Q} + \mathbf{Q}^T}{2}$ , еще ниже – матрица РСС ребер моделируемой сети АС (на Рис. 6 эти матрицы не видны). С использованием сервиса «Поиск решения» отыскиваются значения варьируемых параметров  $\{r'_k\}$  и  $\{r''_k\}$ , доставляющие минимум целевой функции (5). При этом сохраняется точная калибровка графа по РСС вершин, т.е. РСС  $\mathbf{Q}_k$  не изменяется.

Диаграммы, построенные на листе электронных таблиц на Рис. 6, показывают, что при калибровке достигнуто достаточно хорошее соответствие между РСС ребер графа и РСС связей в моделируемой сети АС (верхняя диаграмма – это РСС связей в сети АС, нижняя – РСС ребер в выращенном графе).

Параметры  $\gamma$  и  $P$  определяются таким образом, чтобы обеспечивалось необходимое число треугольников, определяющих требуемый коэффициент кластеризации графа. В результате калибровка графа выполняется одновременно по трем структурным характеристикам – по РСС вершин, по РСС ребер и по коэффициенту кластеризации.

Следует заметить, что успешность калибровки сразу по трем характеристикам зависит от свойств моделируемой сети. При моделировании сети АС калибровка по РСС вершин не ухудшается при улучшении калибровок по РСС ребер и по коэффициенту кластеризации, однако качество калибровок по РСС ребер и по коэффициенту кластеризации находятся в противоречивом отношении: улучшение одной из них ухудшает другую.



#### 4. Анализ устойчивости сети интернет

Рассмотрим задачу устойчивости сети интернет к случайным отказам ее элементов. Как принято в Науке о сетях (Network Science), меру устойчивости определим через долю вершин, относящихся к компоненте связности наибольшего размера при заданной вероятности удалений вершин или ребер графа [9]. При этом объектом моделирования является сеть АС интернета, полученная М. Ньюманом в 2006 году на основе анализа BGP таблиц [5]. Граф этой сети содержит 22963 вершины и 48436 ребер, коэффициент кластеризации графа равен 0,011, будем называть этот натурный граф АС графом сети АС.

В качестве моделей сети АС рассмотрим следующие случайные графы.

– Граф БА [7] с распределением числа ребер в приращении  $\{r_2, r_3\}=\{0,9, 0,1\}$ , такое распределение обеспечивает соответствующее графу АС число вершин и ребер.

– Граф Уотса-Строгатца [27] со средним числом связей у вершины  $m = 2$  (что обеспечивает соответствующее графу АС число вершин и чуть большее число ребер). Также использовался параметр генерации графа Уотса-Строгатца  $p = 0,2$  (это значение параметра  $p$  обеспечивает требуемое значение коэффициента кластеризации графа).

– Граф с НППС-СП, который имеет следующие параметры генерации, полученные путем комплексной калибровки:

$P = 1, \gamma = 0,04096; \{r'_1, \dots, r'_4\} = \{0,3936, 0,4945, 0,1062, 0,00577\};$   
 $\{r''_1, \dots, r''_{30}\} = \{0,0773, 0, 0,2658, 0,3285, 0, 0,127, 0, 0, 0,0465, 0, 0, 0,0706, 0, 0, 0, 0, 0, 0,0055, 0, 0,0036, 0, 0, 0, 0, 0, 0,0063, 0, 0, 0, 0,0687\};$   
 $\{f_1, \dots, f_{31}\} = \{0,0737, 0,1416, 0,7014, 1,455, 2,546, 3,393, 4,198, 4,490, 5,122, 5,783, 6,047, 7,216, 8,516, 8,861, 9,306, 9,704, 10,07, 10,57, 11,06, 11,49, 11,92, 12,18, 12,44, 12,66, 12,86, 13,43, 13,98, 14,11, 14,27, 20,21, 26,2\};$   
 $f_k = 0,829353 \cdot k$  при  $k \geq 32$ .

Заметим, что граф БА и граф Уотса-Строгатца являются популярными моделями сетей, на каждую из статей [7, 27], в которых они предложены, имеется более 20 тысяч ссылок в системе цитирования Scopus. Также заметим, что возможности подбора параметров модели различные. В графе БА можно подобрать число вершин и ребер, при этом будет реализовано асимптотически-степенное РСС вершин, более соответствующее графу АС, чем РСС вершин графа Уотса-Строгатца (Рис. 7).

Но в графе Уотса-Строгатца возможно подобрать соответствующее число вершин и коэффициент кластеризации, приблизительно можно также реализовать заданное число ребер. В графе с НППС-СП возможно реализовать требуемые число вершин, число ребер, РСС вершин, РСС ребер и коэффициент кластеризации (Рис. 8).

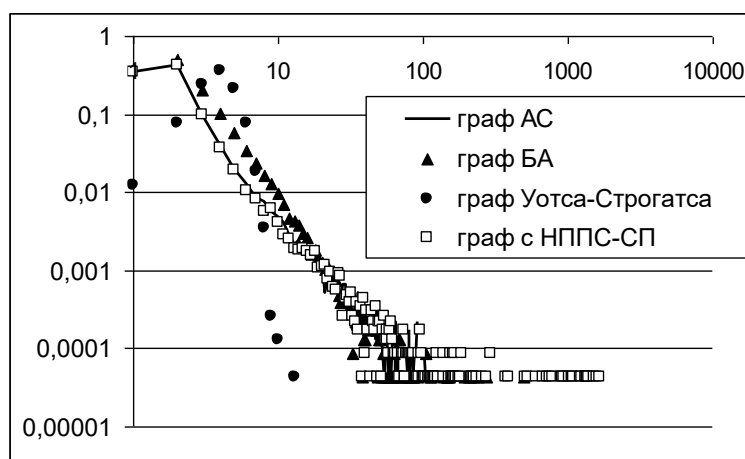


Рис. 7. Эмпирические РСС вершин исследуемых графов

На Рис. 9 и Рис. 10 представлены результаты моделирования устойчивости графов в условиях случайных удалений ребер и вершин: результаты получены и усреднены по 10 статистическим экспериментам.

Оценка числа вершин в максимальной компоненте связности, полученная при использовании графа с НППС, меньше всего отклоняется от результатов, полученных непосредственно на графе сети АС. Это означает, что граф с НППС-СП является более адекватной моделью сети АС.

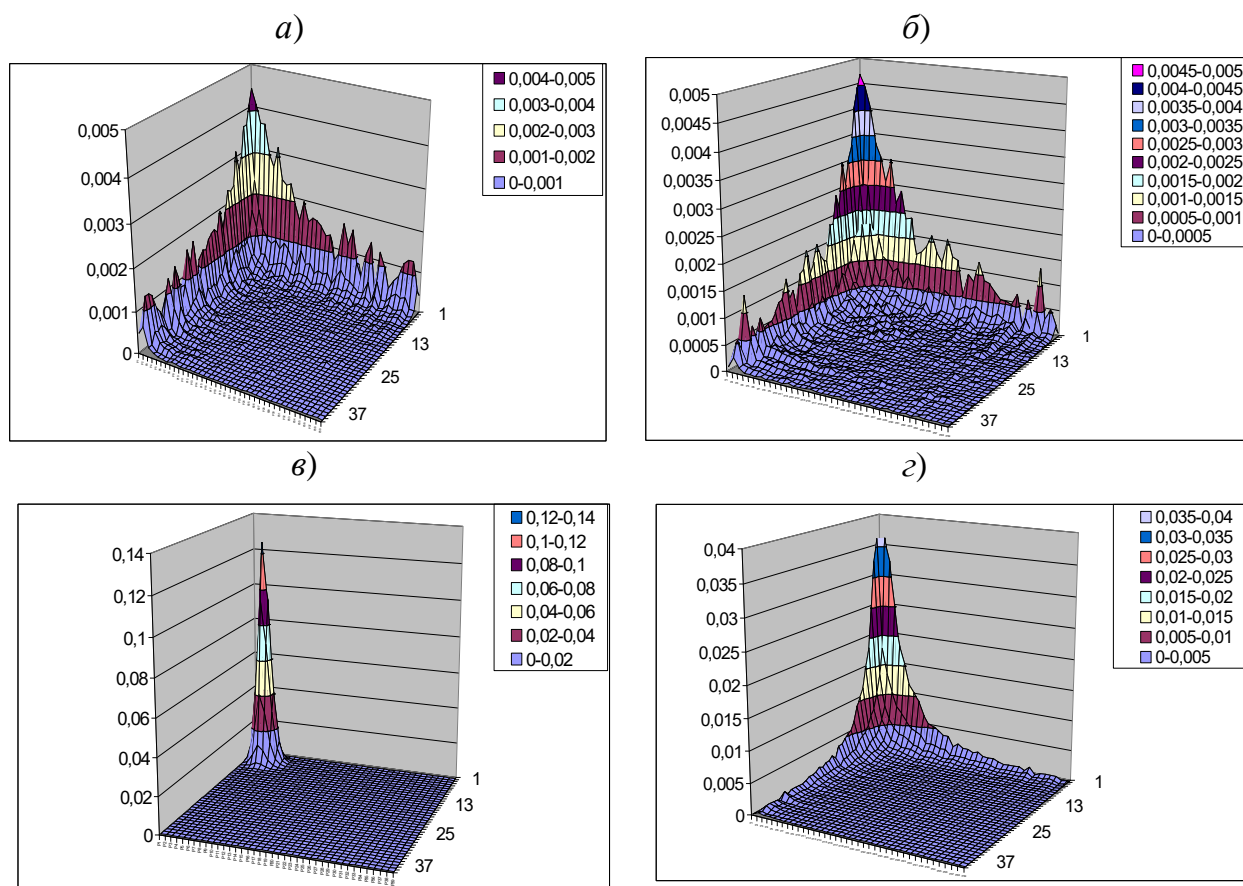


Рис. 8. РСС ребер: а) сети АС, б) графа с НППС-СП, в) графа Уотса-Строгатца, г) графа БА

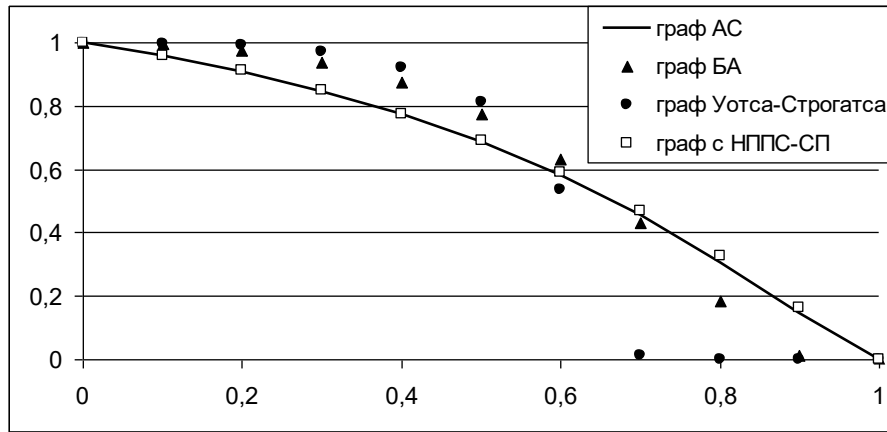


Рис. 9. Доля вершин графа, находящихся в компоненте связности максимального размера, как функция от вероятности удаления ребра графа

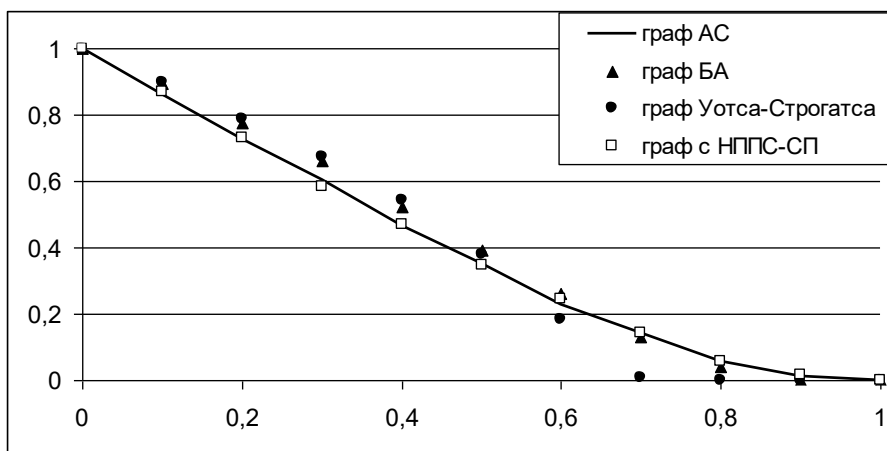


Рис. 10. Доля вершин графа, находящихся в связном компоненте максимального размера, как функция от вероятности удаления вершины графа

## 5. Заключение

1. В работе исследуется модификация графа с нелинейным правилом предпочтительного связывания со сложными стохастическими приращениями (графа с НППС-СП), впервые введенная в работе [4]. Особенностью графа с НППС-СП является наличие помимо приращений с одиночными вершинами так называемых сложных приращений, содержащих сразу две вершины. Это позволяет «внедрять» в граф полные подграфы на трех вершинах («треугольники») и тем самым увеличивать коэффициент кластеризации, решая известную проблему заниженного коэффициента кластеризации в графах предпочтительного связывания. Особенность предложенной модификации заключается в том, что число дуг/ребер в простом и сложном приращениях задаются разными распределениями вероятностей, определяя так называемый гетерогенный граф с НППС-СП. Это позволяет расширить возможности калибровки графа по трем заданным структурным характеристикам: по РСС вершин, РСС ребер и по коэффициенту кластеризации.

2. В работе выведены аналитические рекуррентные соотношения, описывающие РСС вершин, РСС ребер и (приближенно) коэффициент кластеризации гетерогенного графа с НППС-СП.

3. Примеры моделирования таких сетевых процессов как случайные потери узлов и связей в сети свидетельствуют об адекватности гетерогенного графа с НППС-СП как модели графа сети автономных систем интернета. Те же выводы можно сделать, исследуя другие сетевые процессы.

Достоинством разработанной модели является также то, что предложенный случайный граф позволяет моделировать рост конечной сети, и, следовательно, строить прогнозы развития сетевых процессов в ходе роста сетей до заданных размеров. Эта возможность имеет большое значение для решения различных актуальных задач формирования «сетевых сообществ», исследования социальных сетей, влияния на них, и для решения задач информационного противоборства [1]. Такие задачи, связанные с прогнозами развития сети и сетевых процессов, не могут эффективно решаться путем построения последовательности натуральных графов сети, получаемых в разные моменты времени. Такой путь позволял бы строить весьма ограниченные прогнозы, имеющие невысокую точность и требующие больших затрат времени и ресурсов на построение и анализ последовательности натуральных графов сети.

## ЛИТЕРАТУРА

1. Губанов Д.А., Новиков Д.А., Чхартушвили А.Г. Социальные сети: модели информационного влияния, управления и противоборства – М.: МЦНМО, ФИЗМАТЛИТ, 2010. – 228 с.
2. Задорожный В.Н. Случайные графы с нелинейным правилом предпочтительного связывания // Проблемы управления. – 2010. – № 6. – С. 2–11.
3. Задорожный В.Н., Юдин Е.Б. Структурные свойства безмасштабного графа Барабаши–Альберт // Автоматика и телемеханика. – 2012, – № 4. – С. 131–150.
4. Задорожный В.Н., Юдин Е.Б., Юдина М.Н. Калибровка графов предпочтительного связывания по распределениям степеней вершин и коэффициентам кластеризации // Омский научный вестник. – 2017. – Т. 153, №3. – С. 104–109.
5. Структура автономных систем сети Интернет, воссозданная на основе BGP таблиц, 2006 г. [Электронный ресурс] – Режим доступа: <http://www-personal.umich.edu/~mejn/netdata/as-22july06.zip>.
6. Albert R., Barabasi A. Statistical mechanics of complex networks // Reviews of Modern Physics. – 2002. – V. 74. – P. 47–97.
7. Barabasi A., Albert R. Emergence of scaling in random networks // Science. – 1999. – V. 286. – P. 509–512.
8. Barabasi A. Love is all you need: Clauset's fruitless search for scale-free networks [Электронный ресурс] – Режим доступа: <https://www.barabasilab.com/post/love-is-all-you-need>.
9. Barabasi A., Bonabeau E. Scale-Free Networks // Scientific American. – 2003. – V. 288. – P. 60–69.
10. Broido A.D., Clauset A. Scale-free networks are rare // Nature Communications. – 2019. – V. 10. – P. 1017.
11. Dan Yang, Liming Pan, Tao Zhou Lower bound of assortativity coefficient in scale-free networks // Chaos. – 2017. – V. 27. – P. 033113.

12. Dorogovtsev S.N., Mendes J.F.F., Samukhin A.N. Structure of Growing Networks: Exact Solution of the Barabasi-Albert's Model // Physical Review Letters. – 2000. – V. 85. – P. 4633–4636.
13. De Brito J.B., Sampaio Filho C.I.N., Moreira A.A., Andrade J. S. Characterizing the intrinsic correlations of scale-free networks // International Journal of Modern Physics. – 2016. – V. 27, No. 03. – P. 1650024.
14. Csardi G. Generate scale-free graphs according to the Barabasi-Albert model [Электронный ресурс] – Режим доступа: [https://igraph.org/r/doc/sample\\_pa.html](https://igraph.org/r/doc/sample_pa.html).
15. Iskhakov L., Kaminski B., Mironov M., Pralat P., Prokhorenkova L. Clustering Properties of Spatial Preferential Attachment Model // Lecture Notes in Computer Science. – 2018. – V. 11825. – P.30–43.
16. Jackson M.O., Rogers B.W. Meeting strangers and friends of friends: how random are socially generated networks? // American Economic Review. – 2007. – V. 97. – P. 890–915.
17. Lee C., Wilkinson D.J. A Review of Stochastic Block Models and Extensions for Graph Clustering [Электронный ресурс] – Режим доступа: <https://arxiv.org/abs/1905.05738>.
18. Krapivsky P.L., Redner S. Organization of growing random networks // Physical Review E. – 2001. – V. 63, No. 6. – P. 066123-066137.
19. Krapivsky P. L., Redner S. Network growth by copying // Physical Review E. – 71 (3). – 036118.
20. Makarov V.V. et al. Assortative mixing in spatially-extended networks // Scientific Reports. – 2018. – V. 8. – P.13825.
21. Newman M.E.J. Random graphs with clustering // Physical Review Letters. – 2009. – 103. – P. 058701.
22. Pennock D.M. et al. Winners don't take all: Characterizing the competition for links on the web // Proceedings of the National Academy of Science of the United States of America. PNAS. – 2002. – V. 99, No. 8. – P. 5207–5211.
23. Haruna T., Gunji Y.P. Ordinal Preferential Attachment: A Self-Organizing Principle Generating Dense Scale-Free Networks // Scientific Reports. – 2019. – V. 9. – P. 4130.
24. Timar G., Dorogovtsev S.N., Mendes J.F.F. Scale-free networks with exponent one // Physical Review E. – 2016. – V. 94. – P. 022302.
25. Tishby I., Biham O., Katzav E., Kühn R. Generating random networks that consist of a single connected component with a given degree distribution // Physical Review E. – 2019. – 99. – P. 042308.
26. Van Der Pol J. Introduction to Network Modeling Using Exponential Random Graph Models (ERGM): Theory and an Application Using R-Project // Computational Economics. – 2018. – doi:10.1007/s10614-018-9853-2.
27. Watts D.J., Strogatz S. H. Collective dynamics of 'small-world' networks // Nature. – 1998. – V. 393. – P. 440–442.
28. Weaver I.S. Preferential attachment in randomly grown networks // Physica A: Statistical Mechanics and its Application. – 2015. – V. 439. – P. 85–92.
29. Zadorozhnyi V.N., Yudin E.B. Growing network: models following nonlinear preferential attachment rule // Physica A. Statistical Mechanics and its Applications. – 2015. – V. 428. – P. 111–132.
30. Zadorozhnyi V.N., Yudin E.B. Vertex degree distribution and arc endpoints degree distribution of graphs with a linear rule of preferential attachment and Pennock graphs, 2019. [Электронный ресурс] – Режим доступа: <https://arxiv.org/abs/1904.00426>.
31. Zadorozhnyi V.N., Yudin E.B., Yudina M.N. Analytical and numerical methods of calibration for preferential attachment random graphs // International Siberian Conference on Control and Communications. – Kazakhstan, Astana, 2017. – P. 1–6.

---

**Задорожный** Владимир Николаевич, д.т.н., профессор, каф. «Математические методы и информационные технологии в экономике», [zwn2015@yandex.ru](mailto:zwn2015@yandex.ru)

**Юдин** Евгений Борисович, к.т.н., доцент, каф. «Математические методы и информационные технологии в экономике», [yudinev@gmail.com](mailto:yudinev@gmail.com)

**Юдина** Мария Николаевна, ассистент, каф. «Математические методы и информационные технологии в экономике», [mg-and-all@mail.ru](mailto:mg-and-all@mail.ru)



УДК 519.677:532.542.2

## ГИДРАВЛИЧЕСКАЯ МОДЕЛЬ НЕСТАЦИОНАРНОГО ЛАМИНАРНОГО ТЕЧЕНИЯ ПО ТРУБОПРОВОДУ

В. А. Наумов, А. А. Землянов

Калининградский государственный технический университет, Калининград, Россия

**Аннотация.** Рассмотрена одномерная математическая модель важного для инженерных приложений нестационарного ламинарного течения по трубопроводу. Получены дифференциальные уравнения движения и баланса жидкости в безразмерной форме. При постоянном уровне жидкости в баке задача имеет два критерия подобия, ее решение найдено аналитически. При переменном уровне задача имеет три критерия подобия, решение найдено численным методом. Исследовано влияние значений критериев подобия на изменение чисел Рейнольдса.

**Ключевые слова:** нестационарное ламинарное течение, дифференциальные уравнения, критерии подобия.

## HYDRAULIC MODEL OF UNSTEADY LAMINAR FLOW THROUGH A PIPELINE

V.A. Naumov, A.A. Zemlyanov

Kaliningrad State Technical University, Kaliningrad, Russia

**Abstract.** The one-dimensional mathematical model of non-stationary laminar flow through a pipeline, which is important for engineering applications, was considered. Differential equations of fluid motion and balance were obtained in dimensionless form. The problem has two similarity criteria at a constant level of liquid in the tank, and its solution is found analytically. The problem has three criteria of similarity with a variable fluid level, the solution is found numerically. The influence of similarity criteria values on changes in Reynolds numbers has been investigated.

**Keywords:** unsteady laminar flow, differential equations, similarity criteria.

### 1. Введение

Изучению гидравлического удара как нестационарного течения жидкости в трубопроводе посвящено большое количество теоретических и экспериментальных исследований. В инженерных приложениях нередко требуется решать задачи нестационарного движения жидкости в трубопроводе при сравнительно медленном изменении параметров течения (например, перепада давления). В таких случаях используют одномерную модель – уравнение Бернулли для нестационарных течений. Решение нестационарных гидравлических задач для турбулентного течения в трубопроводе рассмотрено в [1-4] и других публикациях.

Гораздо меньше работ опубликовано с решением нестационарных гидравлических задач при ламинарном течении. В [5] с помощью лабораторного и вычислительного эксперимента показано, что при ламинарном нестационарном течении коэффициент гидравлических потерь по длине трубы  $\lambda$  можно рассчитывать по формуле, выведенной для стационарных условий:

$$\lambda = 64 / Re, \quad Re = W \cdot d / \nu,$$

где  $Re$  – число Рейнольдса,  $W$  – скорость движения жидкости (средняя в поперечном сечении),  $d$  – внутренний диаметр трубы,  $\nu$  – коэффициент кинематической вязкости жидкости.

В [6] рассмотрен нестационарный процесс истечения из бака через трубопровод после мгновенного открытия затвора при ламинарном режиме. Пренебрегая потерями в местных гидравлических сопротивлениях и скоростным напором на выходе, автору [6] удалось найти аналитическое решение задачи. Цель данной статьи – решение нестационарной задачи без указанных ограничений и проверка применимости аналитического решения для задач течения высоковязких жидкостей.

## 2. Постановка задачи

Рассмотрим ламинарное течение из открытого бака в атмосферу через горизонтальный трубопровод постоянного диаметра (рис. 1). Сначала задвижка закрыта, жидкость неподвижна. После открытия задвижки жидкость начинает движение, скорость жидкости растет, процесс будет нестационарным даже при поддержании постоянного уровня воды в баке  $H$ .

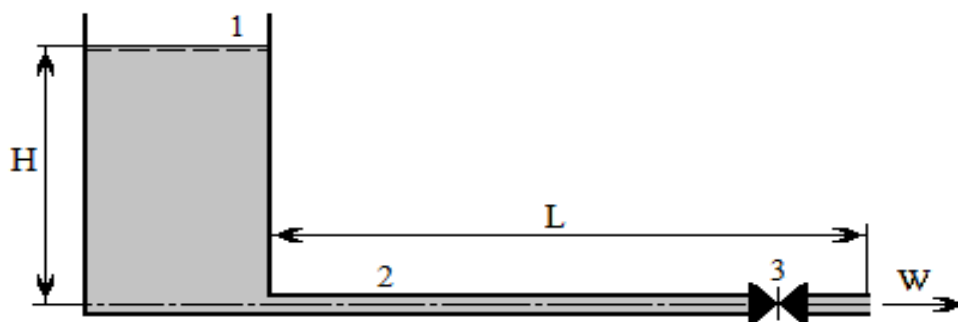


Рис. 1. Схема течения: 1 – бак с жидкостью, 2 – трубопровод, 3 – задвижка

Воспользуемся уравнением Бернулли для нестационарного течения жидкости в трубе (см., например, [1-4]):

$$H(t) = \frac{W^2}{2g} (\alpha_1 + \zeta) + \frac{\alpha_2}{g} \int_0^L \frac{\partial W}{\partial t} dX, \quad (1)$$

где последнее слагаемое обусловлено нестационарным характером течения;  $X$  – координата, отсчитываемая вдоль оси трубы, м;  $g$  – ускорение свободного падения, м/с<sup>2</sup>;  $L$  – длина трубы, м;  $t$  – время, с;  $\alpha_1, \alpha_2$  – коэффициенты неравномерности профиля продольной скорости жидкости;  $\zeta$  – обобщенный коэффициент гидравлического сопротивления:

$$\zeta = \lambda \cdot L / d + \Sigma \zeta_M,$$

где  $\zeta_M$  – коэффициенты местных гидравлических потерь.

В гидравлической системе на рис. 1 следует учитывать местные потери на вход из бака в трубопровод (коэффициент потерь  $\zeta_1$ ) и задвижку ( $\zeta_2$ ). Коэффициенты местных гидравлических потерь при ламинарном режиме течения рассчитываем по формуле, приведенной в [7]:

$$\zeta_M = B_M + A_M / Re. \quad (2)$$

Коэффициенты в (2) принимаем, как в [7]:  $A_1 = 30$ ;  $B_1 = 0,5$ ;  $A_2 = 150$ ;  $B_2 = 0,15$ .

Учитывая медленное изменение условий течения, можно считать скорость жидкости постоянной по длине трубопровода. Тогда из (1) с учетом (2) следует

$$\frac{\alpha_2 \cdot L}{g} \cdot \frac{dW}{dt} = H(t) - \frac{v \cdot W}{2g \cdot d} \cdot \left( \frac{64L}{d} + A \right) - \frac{W^2}{2g} \cdot (\alpha_1 + B). \quad (3)$$

В рассматриваемом случае  $\alpha_1 = 2$ ;  $A = A_1 + A_2 = 180$ ;  $B = B_1 + B_2 = 0,65$ .

При изменении уровня потребуется уравнение баланса жидкости:

$$S_1 \cdot \frac{dH}{dt} = -W(t) \cdot S_0, \quad S_0 = \frac{\pi \cdot d^2}{4}, \quad (4)$$

где  $S_0$  – площадь поперечного сечения трубы,  $S_1$  – площадь поперечного сечения бака. Чтобы не загромождать изложение, полагаем  $S_1 = \text{const}$ .

Начальные условия к дифференциальным уравнениям (3), (4):

$$W(0) = 0, \quad H(0) = H_0. \quad (5)$$

### 3. Решение задачи

#### 3.1. Установившаяся скорость при постоянном уровне

Пусть можно считать  $H_0 = \text{const}$ . Не обязательно, чтобы физически уровень не менялся. Достаточно выполнения условия  $S_0 \cdot L \ll S_1 \cdot H_0$ . Тогда из (3) следует квадратное уравнение для установившейся скорости  $W_1$ :

$$(\alpha_1 + B) \cdot W_1^2 + \frac{v \cdot W_1}{d} \cdot \left( \frac{64L}{d} + A \right) - 2g \cdot H_0 = 0. \quad (6)$$

Уравнение (6) имеет положительный корень:

$$W_1 = \frac{\sqrt{(\beta \cdot v/d)^2 + 2g \cdot H_0(\alpha_1 + B) - \beta \cdot v/d}}{\alpha_1 + B}, \quad \beta = 32 \cdot \delta + \frac{A}{2}, \quad \delta = \frac{L}{d}. \quad (7)$$

По (7) получим число Рейнольдса в установившемся режиме:

$$Re_1 \equiv \frac{W_1 d}{v} = \frac{\sqrt{\beta^2 + 2\Theta \cdot (\alpha_1 + B) - \beta}}{\alpha_1 + B}, \quad \Theta = \frac{g \cdot H_0 \cdot d^2}{v^2}. \quad (8)$$

Автор [6] полагает  $\alpha_1 + B = 0$ ,  $A = 0$ . Тогда установившаяся скорость и число Рейнольдса будут равны:

$$W_1^o = \frac{g \cdot H_0 \cdot d^2}{32v \cdot L}, \quad Re_1^o = \frac{\Theta}{32 \cdot \delta}. \quad (9)$$

Для сравнения введем их отношение:

$$E(\Theta, \delta) = W_1^o / W_1 = Re_1^o / Re_1. \quad (10)$$

На рис. 2-3 представлены результаты расчетов по формулам (8)-(10). Видно, что различие результатов расчета чисел Рейнольдса в установившемся режиме течения падает с уменьшением величины безразмерного комплекса  $\Theta$  и увеличением относительной длины трубопровода  $\delta$ . Полученной в [6] приближенной формулой для установившейся скорости (9) можно пользоваться при  $\delta > 400$ . Небольшие величины  $\delta$  требуют проверки зависимости скорости от  $\Theta$ . При неблагоприятных условиях расчет по формуле (9) может зависеть скорость более чем в два раза (см. рис. 3).

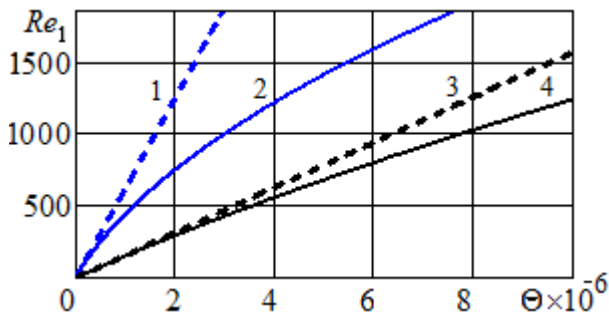


Рис. 2. Расчет чисел Рейнольдса в установившемся режиме: 1, 3 – по (9); 2, 4 – по (8); 1, 2 –  $\delta = 50$ , 3, 4 –  $\delta = 200$

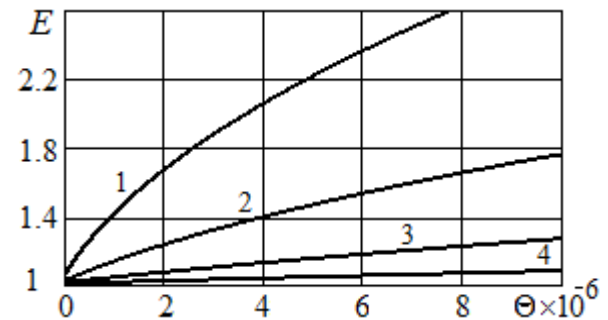


Рис. 3. Расчет отношения чисел Рейнольдса по (10): 1 –  $\delta = 50$ ; 2 –  $\delta = 100$ ; 3 –  $\delta = 200$ , 4 –  $\delta = 400$

### 3.2. Задача Коши при постоянном уровне

Умножим обе части уравнения (3) на  $g \cdot (d/v)^2$  и введем безразмерное время

$$\tau = t \cdot v / (L \cdot \alpha_2 \cdot d). \quad (11)$$

Тогда из (3), получим дифференциальное уравнение для числа Рейнольдса:

$$\frac{d Re}{d \tau} = \Theta - \left( 32 \cdot \delta + \frac{A}{2} \right) \cdot Re - \frac{1}{2} (\alpha_1 + B) \cdot Re^2, \quad Re(0) = 0. \quad (12)$$

По (12) видно, что критериями подобия задачи являются  $\Theta$  и  $\delta$ . Разумеется, при заданной гидравлической схеме и величинах  $A, B$ . Задача Коши (12) имеет аналитическое решение. На рис. 4 представлено изменение числа Рейнольдса в процессе истечения при различных значениях критериев подобия.

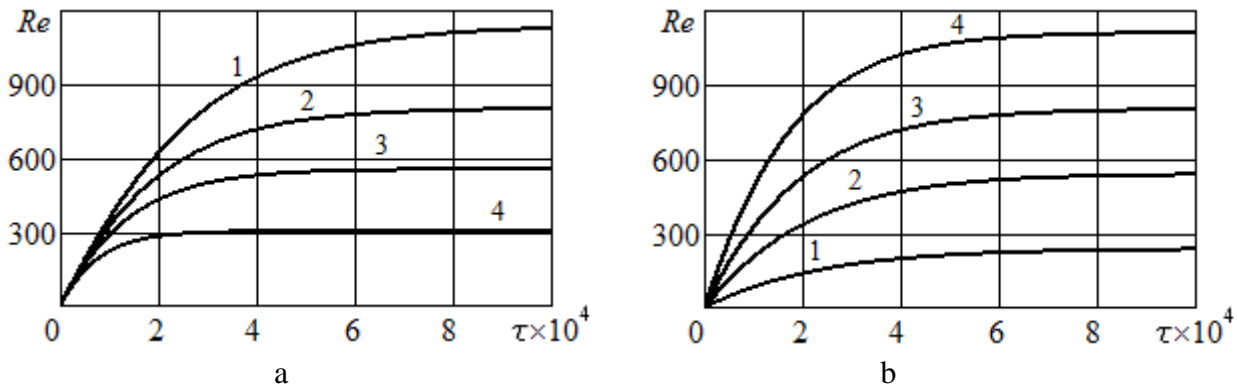


Рис. 4. Изменение чисел Рейнольдса при различных значениях критериев подобия.

а –  $\Theta = 4 \cdot 10^6$ : 1 –  $\delta = 60$ ; 2 –  $\delta = 120$ ; 3 –  $\delta = 200$ ; 4 –  $\delta = 400$ ;  
 б –  $\delta = 1$  –  $\Theta = 10^6$ ; 2 –  $\Theta = 2,5 \cdot 10^6$ ; 3 –  $\Theta = 4 \cdot 10^6$ ; 4 –  $\Theta = 6 \cdot 10^6$

### 3.3. Задача Коши при переменном уровне

Умножим обе части (4) на  $(d/v)$ , время выразим из (11); получим дифференциальное уравнение баланса жидкости в безразмерной форме:

$$\frac{d h}{d \tau} = -Re(\tau) \cdot \Psi, \quad h = \frac{H}{H_0}, \quad h(0) = 1, \quad \Psi = \frac{\alpha_2 S_0 L}{S_1 \cdot H_0}, \quad (13)$$

где  $\Psi$  – третий критерий подобия, который характеризует отношение объемов жидкости в трубопроводе и в баке.

Уравнение (12) в задаче с переменным уровнем должно быть записано так

$$\frac{d Re}{d \tau} = \Theta \cdot h(\tau) - \left( 32 \cdot \delta + \frac{A}{2} \right) \cdot Re - \frac{1}{2} (\alpha_1 + B) \cdot Re^2, \quad Re(0) = 0. \quad (14)$$

Задача Коши (13)-(14) решена численным методом в среде Mathcad [8].

## 4. Заключение

Поставлена задача Коши для нестационарного истечения жидкости из бака через трубопровод в атмосферу при ламинарном режиме. При переменном уровне жидкости в баке задача имеет три критерия подобия  $\delta, \Theta, \Psi$ ; при постоянном – два критерия. Полученной ранее приближенной формулой для установившейся скорости можно пользоваться при  $\delta > 400$ , независимо от  $\Theta$ . В неблагоприятных условиях расчет по такой формуле может зависеть скорость более

чем в два раза. Исследовано влияние значений критериев подобия на изменение чисел Рейнольдса в процессе истечения. Увеличение  $\delta$  приводит к снижению  $Re$ , возрастание  $\Theta$  – к росту чисел Рейнольдса.

## ЛИТЕРАТУРА

1. *Fox J.A.* Hydraulic analysis of unsteady flow in pipe networks. London: Macmillian Press Ltd. 1977. 234 p.
2. *Pejovic S., Boldy A.P.* Guidelines to hydraulic transient analysis of pumping systems. Belgrade; Coventry. P & B Press. 1992. 181 p.
3. *Наумов В.А.* Механика движения неоднородных сред: учебник. Калининград: Изд-во КГТУ, 2005. 125 с.
4. *Бабаев А.М.* Гидравлика: учебное пособие. Москва: Изд-во ЭКСМО, 2008. 192 с.
5. *Турченко М.В., Андреев В.В., Дунцев А.В.* Моделирование коэффициента гидравлического трения при неустановившемся течении жидкости с помощью искусственных нейронных сетей // Труды Нижегородского гос. техн. университета. 2011. № 3(90). С. 127-134.
6. *Зезин В.Г.* Нестационарные процессы гидродинамики: учебное пособие. Челябинск: Изд-во Южно-Уральского государственного университета, 2011. 74 с.
7. *Альтишль А.Д., Калицун В.И., Майрановский Ф.Г., Пальгунов П.П.* Примеры расчетов по гидравлике: учебное пособие. Москва: Стройиздат, 1977. 255 с.
8. *Наумов В.А.* Прикладная математика: учебное пособие по решению профессиональных задач в среде Mathcad. Калининград: Изд-во ФГОУ ВПО «КГТУ», 2014. 144 с.

---

**Наумов** Владимир Аркадьевич, д.т.н., профессор, каф. «Водные ресурсы и водопользование» van-old@mail.ru

**Землянов** Антон Алексеевич, студент, zemlyanov99@gmail.com

УДК 004.912+004.738.5

## СОЗДАНИЕ ПРОГРАММЫ ДЛЯ ПОСТРОЕНИЯ ГИСТОГРАММЫ НА ОСНОВЕ ГРАФЕМАТИЧЕСКОГО АНАЛИЗА ДАННЫХ ИЗ ОТКРЫТОГО ИНТЕРНЕТ-ИСТОЧНИКА

*А. Н. Абдрахманов*

Омский государственный университет путей сообщения, Омск, Россия

**Аннотация.** В ходе исследования создается программа для графематического анализа текстовой информации. Она должна принимать на входе текстовые данные, разбивать их на составные части, выполнять вычисления на их основе и строить гистограмму. Для этих целей выбран язык программирования Python, который включает в себя набор необходимых библиотек. В качестве эксперимента собраны данные российских физических лиц, входящих в перечень действующих террористов и экстремистов на официальном сайте Федеральной службы по финансовому мониторингу.

**Ключевые слова:** графематический анализ, Python, чанкинг.

## CREATING A PROGRAM FOR BUILDING A HISTOGRAM BASED ON GRAPHEMATIC ANALYSIS OF DATA FROM AN OPEN INTERNET SOURCE

*A. N. Abdrakhmanov*

Omsk State Transport University, Omsk, Russia

**Abstract.** During the research, we will create a program for graphematic analysis of text information. It must accept text data as input, parse it into its component parts, perform calculations based on it, and build a histogram. For this purpose, the Python programming language was chosen, which includes a set of necessary libraries. As an experiment, we collected data from Russian individuals included in the list of active terrorists and extremists on the official website of the Federal financial monitoring service.

**Keyword:** graphemic analysis, Python, chunking.

### 1. Введение

Большинство публикаций о машинном обучении и обработке естественного языка носят исследовательский характер и не подходят для разработки приложений. Например, несмотря на наличие большого количества превосходных инструментов для машинного обучения на текстовых данных, имеющиеся источники информации – документация, руководства и статьи в сети интернет, как правило, опираются на искусственно подобранные массивы данных, исследовательские инструменты и экспериментальный код.

Сложность исследований заключается в том, что язык – это неструктурированные данные, которые используются людьми для общения между собой. Но, несмотря на отсутствие машиночитаемой структуры, неструктурированные данные не являются случайными. Напротив, они подчиняются лингвистическим правилам, которые делают эти данные понятными для людей. Текст является одной из самых распространенных форм представления данных.

## 2. Постановка задачи

Большинство современных программ по обработке естественного языка относятся к модульному типу, в котором каждому этапу анализа или синтеза текста соответствует отдельный модуль программы. В случае анализа текста отдельные модули выполняют:

- графематический анализ (сегментацию), то есть выделение в тексте предложений и словоформ, точнее токенов, так как в тексте могут быть не только слова, но и переходы от символов к словам;
- морфологический анализ – переход от словоформ к их леммам (словарным формам лексем) или основам (ядерным частям слова, за вычетом словоизменятельных морфем);
- синтаксический анализ – выявление синтаксических связей слов и грамматической структуры предложений;
- семантический и прагматический анализ, при котором определяется смысл фраз и соответствующая реакция системы, в рамках которой работает программа [1].

В рамках данной статьи требуется создать программу для графематического анализа текстовой информации. Программа принимает на входе текстовые данные, разбирает их на составные части, выполняет вычисления на основе этих частей и вновь собирает их, возвращая осмысленный и адаптированный результат.

Поскольку методы анализа текста применяются в первую очередь в машинном обучении, необходим язык программирования с богатым набором научных и вычислительных библиотек. Именно по этой причине был выбран язык Python. В качестве эксперимента собираются данные российских физических лиц, входящих в перечень действующих террористов и экстремистов на официальном сайте Федеральной службы по финансовому мониторингу [2]. Гистограмма строится с группировкой лиц по возрасту для определения среднего возраста действующих террористов и экстремистов.

## 3. Результаты экспериментов

Прежде чем собрать данные, необходимо определить структуры страницы, чтобы было понятно, в каких элементах хранится необходимая нам информация. В нашем случае используются элементы `li` в классе «`russianFL`». Обычно текстовые данные приходится разбивать на отдельные информационные блоки (чанки) для последующего анализа. Этот процесс называется чанкингом и часто используется при анализе текста. Условия разбиения текста на блоки могут меняться в зависимости от конкретной задачи. Чанкинг не следует смешивать с токенизацией, которая также связана с разбиением текста на отдельные части. В процессе чанкинга не придерживаются никаких ограничений, а результирующие блоки обязаны иметь смысл [3].

При работе с большими документами важно разбивать текст на блоки для извлечения смысловой информации. В нашем случае каждую строку разбиваем на 5 блоков:



- 1) фамилия, имя, отчество;
- 2) фамилия, имя, отчество в случае смены;
- 3) дата рождения;
- 4) возраст;
- 5) место регистрации.

Для работы парсера необходимо использовать библиотеки Requests, BeautifulSoup, Csv, Datetime, Matplotlib, Pandas, NumPy. Для установки библиотек в командной строке вводим `pip install` и название библиотеки.

Библиотека Requests является стандартным инструментом для составления HTTP-запросов в Python [4].

BeautifulSoup – это библиотека Python для извлечения данных из файлов HTML и XML. Часто используется для скрапинга веб-страниц. BeautifulSoup позволяет трансформировать сложный HTML-документ в сложное дерево различных объектов Python. Это могут быть теги, навигация или комментарии [5].

Модуль CSV дает возможность реализовать чтение и запись табличных данных в формате CSV. Данный формат является наиболее распространенным форматом импорта и экспорта для электронных таблиц и баз данных [6].

Модуль Datetime предоставляет классы для работы с датами и временем. Поддерживается и стандартный способ представления времени, но модуль позволяет эффективно извлекать атрибуты для форматирования и обработки выходных данных [7].

Matplotlib – это обширная библиотека для создания статических, анимированных и интерактивных визуализаций в Python. Получаемые изображения могут быть использованы в качестве иллюстраций в публикациях [8].

Pandas является библиотекой с открытым исходным кодом. Предоставляет высокопроизводительные, простые в использовании структуры данных и инструменты анализа данных для языка программирования Python [9].

NumPy – это библиотека языка Python, добавляющая поддержку больших многомерных массивов и матриц вместе с большой библиотекой высокоуровневых математических функций для операций с этими массивами [10].

Пример собранных данных представлен на Рис. 1, в качестве разделителя используем запятую.

А1														fx АБАДИЕВ МАГОМЕД МИКАИЛОВИЧ,,09.11.1982,37,Г. НАЗРАНЬ ЧИАССР													
	A	B	C	D	E	F	G	H	I	J	K	L	M														
1	АБАДИЕВ МАГОМЕД МИКАИЛОВИЧ,,09.11.1982,37,Г. НАЗРАНЬ ЧИАССР																										
2	АБАЕВ АРТУР АБУАЛИЕВИЧ,КИСКАЕВ АРТУР АБУАЛИЕВИЧ,03.08.1983,36,Г. ХАСАВЮРТ ДАГЕСТАНСКОЙ ССР																										
3	АБАЕВ АСЛАНБЕК БАУДИНОВИЧ,,29.05.1983,36,Г. ГРОЗНЫЙ ЧИАССР																										
4	АБАЕВ ИДРИС МОВСАРОВИЧ,,02.01.1986,34,С. СУВОРОВ-ЮРТ ГУДЕРМЕССКОГО РАЙОНА ЧИАССР																										
5	АБАЕВ МУХАДИН ЮРЬЕВИЧ,,30.09.1981,38,Г. НАЛЬЧИК КАБАРДИНО-БАЛКАРСКОЙ РЕСПУБЛИКИ																										
6	АБАЗОВ АЗАМАТ МУАЕДОВИЧ,,06.09.1985,34,С. ЗАЮКОВО БАКСАНСКОГО РАЙОНА КБР																										
7	АБАЗОВ АХМЕД ЭДУАРДОВИЧ,,30.08.1992,27,С. КЫЗБУРУН-3 БАКСАНСКОГО РАЙОНА КАБАРДИНО-БАЛКАРСКОЙ РЕСПУБЛИКИ																										
8	АБАЗОВ БЕСЛАН СУПЬЯНОВИЧ,,15.08.1993,26,С. ШАЛИ ШАЛИНСКОГО РАЙОНА ЧЕЧЕНСКОЙ РЕСПУБЛИКИ																										
9	АБАЗОВ МУРАТ АРСЕНОВИЧ,,10.05.1989,30,Г. НАЛЬЧИК КАБАРДИНО-БАЛКАРСКОЙ РЕСПУБЛИКИ																										
10	АБАЗОВ МУРАТ ВАЛЕРЬЕВИЧ,,20.12.1982,37,Г. НАЛЬЧИК КАБАРДИНО-БАЛКАРСКОЙ РЕСПУБЛИКИ																										

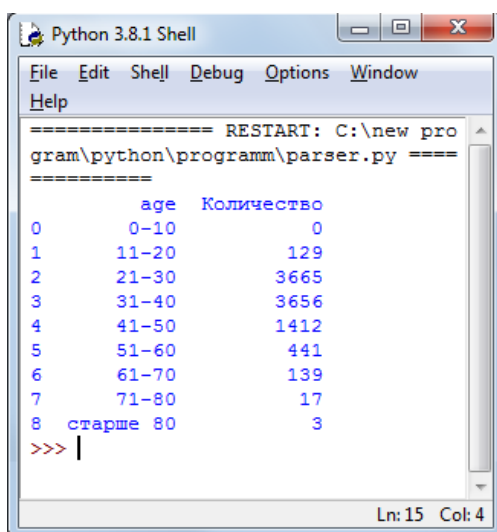
Рис. 1. Данные, собранные в результате работы приложения

Для анализа строки используем признаки, характерные для данного набора данных. Например, фамилия, имя и отчество отделяются либо знаком «\*», либо «,», фамилия, имя и отчество в случае смены указаны в скобках, после даты рождения, если они имеются, стоят символы «г.р.». На основе именно этих признаков разделяем текст на блоки. Для расчета возраста по дате рождения будем использовать условные операторы.

## 5. Обсуждение результатов

После структуризации данных в виде CSV-файла переходим к построению гистограммы. Для удобства используются следующие возрастные диапазоны: 0-10, 11-20, 21-30, 31-40, 41-50, 51-60, 61-70, 71-80 и старше 80 лет.

Данные, необходимые для построения гистограммы, представлены на Рис. 2.



	age	Количество
0	0-10	0
1	11-20	129
2	21-30	3665
3	31-40	3656
4	41-50	1412
5	51-60	441
6	61-70	139
7	71-80	17
8	старше 80	3

Рис. 2. Исходные данные для построения гистограммы

Полученная гистограмма представлена на Рис. 3.

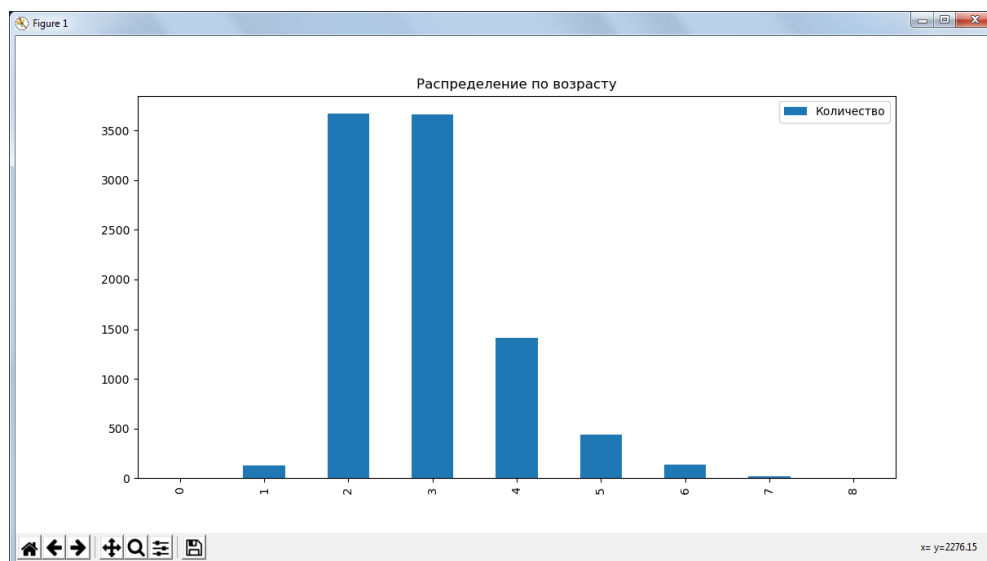


Рис. 3. Гистограмма «Распределение лиц по возрасту»

## 6. Выводы и заключение

В ходе научных исследований разработано приложение, собирающее данные из открытого интернет-источника с последующими их анализом и визуализацией.

На основе полученных данных видим, что большая часть террористов и экстремистов – это лица в возрасте от 20 до 40 лет. Данная информация может использоваться для поиска таких лиц в социальных сетях и позволяет ограничивать круг поиска, учитывая возрастные характеристики.

## ЛИТЕРАТУРА

1. *Большакова Е.И., Воронцов К.В., Ефремова Н.Э., Клышинский Э.С., Лукашевич Н.В., Сатин А.С.* Автоматическая обработка текстов на естественном языке и анализ данных: учеб. пособие. М.: Изд-во НИУ ВШЭ. – 2017. – 269 с.
2. Федеральная служба по финансовому мониторингу. Перечень террористов и экстремистов (действующие). URL: <http://www.fedsfm.ru/documeNts/terrorists-catalog-portal-act> (дата обращения: 14.01.2020).
3. *Практик Джоши.* Искусственный интеллект с примерами на Python: Пер. с англ. СПб.: ООО «Диалектика» – 2019. – 448 с.
4. Urllib.request – Extensible library for opening URLs. URL: <https://docs.python.org/3/library/urllib.request.html> (дата обращения: 13.01.2020).
5. BeautifulSoup Documentation. URL: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/> (дата обращения: 13.01.2020).
6. CSV – CSV File Reading and Writing. URL: <https://docs.python.org/3/library/csv.html> (дата обращения: 13.01.2020).
7. Datetime – Basic date and time types. URL: <https://docs.python.org/3/library/datetime.html> (дата обращения: 13.01.2020).
8. Matplotlib: Vizualization with Python. URL: <https://matplotlib.org/> (дата обращения: 13.01.2020).
9. Pandas documentation. URL: <https://pandas.pydata.org/docs/> (дата обращения: 13.01.2020).
10. NumPY Documentation. URL: <https://numpy.org/doc/> (дата обращения: 13.01.2020).

---

**Абдрахманов** Амангельды Нурзадаевич, аспирант ОмГУПС, [galiokpar@mail.ru](mailto:galiokpar@mail.ru)

УДК 519.2:004.421.5:004.7

## ИССЛЕДОВАНИЕ СЛУЧАЙНЫХ ГРАФОВ МЕТОДОМ ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ

В. А. Бадрызлов

Омский государственный технический университет, Омск, Россия

**Аннотация.** *Методом имитационного моделирования выполнено исследование распределения степеней связности вершин для известных классов случайных графов, применяемых для моделирования растущих сетей. Представлено описание программного средства для выполнения имитационных экспериментов.*

**Ключевые слова:** *случайный граф, предпочтительное связывание, имитационное моделирование.*

## THE STUDY OF RANDOM GRAPHS BY MEANS OF SIMULATION

V. A. Badryzlov

Omsk State Technical University, Omsk, Russia

**Abstract.** *The simulation method is used to study the distribution of degrees of vertex connectivity for known classes of random graphs used to model growing networks. The description of a software tool for performing simulation experiments is presented.*

**Keywords:** *random graph, preferential attachment, simulation*

### 1. Введение

Растущие сети, такие как Интернет, современные социальные сети, сети распространения слухов и инфекций успешно моделируются случайными графами с предпочтительным связыванием. Модели сетей в виде случайных графов могут использоваться для анализа разнообразных сетевых процессов, происходящих в реальных сетях, например, для анализа процессов распространения информации, оценки устойчивости сетей к удалению узлов и ребер, поиска наиболее важных узлов и т.д.

В 1999 г. Р. Альберт и А. Барабаша предложили графы (графы БА), объясняющие механизм роста Интернета. В основу генерации графов Р. Альберт и А. Барабаша заложили правило предпочтительного связывания. В оригинальной формулировке авторов оно выглядит следующим образом [1C:\Users\Владимир\Desktop\2020\Новые работы от 2020\3 СУИТиММ\СУИТиММ-2020\ИТОГ\Библиографический список.doc - barabasi Albert 1999]: «Учитывая растущий характер сети и начиная с небольшого числа  $m_0$  вершин, на каждом шаге времени мы добавляем новую вершину с  $m < m_0$  ребрами, которые связывают новую вершину с  $m$  вершинами, уже существующими в системе. Используя предпочтительное связывание, мы полагаем, что вероятность  $P$  того, что новая вершина будет связана с вершиной  $i$ , зависит от связности  $k_i$  этой вершины так,

что  $P(k_i) = k_i / \sum_j k_j$ . После  $t$  шагов времени модель приводит к случайной сети с  $t + m_0$  вершинами и  $mt$  ребрами».

Известна финальная вероятность  $Q_k$  того, что случайно выбранная вершина графа БА имеет степень связности  $k$  [2С:\Users\Владимир\Desktop\2020\Новые работы от 2020\3 СУИТиММ\СУИТиММ-2020\ИТОГ\Библиографический список.doc - Alber Barabasi Statistical mechanics ofС:\Users\Владимир\Desktop\2020\Новые работы от 2020\3 СУИТиММ\СУИТиММ-2020\ИТОГ\Библиографический список.doc - Alber Barabasi Statistical mechanics of]:

$$Q_k = \frac{2m(m+1)}{k(k+1)(k+2)}, \quad k \geq m. \quad (1)$$

Граф БА, таким образом, имеет степенной закон распределения степеней связности (РСС) вершин:

$$Q_k \sim \frac{c}{k^3}, \quad (2)$$

где  $c$  – некоторая константа.

Идея предпочтительного связывания оказалась плодотворной, появилось множество модификаций случайных графов с предпочтительным связыванием. Обзор некоторых из графов представлен в работах [3, 4]. Среди случайных графов можно выделить случайные графы с нелинейным правилом предпочтительного связывания (графы с НППС) [5]. Теория этих графов развита в работах [6, 7]. Одним из достижений разработанной теории является метод калибровки генераторов случайных графов по эмпирическим данным. Метод на основании известного РСС узлов реальной сети позволяет подобрать параметры генератора случайных графов таким образом, что РСС вершин построенного случайного графа совпадает с РСС моделируемой сети [6].

## 2. Постановка задачи

Согласно теории случайных графов с НППС генерация графов происходит следующим образом [6]. Строится небольшой произвольный граф-затравка. В каждый момент времени  $t = t_1, t_2, \dots$  к графу добавляется приращение графа – новая вершина с несколькими инцидентными ей ребрами. Свободные концы ребер приращения присоединяются к вершинам графа, которые выбираются случайно с вероятностями, пропорциональными заданной функции предпочтения (весу)  $f(k)$ , где  $k$  – степень связности вершины. В качестве функции предпочтения вершины случайного графа с НППС может быть выбрана произвольная неотрицательная функция степени связности вершины. Вероятность  $p_i$  того, что ребро приращения выберет для присоединения вершину  $i$  графа, пропорциональна  $f(k_i)$ :

$$p_i = \frac{f(k_i)}{\sum_j f(k_j)}, \quad i = 1, \dots, N, \quad (3)$$

где  $N$  – число вершин в графе.

Число ребер  $x$  в приращении случайного графа с НППС является случайной величиной, которая имеет дискретное распределение вероятностей  $\{r_k\}$ . Вероятности  $r_k = P(x = k) \geq 0$  наличия в приращении  $k$  ребер заданы для  $g \leq k \leq h$ , где величина  $g \geq 1$  определяет минимальное число ребер в приращении, а  $h$  – максимальное число ребер. Кроме того,  $r_g + \dots + r_h = 1$ . Распределение вероятностей  $\{r_k\}$  определяет среднее число ребер  $m = M(x) = \sum k r_k < \infty$  в приращении. Функция предпочтения  $f(k)$  и распределение вероятностей  $\{r_k\}$  числа ребер в приращении задают алгоритм генерации растущего случайного графа с НППС.

Для генерации случайных графов с предпочтительным связыванием разработана программа [8], выполняющая следующие действия:

- задание параметров графа-затравки (число вершин и ребер);
- задание функции предпочтения для вершин графа;
- задание распределения вероятностей числа ребер в приращении;
- генерация случайного графа с заданными параметрами;
- построение РСС вершин графа в заданные моменты времени;
- отслеживание степени связности выделенной вершины графа;
- вывод в файл или на консоль списка ребер графа;
- загрузка из внешнего файла списка вершин и ребер графа.

Ставится задача – используя программу и метод имитационного моделирования, выполнить исследование РСС вершин для нескольких видов случайных графов, применяемых в моделировании растущих сетей. Только для некоторых из этих графов известны финальные распределения вероятностей  $Q_k$ , как, например, в формуле (1), распределения вероятностей  $q_k(t)$ , возникающие в произвольные моменты времени  $t$ , до перехода графов в стационарные состояния, также слабо изучены.

В круг исследуемых графов включены следующие графы:

- граф 1 с функцией предпочтения  $f(k) = k$ ;
- граф 2 с функцией предпочтения  $f(k) = k^\gamma$ ,  $\gamma = 0,9$ ;
- граф 3 с функцией предпочтения  $f(k) = k^\gamma$ ,  $\gamma = 1,1$ ;
- граф 4 с функцией предпочтения  $C \cdot \ln(k)$ ,  $C = 0,1$ .

Граф 1 – модификация графов БА, состоящая в том, что количество ребер в приращении не фиксированная величина, а случайная. Графы 2 и 3 с функцией  $f(k) = k^\gamma$  исследуются в работах [9, 10]. Важность показателя  $\gamma$  и его влияние на свойства генерируемых графов с соответствующей функцией предпочтения показана в работах [11, 12].

Граф 4 имеет функцию предпочтения, воспроизводящую закон субъективной ценности Даниила Бернулли [13]. Закон утверждает, что в зависимости от величины  $\omega$  богатства человека ценность  $U(\omega)$  этого богатства для человека определяется выражением  $U(\omega) = k \cdot \ln(\omega)$ , где  $k$  – коэффициент, определяющий единицу полезности. Подобный подход соответствует и основному психофизическому закону Вебера-Фехнера [14], определяющему восприятие человеком звука, света, давления и других раздражителей. В математическом виде закон имеет вид:

$$S = C \cdot \ln(I) + K, \quad (4)$$

где  $S$  – интенсивность ощущения,  $I$  – сила раздражителя,  $C$  и  $K$  – константы. Величина константы  $C$  (константа Вебера) определяется тем, для какого органа чувств или ощущений записан закон Вебера-Фехнера. В частности, для ощущений громкости звука  $C = 0,1$ , для других ощущений человека (высота звука, яркость света, вес предметов и др.) константа Вебера имеет значение от 0,003 до 0,2.

Соответствующая закону (4) функция предпочтения  $f(k) = C \cdot \ln(k)$  специфическим образом задает предпочтительность вершин случайного графа для присоединения ребер приращения. Например, при  $C = 1$  вероятности присоединения к вершинам со степенью связности 100 010 и 100 100, практически не различаются, в то время как для вершин со степенями связности 10 и 100 вероятности присоединения различаются почти в 2 раза. Это обеспечивает относительно высокую чувствительность процедуры выбора предпочтительной вершины к изменению степени при малых значениях  $k$  и низкую чувствительность к изменению степени при больших значениях  $k$ .

### 3. Результаты моделирования

Все исследуемые графы выращиваются из графа-затравки, имеющего вид кольца из 5 вершин. На каждом шаге модельного времени генерируется приращение графов со случайным количеством ребер. Число ребер  $x$  в приращении задается дискретным распределением вероятностей – равновероятно в приращении могут появиться от 1 до 5 ребер. Многократное выращивание графов позволяет построить усредненное РСС вершин – одну из важнейших характеристик как случайных графов, так и моделируемых с их помощью реальных сетей. Распределения  $q_k(t)$  степеней связности графов 1-4 в момент времени  $t = 10000$  изображены ниже на Рис. 1-4 соответственно.

Дадим качественную характеристику полученным РСС. Хорошо известно, что график степенного распределения в логарифмической шкале представляет собой прямую линию. Для графов 1-3 можно предположить, что распределения степеней в средней части близки к степенному, а РСС вершин графа 4 значительно от него отличается. Для всех исследуемых графов отмечается особенность – часть РСС, построенная для степеней  $k = 1, 2, \dots, 5$  не укладывается в «схему» степенного распределения. Однако эта часть графика более всего соответствует стационарным вероятностям, так как постоянно, на каждом шаге моделирования, появляющиеся новые вершины по условиям проведения экспериментов имеют равновероятно от 1 до 5 ребер в приращении и в каждом из графов таких вершин более половины от общей численности вершин. Очевидно, что при иных распределениях вероятностей  $\{r_k\}$  числа ребер в приращениях будет генерироваться специфическая нелинейная начальная часть РСС вершин графов.

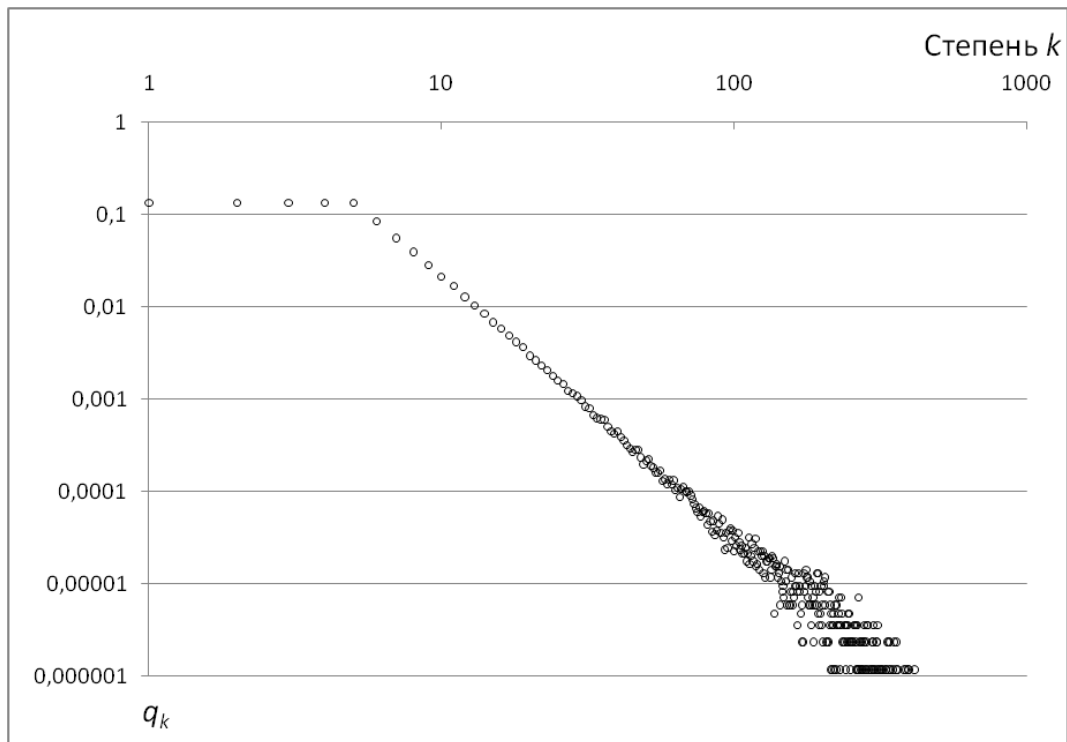


Рис. 1. Распределение степеней  $q_k(t)$  при  $t = 10^4$  в графе с функцией предпочтения  $f(k) = k$

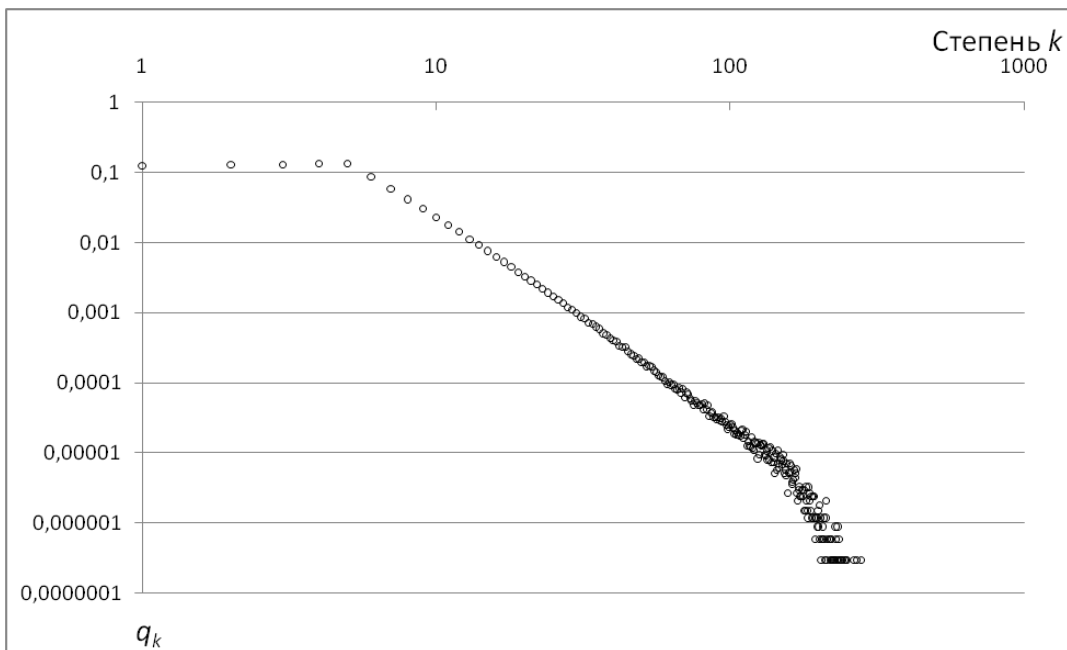


Рис. 2. Распределение степеней  $q_k(t)$  при  $t = 10^4$  в графе с функцией предпочтения  $f(k) = k^\gamma$ ,  $\gamma = 0,9$

Правая часть РСС, соответствующая большим степеням связности вершин, также имеет особенности. В ряду распределения начинаются пропуски из-за того, что в выращенных графах отсутствуют вершины с некоторыми значениями степеней  $k$ . Количество вершин с большими  $k$  еще невелико, поэтому оценки соответствующих вероятностей  $q_k$  имеют высокую относительную погрешность. В целом по построенным графикам можно отметить, что функция предпочтения



вершин определяет РСС вершин графа, а теория случайных графов с НППС позволяет по эмпирическим данным о РСС сетей определять параметры генераторов случайных графов для построения адекватных моделей.

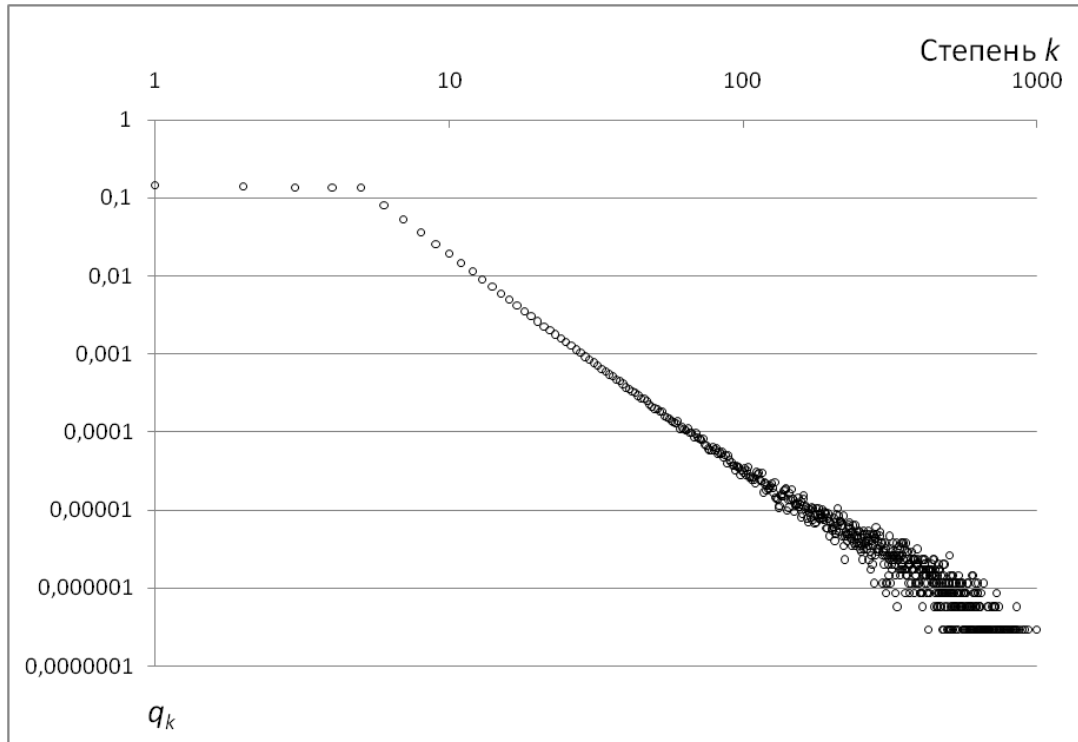


Рис. 3. Распределение степеней  $q_k(t)$  при  $t = 10^4$  в графе с функцией предпочтения  $f(k) = k^\gamma$ ,  $\gamma = 1,1$

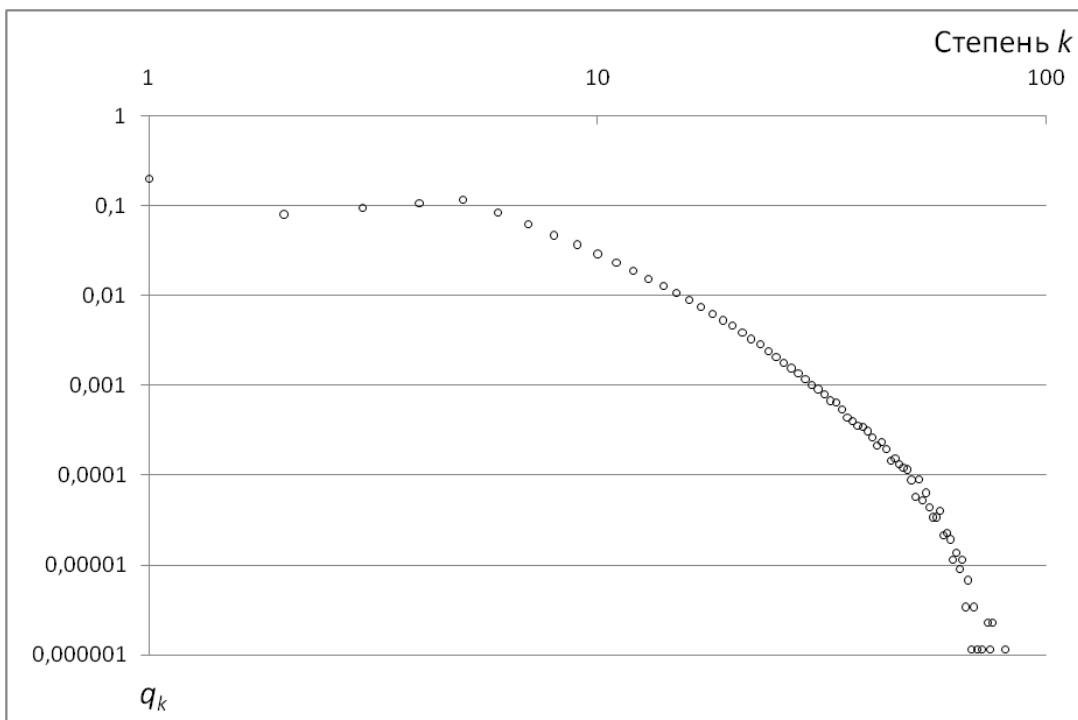


Рис. 4. Распределение степеней  $q_k(t)$  при  $t = 10^4$  в графе с функцией предпочтения  $C \cdot \ln(k)$ ,  $C = 0,1$

#### 4. Средства моделирования

Программа «Генерация случайного графа с предпочтительным связыванием v.1» существует в форме модели, разработанной для системы имитационного моделирования AnyLogic. Вершины графа в модели представлены как агенты в рамках парадигмы агентного моделирования. Возникающие между агентами связи – это ребра графа. Элементы модели представлены на Рис. 5.

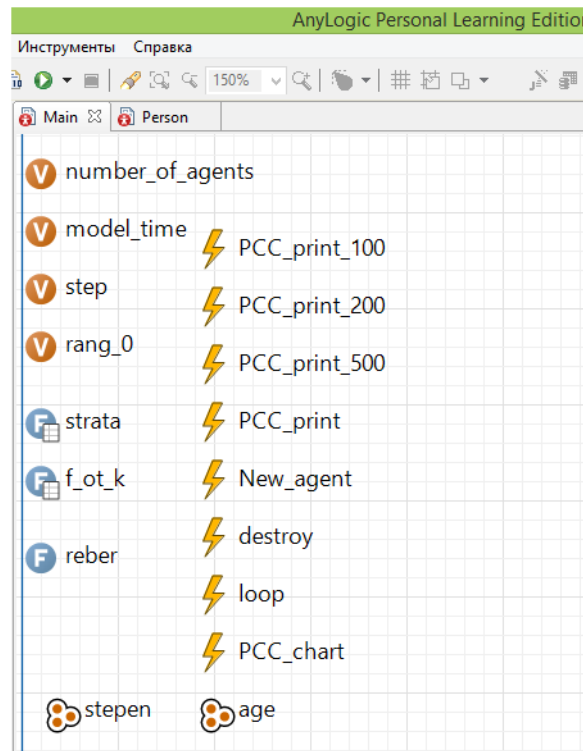








Рис. 5. Элементы модели

Выделяются следующие классы элементов модели:

- среда обитания агентов ;
- переменные модели ;
- функции  и ;
- коллекции ;
- события .

Среда обитания **graf** агентов модели задается как прямоугольная область, в которой сначала располагаются агенты, образующие граф-затравку. Каждый агент визуально представляется как окружность малого радиуса, а связи между агентами (вершинами графа) строятся в виде отрезков прямых линий.

Изначально для построения графа-затравки в среде обитания **graf** задаются переменные:

**number\_of\_agents** – число вершин в графе-затравке;

**rang\_0** – степень связности вершин графа-затравки.

Переменные, хранящие значения времени:

**step** – шаг модельного времени;

**model\_time** – начальное время в модели.

Функция **reber** задает в модели распределение вероятностей  $r = \{r_1, r_2, \dots\}$  числа ребер в приращении. Табличная функция **f\_ot\_k** задает функцию предпочтения вершин в табличном виде. Для конкретных графов функции **reber** и **f\_ot\_k** могут быть переопределены так, как это необходимо для исследователя. Примеры задания функций **reber** и **f\_ot\_k** показаны на Рис. 6 и 7 соответственно.

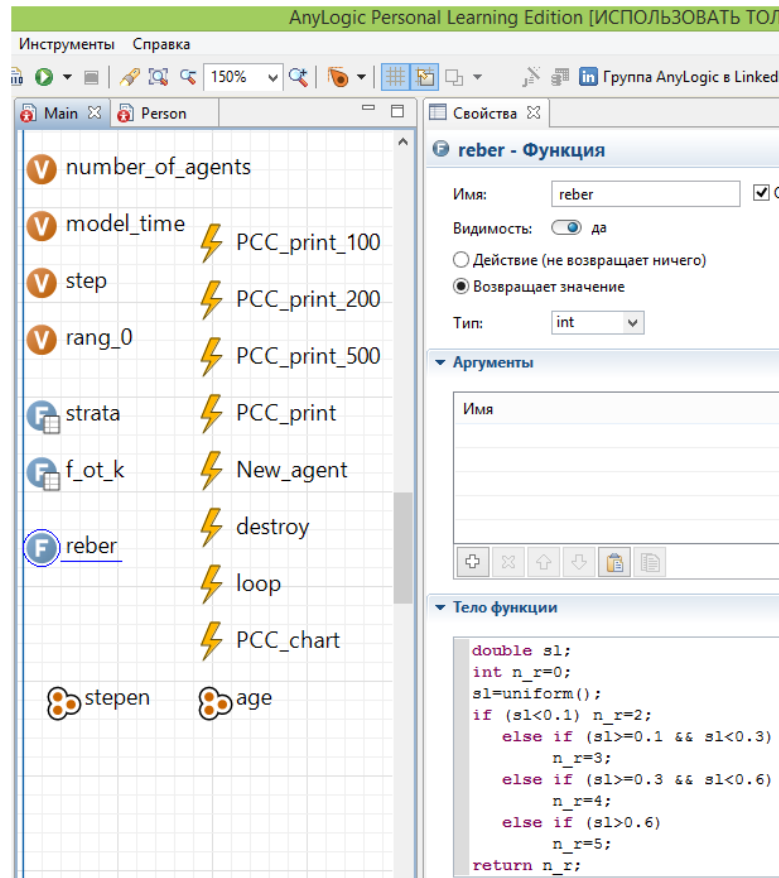


Рис. 6. Задание распределения вероятностей числа ребер в приращении с помощью функции **reber**

Коллекция **age** хранит возраст вершин графа, а коллекция **stepen** – степени связности вершин. Управление процессом моделирования осуществляется с помощью событий. Для событий задается время их выполнения, либо однократно, либо с заданной периодичностью. Период между реализациями событий может задаваться как случайная величина.

Событие **loop** задает бесконечную последовательность экспериментов по генерации случайного графа, которую можно прервать по желанию исследователя.

Событие **destroy** выполняется либо с определенной периодичностью, либо в случайные моменты времени. Событие состоит в том, что происходит удаление какой либо вершины графа со всеми ее ребрами. Выбор вершины для удаления может осуществляться различными способами, например, случайным образом. Возможно удаление вершины с максимальной степенью или вершины, удовлетворяющей какому-либо особому условию.

События **PCC\_print\_100**, **PCC\_print\_200**, **PCC\_print\_500** – однотипные события, состоящие в том, что в заданные моменты времени рассчитывается РСС вершин графа и выводится для анализа на консоль или в файл. События **PCC\_chart** и **PCC\_print** выполняются с заданной периодичностью и при каждом срабатывании выводят РСС вершин графа, что позволяет изучать динамику изменения РСС.

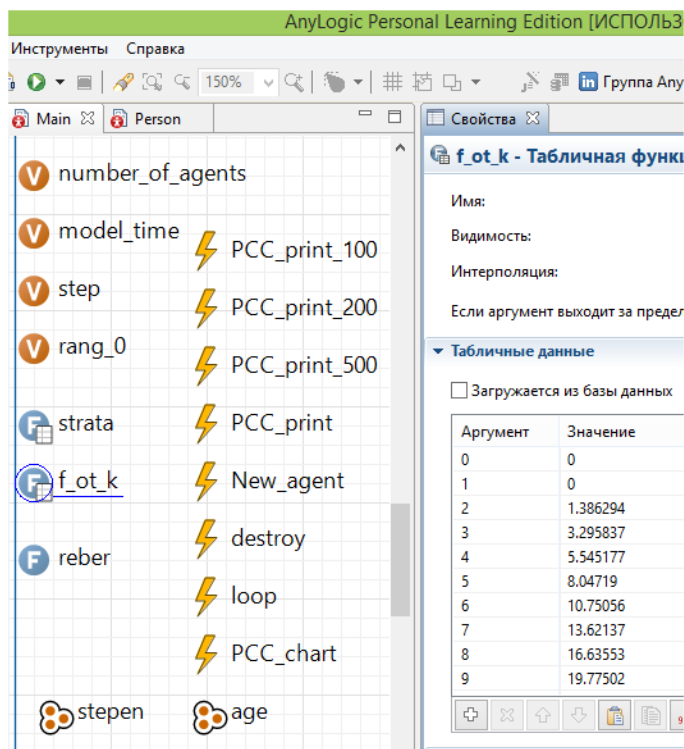


Рис. 7. Задание функции предпочтения вершин **f\_ot\_k** в табличном виде

Основное событие в модели – **New\_agent**. Оно запускается на каждом шаге модельного времени и состоит в том, что появляется новая вершина графа, разыгрывается число ребер в приращении, устанавливаются связи новой вершины с уже существующими, то есть производится вся совокупность действий по созданию и добавлению к случайному графу очередного приращения. При выполнении события учитываются значение функции **reber** и значение функции предпочтения **f\_ot\_k**.

## 5. Заключение

Для моделирования растущих сетей эффективно применяются случайные графы с предпочтительным связыванием. Особое место среди них занимают графы с НППС. Теория этих графов позволяет определять параметры генератора случайных графов таким образом, чтобы у построенной модели распределение степеней вершин было идентичным распределению степеней узлов реальной моделируемой сети. Для выращивания случайных графов с найденными параметрами генерации разработана программа, возможности которой продемонстрированы выше на примере построения РСС вершин нескольких известных классов случайных графов.

Разработанная программа имеет потенциал для исследования не только процессов роста сетей, но и для исследования процессов распространения информации в сетях, для исследования процессов изменения свойств сетей в результате случайной убыли вершин или ребер графа.

## ЛИТЕРАТУРА

1. *Barabasi A.L., Albert R.* Emergence of scaling in random networks // *Science*. – 1999. – V.286. – P. 509-512.
2. *Dorogovtsev S.N., Mendes J.F., Samukhin A.N.*, Structure of growing networks with preferential linking // *Physical Review Letters*. – 2000. – V.85(21). – P.4633-4636.
3. *Бадрызлов, В.А.* Классификация случайных графов с предпочтительным связыванием / В.А. Бадрызлов // *Омский научный вестник*. – 2017. - №4 (154). – С. 124-128.
4. *Щербакова, Н.Г.* Модели сетей с предпочтительным присоединением / Н.Г. Щербакова // *Проблемы информатики*. – 2019. – №3 (44). – С.46-61.
5. *Задорожный, В.Н.* Случайные графы с нелинейным правилом предпочтительного связывания / В.Н. Задорожный // *Проблемы управления*. – 2010. – № 6. – С. 2-11.
6. *Задорожный, В.Н.* Структурные свойства безмасштабного графа Барабаши-Альберт / В.Н. Задорожный, Е.Б. Юдин // *Автоматика и телемеханика*. – 2012. – №4. – С.131-150.
7. *Задорожный, В.Н.* Калибровка случайных графов предпочтительного связывания по распределениям степеней вершин и ребер / В.Н. Задорожный, Е.Б. Юдин // *Омский научный вестник*. – 2017. – № 1 (151). – С. 114–119.
8. *Бадрызлов, В.А.* Генерация случайного графа с предпочтительным связыванием v.1» / В.А. Бадрызлов // *Хроники объединенного фонда электронных ресурсов «Наука и образование»*. – 2017. – №10(101). – С.12.
9. *Krapivsky P.L., Redner S., Leyvraz F.* Connectivity of Growing Random Networks // *Phys. Reviews*. – 2000. – V.85. – P.4626-4632.
10. *Krapivsky P.L., Redner S.* Organization of growing random networks // *Phys. Reviews E*. – 2001. – V.63. – P.066123.
11. *Задорожный, В.Н.* Исследование трех актуальных классов растущих графов / В.Н. Задорожный, Е.Б. Юдин, В.А. Бадрызлов // *Информационные технологии и автоматизация управления: материалы IX Всерос. науч.-практ. конф. студентов, аспирантов, работников образования и промышленности (Омск, 17-19 мая 2018 г.) / Минобрнауки России, ОмГТУ, каф. АСОИУ*. – Омск: Изд-во ОмГТУ, 2018. – С. 59-77.
12. *Задорожный, В.Н.* Распределения степеней вершин в растущих графах: методы расчета с контролем погрешностей / В.Н. Задорожный, Е.Б. Юдин, В.А. Бадрызлов // *Омский научный вестник*. – 2018. – №4(160). – С. 156-160.
13. *Ватник, П.* Даниил Бернулли – экономист [Электронный ресурс]. – URL: <https://seinst.ru/page34/> (дата обращения: 15.04.2020).
14. *Маклаков, А.Г.* Общая психология: Учебник для вузов / А.Г. Маклаков. – СПб.: Питер, 2016.

---

**Бадрызлов** Владимир Александрович, старший преподаватель, каф. «Математические методы и информационные технологии в экономике», v\_bad@mail.ru

УДК 519+005.5:378.342

## ОПТИМИЗАЦИЯ И АВТОМАТИЗАЦИЯ БИЗНЕС-ПРОЦЕССА НАЗНАЧЕНИЯ ПОВЫШЕННОЙ СТИПЕНДИИ В ОМГТУ

*И. М. Баркова, Ю. П. Галдина, К.В. Кравченко, Ю.А. Бахмутский*

Омский государственный технический университет, Омск, Россия

**Аннотация.** *Проведен анализ процесса назначения повышенной стипендии в Омском государственном техническом университете. Осуществлен обзор программных средств автоматизации этого бизнес-процесса с обоснованием выбора оптимального варианта.*

**Ключевые слова:** *бизнес-процесс, стипендия, стипендиальная комиссия.*

## OPTIMIZATION AND AUTOMATION OF THE BUSINESS PROCESS OF APPOINTING AN INCREASED SCHOLARSHIP AT OMSTU

*I. Barkova, J. Galdina, K. Kravchenko, Y. Bakhmutsky*

Omsk State Technical University, Omsk, Russia

**Abstract.** *Scholarship increase research was performed within Omsk State Technical University. An observation of software automation tools for the business process was carried out with the rationale for choosing the best option.*

**Keywords:** *business-process, scholarship, scholarship commission.*

### 1. Введение

Назначение стипендии является одной из основных задач организации учебного процесса.

Объем работ различных задач, выполняемых сотрудниками деканата, постоянно растет, появляются дополнительные формы отчетности. Увеличение штата сотрудников деканата невозможно в оперативном режиме и связано с дополнительными и регулярными затратами. Поэтому целесообразно сконцентрировать усилия на последовательной оптимизации и автоматизации существующих процессов для максимального высвобождения рабочего времени имеющихся трудовых ресурсов. В этом и состоит актуальность данной задачи.

Сбор и обработка всех документов требуют больших временных затрат от сотрудников деканатов, а подготовка документов и их дальнейшее исправление – от студентов. Также возможна потеря документов в силу преобладания человеческого фактора. Уменьшение времени работы с документами и снижение рисков потери данных подразумевает необходимость оптимизации и автоматизации бизнес-процесса назначения стипендии.



Модель текущего бизнес-процесса наглядно показывает, как происходит назначение повышенной стипендии. Отражена подача документов, оформленных по примерному шаблону заполнения. Далее происходит проверка данных документов и подготовка документов для стипендиальной комиссии в бумажном виде. Затем комиссия рассматривает документы каждого претендента и формирует списки студентов, которым будет назначена повышенная стипендия. Данные списки передаются в деканат. На основании этих списков формируется проект приказа о назначении стипендии и передается проректору по ВР и СВ. Проректор согласует проект приказа и после согласования утверждает приказ. На основании утвержденного приказа стипендиальная группа бухгалтерии осуществляет выплату стипендий студентам.

Проблемы, выявленные при анализе модели текущего бизнес-процесса:

- существует риск потери документов;
- отсутствует единый шаблон по оформлению документации. Это приводит к появлению ошибок и увеличению времени, затрачиваемого на подачу документов;
- затрачивается время на составление проекта приказа и согласование данного проекта.

Для устранения выявленных недостатков следует выбрать программный продукт, применение которого позволило бы провести автоматизацию и оптимизацию рассматриваемого процесса. В Табл. 1 приведен обзор существующих программных продуктов и их сравнение по показателям, которые являются ключевыми с точки зрения поставленной задачи.

**Таблица 1.** Обзор программных продуктов

Показатели	1С: Документооборот	Галактика	Парус-ВУЗ
Функционал	удобный/понятный	удобный/понятный	удобный/понятный
Стоимость	низкая	средняя	средняя
Тех. поддержка и обновление	средняя	высокая	высокая
Обучение пользователей	легкое обучение	среднее обучение	среднее обучение
Технические специалисты	много	средне	средне
Внесение изменений в систему	возможно	нет	нет
Внедрены в ОмГТУ	да	да	нет
Хранение документов в информационной базе	да	да	да
Разработка шаблонов документов	да	возможно, но с затруднениями	возможно, но с затруднениями
Согласование проекта документа	да	нет	нет
Утверждение проекта документа	да	нет	нет
Регистрация документа	да	нет	нет
Исполнение документа	да	нет	нет
Отправка документов	да	да	да



На основании обзора можно сделать вывод, что программа 1С: Документооборот наиболее подходит для автоматизации данного процесса, так как имеет ряд преимуществ перед другими программами. Главными преимуществами являются открытость кода, которая позволяет легко вносить необходимые изменения, и активное использование данной системы в университете.

В ходе исследования было выявлено, что одним из путей устранения существующих недостатков является применение Электронного документооборота. Благодаря данному инструменту возможно создание стабильного шаблона оформления документов [2], что, в свою очередь, позволит исключить ошибки и сократить время, затрачиваемое на их исправление. Студенты смогут загружать всю необходимую документацию в систему, а также оперативно получать уведомления о возможности внесения исправлений. Будет исключена возможность потери документов, а также упростится поиск документов по студентам. Стипендиальная комиссия сможет просматривать и формировать списки студентов и отправлять их в деканат через систему документооборота. В базе будет создан шаблон проекта приказа на стипендию, который будет заполняться сотрудником деканата путем внесения только динамически меняющихся данных. Также будет сформирован маршрут согласования приказа о назначении стипендии, что значительно сократит время на согласование документа. С помощью документооборота сотрудники деканата смогут оперативно информировать студентов о времени начала и окончания приема документов, а также результатах назначения стипендии.

Целевой бизнес-процесс с учетом внедрения системы электронного документооборота представлен на диаграмме в нотации BPMN на Рис. 2.

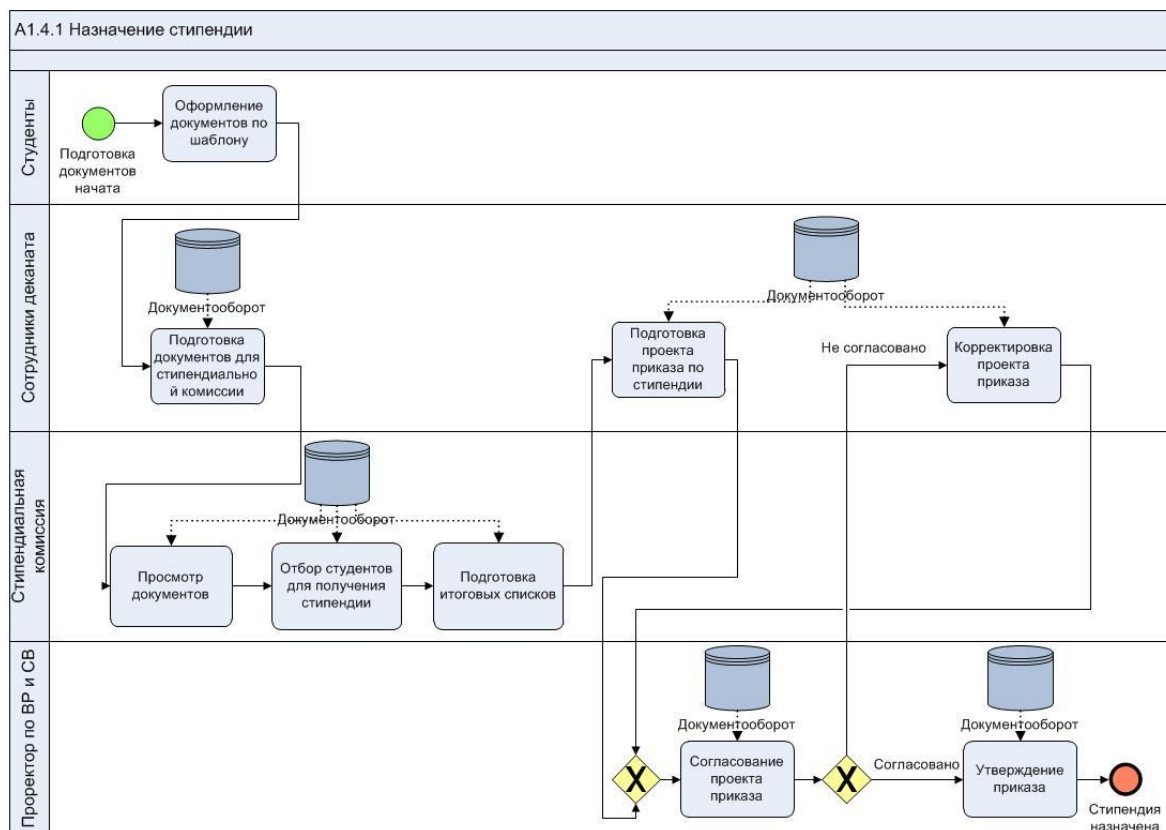


Рис. 2. Диаграмма процесса в нотации BPMN «как будет»

Внедрение системы Электронного документооборота позволит:

- обеспечить оптимальное хранение документов в базе;
- сократить время, затрачиваемое на подачу и прием документов, подготовку и согласование проекта приказа.

### 3. Заключение

Таким образом, на основании анализа процесса назначения стипендии в статье построена текущая модель бизнес-процесса и выявлены недостатки его организации. В качестве рекомендации предложена целевая модель процесса и описаны достоинства применения электронного документооборота для решения проблем, выявленных в данном процессе.

### ЛИТЕРАТУРА

1. Положение «О стипендиальном обеспечении и других формах материальной поддержки обучающихся ОмГТУ». Режим доступа: [https://www.omgtu.ru/educational\\_activities/dokumenty\\_smk/Pologeniya/П%20ОмГТУ%2061.11-2017%20О%20стипендиальном%20обеспечении%20и%20других%20%20формах%20материальной%20поддержки%20обучающихся%20ОмГТУ%20\(зашифрован\).pdf](https://www.omgtu.ru/educational_activities/dokumenty_smk/Pologeniya/П%20ОмГТУ%2061.11-2017%20О%20стипендиальном%20обеспечении%20и%20других%20%20формах%20материальной%20поддержки%20обучающихся%20ОмГТУ%20(зашифрован).pdf) (дата обращения: 13.03.2020).
2. Ульянцева С.Э. Управление документами: быстро, эффективно, своими силами. На примере «1С:Документооборота 8». Режим доступа: <https://online.1c.ru/books/book/20337851/> (дата обращения: 05.04.2020).

---

**Баркова** Ирина Михайловна, магистрант, barkowa.ira2018@yandex.ru

**Галдина** Юлия Павловна, магистрант, juliagaldina96@mail.ru

**Кравченко** Ксения Владимировна, старший преподаватель, каф. «Математические методы и информационные технологии в экономике», trr474747@mail.ru

**Бахмутский** Юрий Андреевич, заведующий кафедрой, каф. «Математические методы и информационные технологии в экономике», shd@satoryomsk.ru

УДК 004.624:65.011.56

## РАЗРАБОТКА БЕСПРОВОДНОГО ДАТЧИКА ТЕМПЕРАТУРЫ С СОВМЕСТНЫМ ИСПОЛЬЗОВАНИЕМ МИКРОКОМПЬЮТЕРА ESP32

*А. А. Берестин, С. Д. Омельченко, А. З. Кулганатов*

Южно-Уральский государственный университет, Челябинск, Россия

**Аннотация.** *Рассматривается опыт разработки беспроводного датчика температуры на базе микроконтроллера ESP32 и дается всесторонний обзор возможностей в области измерения и обработки данных на платформе ESP32.*

**Ключевые слова:** *датчик температуры, микроконтроллер ESP32.*

## DEVELOPMENT OF A WIRELESS TEMPERATURE SENSOR WITH JOINT USE OF THE ESP32 MICROCOMPUTER

*A. A. Berestinov, S. D. Omelchenko, A. Z. Kulganatov*

South Ural State University, Chelyabinsk, Russia

**Abstract.** *The experience of developing a wireless temperature sensor based on the ESP32 microcontroller is examined and a comprehensive overview of the possibilities in the field of measurement and data processing on the ESP32 platform is given.*

**Keywords:** *temperature sensor, microcontroller ESP32.*

### 1. Введение

В настоящее время IoT, системы автоматизации умного дома и встроенные системы стремительно развиваются. Это тесно связано с разработкой доступных аппаратных модулей и процессоров. Производятся платы разработки, на которых интерфейс связи и периферийные устройства реализованы вместе с основным процессором. В настоящее время популярность чипа ESP32 растет, и развиваются не только аппаратные варианты реализации устройств на базе этого чипа, но и программные решения для данного микроконтроллера. Широкое сообщество разработчиков, а также ученых занимаются использованием чипа ESP32 более высокого поколения в качестве преемника микроконтроллера ESP8266. Последние научные статьи доказывают популярность данного чипа ESP32 в различных областях.

В статье описывается применение Интернета вещей с микроконтроллером ESP32 для реализации обработки данных и беспроводной связи. Была опробована идея использования низкоуровневой аппаратной и программной оптимизации для создания программной архитектуры, чтобы максимально использовать это оборудование и снизить стоимость более дорогих аппаратных решений.

Наша цель в архитектурном дизайне была достигнута. Мы создали основу общей структуры и определили сценарии ее использования, а затем реализовали эти сценарии с использованием компонентов.

В этой работе мы сосредоточимся на разработке приложений для устройств, доступных на рынке. За основу был взят чип ESP32 (ESP-WROOM-32) от производителя Espressif и датчик температуры DS18S20. Таким образом, используются традиционные и доступные низкоуровневые модули, доступные на рынке. Основная цель - разработать устройство IoT на основе чипа ESP32. Предлагаемые сценарии использования:

Удаленный доступ;

Контроль и диагностика устройства;

Синхронизация нескольких устройств одновременно.

## 2. Постановка задачи

В настоящее время в системах диспетчеризации зданий начинают активно использоваться беспроводные протоколы передачи данных. Одним из наиболее популярных протоколов, является Wi-Fi, точка доступа есть практически в каждом помещении, следовательно, при использовании беспроводных датчиков уменьшаются расходы на монтаж проводных сетей.

Целью данной работы является разработка беспроводного термодатчика, для систем диспетчеризации зданий.

## 3. Теория

ESP32 - это мощный микроконтроллер со встроенным Wi-Fi 802.11 b/g/n, двухрежимным Bluetooth версии 4.2 и различными периферийными устройствами. Он является передовым преемником чипа 8266, в основном, в реализации двух ядер с тактовой частотой до 240 МГц. По сравнению со своим предшественником, кроме этих функций, он также увеличивает количество выводов GPIO с 17 до 36, количество каналов ШИМ на 16 и оснащен 4 МБ флэш-памяти.

Чип ESP32 был разработан компанией Espressif Systems, которая в настоящее время предлагает несколько версий ESP32 SoC в форме ESP32 Developer Kit, ESP32 Wrover Kit, который также включает в себя SD-карту и 3,2-дюймовый жидкокристаллический-дисплей и последний, но не менее важный комплект ESP32 Azure IoT с USB-мостом и другими встроенными датчиками.

ESP32 включает в себя два ядра (процессор Xtensa LX6, выполненный по 40 нм технологии). Ядра процессора могут контролироваться индивидуально. Имеется 520 КБ встроенной SRAM для данных и инструкций. Некоторые модули SoC, такие как ESP32-Wrover, оснащены 4 МБ внешней флэш-памяти SPI и дополнительными 8 МБ SPI PSRAM (псевдостатическая ОЗУ). С возможностью использовать SPI, I2S, I2C, CAN, UART, Ethernet MAC и IR в различных количествах, в зависимости от типа платы. Стандартное оборудование также включает

в себя датчик Холла, датчик температуры и сенсорный датчик, другие встроенные датчики реализованы в Azure IoT и Developer Kit. Микроконтроллер также обеспечивает криптографическое аппаратное ускорение: AES, SHA-2, RSA, криптография с эллиптическими кривыми (ECC) и генератор случайных чисел (RNG). На Рис. 1 показаны варианты микроконтроллеров ESP32, с которыми мы работаем в Департаменте систем управления и приборостроения.

Платы ESP32 производятся в прототипах, которые можно использовать в приложениях для умного дома, автоматизации, в носимых устройствах, аудио-приложениях, облачных IoT-приложениях и т.д. Чип ESP32 от компании Espressif. ESP32 может работать как полностью автономная система или как ведомое устройство для главного MCU, уменьшая накладные расходы стека связи на главном процессоре приложений. Микросхема включает в себя встроенный Wi-Fi, Bluetooth и другие интерфейсы связи, такие как SPI и I2C / UART.



Рис. 1. Различные модели плат ESP32 [3]

#### 4. Результаты экспериментов

Для решения данной задачи используется микроконтроллер ESP32-WROOM-32U [1], имеющий встроенный Wi-Fi 802.11b/g/n/e/I и Bluetooth v4.2, при этом обладающий низким энергопотреблением.

На базе выбранного микроконтроллера была собрана экспериментальная установка, изображенная на Рис. 2.

#### 5. Обсуждение результатов

Данная экспериментальная установка позволила успешно передать данные с 2 термодатчиков DS18S20, подключенных по шине 1-wire. Каждый термодатчик имеет свой индивидуальный адрес, что позволяет подключать несколько датчиков, используя всего один вход контроллера. Датчик измеряет температуру в диапазоне от -55 °C до +125 °C [2].

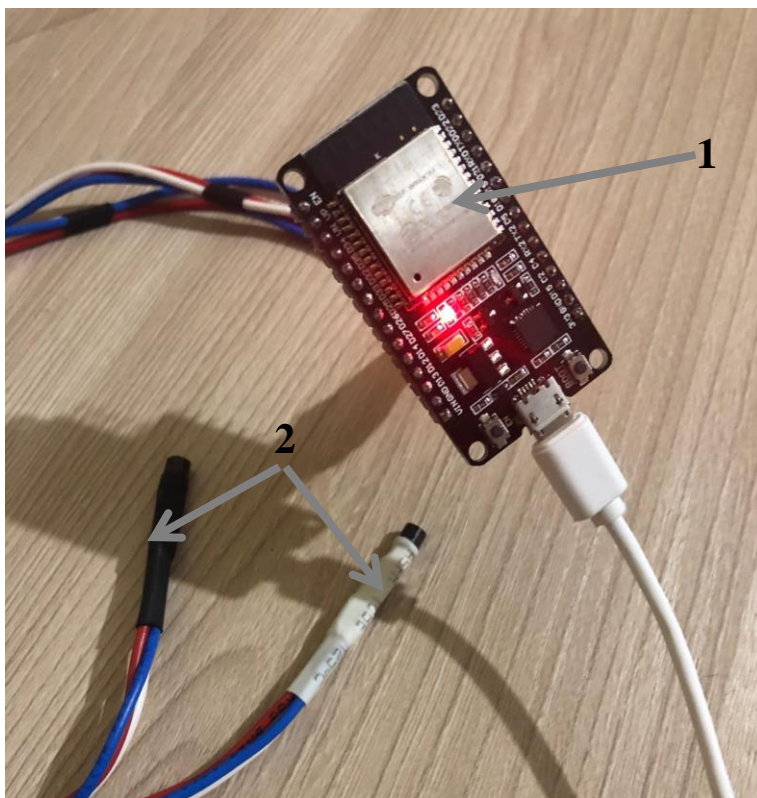


Рис. 2. Экспериментальная установка для беспроводной передачи данных с датчика температуры: 1 – микроконтроллер, 2 – термодатчики

## 6. Выводы и заключение

Основными преимуществами данной экспериментальной установки являются: использование беспроводной связи, мобильность, небольшие размеры, низкое энергопотребление, возможность расширения функционала (наличие резервных входов для подключения датчиков влажности, освещенности и т.д.), низкая цена.

## ЛИТЕРАТУРА

1. ESP32. URL: [http://www.gamma.spb.ru/images/pdf/esp32\\_datasheet\\_en.pdf](http://www.gamma.spb.ru/images/pdf/esp32_datasheet_en.pdf).
2. DS18S20. URL: <https://static.chipdip.ru/lib/073/DOC000073557.pdf>.
3. Рисунок платы разработок ESP32. URL: <https://makeradvisor.com/esp32-development-boards-review-comparison/>.

---

**Берестин** Амиржан Ахметжанович, студент, [berestinov.ru@mail.ru](mailto:berestinov.ru@mail.ru)

**Омельченко** Сергей Дмитриевич, студент, [sergei.omel.97@gmail.com](mailto:sergei.omel.97@gmail.com)

**Кулганатов** Аскар Зайдакбаевич, студент, [kulganatov97@gmail.com](mailto:kulganatov97@gmail.com)

УДК 004.9: 65.011.56

## АНАЛИЗ СИСТЕМ ЭЛЕКТРОННОГО ДОКУМЕНТООБОРОТА ДЛЯ АВТОМАТИЗАЦИИ ПРОЦЕССА «ВНУТРЕННИЙ АУДИТ»

*Е. О. Будзинская, Е. А. Калиберда, А.А. Новиков*

Омский государственный технический университет, Омск, Россия

**Аннотация.** *Проводится сравнительный анализ систем электронного документооборота как инструмента разработки модуля для автоматизации процесса «Внутренний аудит». Определяются достоинства и недостатки различных систем электронного документооборота. На основе проведенного анализа делается вывод о наиболее предпочтительной, в соответствии с требуемыми функциям и критериями, системе.*

**Ключевые слова:** *интегрированная система менеджмента, система электронного документооборота, внутренний аудит.*

## ANALYSIS OF ELECTRONIC DOCUMENT MANAGEMENT SYSTEMS FOR AUTOMATING THE «INTERNAL AUDIT» PROCESS

*E. O. Budzinskaya, E. A. Kaliberda, A.A. Novikov*

Omsk state technical University, Omsk, Russia

**Abstract.** *A comparative analysis of electronic document management systems as a tool for developing a module for automating the Internal Audit process is presented. The advantages and disadvantages of various electronic document management systems are determined. Based on the analysis, it is concluded that the system is most preferable, in accordance with the required functions and criteria.*

**Keywords:** *integrated management system, electronic document management system, internal audit, ISO.*

### 1. Введение

Внутренний аудит (ВА) является высшей формой контроля руководством системы менеджмента качества предприятия. Он проводится для того, чтобы определить соответствие деятельности и результатов в области интегрированной системы менеджмента (ИСМ) [1].

Результаты внутренних проверок Компании служат основой данных для анализа со стороны руководства и позволяют Компании проработать несоответствия/соответствия стандартам ISO. Поэтому важным вопросом для Компании является планирование, организация, проведение и анализ результатов внутреннего аудита.

Настоящая статья посвящена анализу существующих систем электронного документооборота (СЭД) для автоматизации процесса «Внутренний аудит».



## 2. Постановка задачи

Предварительным этапом анализа систем аналогов и выбора подходящей платформы для автоматизации процесса «Внутренний аудит» является формирование основного функционала системы. В Компании предъявляются следующие основные требования к функционированию системы:

- система должна использовать существующую инфраструктуру Компании;
- утверждение комплекта документов должно осуществляться с применением стандартного механизма согласования СЭД;
- формирование программы внутреннего аудита в СЭД;
- согласование программы внутреннего аудита, а также информирование о согласовании программы внутреннего аудита через электронную почту;
- возможность загрузки программы внутреннего аудита в MS «Excel».

## 3. Теория

Для того чтобы выбрать СЭД, соответствующую поставленным требованиям, необходимо провести сравнительный анализ трех систем СЭД: «TESSA», «1С: Битрикс 24» и «Дело».

СЭД «TESSA» – гибкая платформа для создания решений по автоматизации документооборота и бизнес-процессов компаний в различных сферах бизнеса. Разработчиком программы является компания Syntellekt (СИНТЕЛЛЕКТ). Основным направлением компании является разработка и внедрение решений по автоматизации документооборота и бизнес-процессов, интеграция приложений и информационных систем, а также разработка заказного программного обеспечения. На платформе «TESSA» могут быть автоматизированы различные функциональные блоки документооборота: электронный архив документов; автоматизация работы канцелярии; автоматизация процессов, специализированных для различных сфер бизнеса. Система оптимизирована для работы с большим объемом данных.

Основные преимущества:

- высокая скорость работы;
- современный, удобный и интуитивно понятный интерфейс;
- возможность настройки и разработки расширений для системы;
- возможность организации сложных бизнес-процессов с наглядной визуализацией хода выполнения процесса;
- возможность интеграции с другими информационными системами.

СЭД «TESSA» предоставляет пользователям возможность согласовывать документы при помощи любого почтового клиента, без доступа в приложение, дает пользователям загружать необходимые документы. Также при необходимости есть функция замещения и делегирования [2].



В программном продукте «1С: Битрикс 24» отсутствуют такие функциональные стандарты как возможность выполнять замещение и администрирование. Также отсутствует анализ и отчетность по просроченным документам, мобильное согласование, формирование документа по шаблону в СЭД, визуализация бизнес-процессов. К безусловным преимуществам системы можно отнести: возможность настройки конфигурации под требования заказчика, наличие мобильных приложений, регулярный выпуск обновлений [3].

Программный комплекс «Дело» разработан компанией разработанная «Электронные Офисные Системы» (ЭОС). Данная система предназначена для автоматизации документооборота и делопроизводства на предприятии. Система позволяет отслеживать перемещение электронного документа и создавать проекты документации. Помимо этого, система отличается общей отлаженностью и широкой функциональностью [4]. СЭД позволяет регистрировать документы, применять электронную подпись, интегрировать с различными системами, создавать стандартные отчеты, мониторить документы. Однако СЭД имеет ряд своих недостатков: несколько устаревший интерфейс и определенная сложность в освоении новым пользователем программы. У системы «Дело» отсутствуют такие функции как возможность работать в режиме «оффлайн», работать с договорами, выполнять делегирование и замещение, визуализирование бизнес-процессов, анализ отчетности по просроченным документам, автоматическое уведомление информации о документе в электронную почту.

Для выполнения сравнительного анализа были выбраны необходимые функции и критерии, которым должна соответствовать выбранная СЭД. Сравнительный анализ представлен в таблице 1.

**Таблица 1.** Сравнительный анализ

Выполняемые функции и критерии	Системы аналоги		
	«TESSA»	«1С: Битрикс 24»	«Дело»
1	2	3	4
Учет российского законодательства	+	+	+
Совместимость с Windows	+	+	+
Импорт/экспорт данных в различные форматы	+	+	+
Масштабируемость	+	+	+
Высокая производительность	+	+	-
Быстродействие	+	+	-
Поддержка пользователя	+	+	-
Простота обслуживания	+	-	-
Повышенная отказоустойчивость	+	+	-
Единое централизованное хранилище документов	+	+	+
Выгрузка документов	+	+	+

Продолжение таблицы 1

1	2	3	4
Выгрузка отчета по найденным документам	+	+	+
Анализ и отчетность по просроченным документам	+	-	-
ЭЦП	+	+	+
Интегрирование с SAP	+	+	+
Мобильное согласование	+	-	-
Автоматическое уведомление информации о документе в электронную почту	+	+	-
Формирование документа по шаблону в СЭД	+	-	+
Визуализация бизнес-процессов	+	-	-
Выполнение делегирование и замещение	+	-	-
Итог	20	14	10

## 6. Выводы и заключение

В статье рассмотрены три системы электронного документооборота – «TESSA», «1С: Битрикс 24» и «Дело». Все три системы имеют обширный функционал, который позволяет решать задачи, связанные с процессом «Внутренний аудит». В результате анализа установлено, что система «TESSA» в большей степени соответствует требованиям Компании к СЭД и превосходит сравниваемые системы по нескольким требуемым критериям.

## ЛИТЕРАТУРА

1. Внутренние аудиты системы менеджмента [Электронный ресурс]. – Режим доступа: <https://quality.eup.ru/DOCUM/audit1.html> (дата обращения: 22.10.2019).
2. TESSA [Электронный ресурс]. – Режим доступа: <https://mytessa.ru/system/about/> (дата обращения: 22.10.2019).
3. Документооборот [Электронный ресурс]. – Режим доступа: <https://www.1c-bitrix.ru/products/cms/features/workflow.phphtml> (дата обращения: 22.10.2019).
4. Дело [Электронный ресурс]. – Режим доступа: [https://www.eos.ru/eos\\_products/eos\\_delo/](https://www.eos.ru/eos_products/eos_delo/) (дата обращения: 22.10.2019).

**Будзинская** Елизавета Олеговна, студентка, budzinskay-97@mail.ru

**Калиберда** Елена Анатольевна, к.т.н., доцент, каф. «Математические методы и информационные технологии в экономике», [elekaliberda@rambler.ru](mailto:elekaliberda@rambler.ru)

**Новиков** Артем Андреевич, студент, [novikov.art30@gmail.com](mailto:novikov.art30@gmail.com)

УДК 004.94: 65.011.56

## ВЫБОР ПРОГРАММНОГО ПРОДУКТА ДЛЯ АВТОМАТИЗАЦИИ ПРОЦЕССА УПРАВЛЕНИЯ ВУЗОМ

Ю. П. Галдина, К. В. Кравченко, И. М. Баркова

Омский государственный технический университет, Омск, Россия

**Аннотация:** Рассмотрены программные продукты для автоматизации деятельности учебных заведений: Магеллан, 1С: Университет ПРОФ и Infosuite. Проанализирован функционал данных систем, приведено обоснование выбора программного продукта для автоматизации управления бизнес-процессами вуза.

**Ключевые слова:** управление вузом, автоматизация деятельности учебных заведений.

## CHOOSING A SOFTWARE PRODUCTS FOR AUTOMATING THE UNIVERSITY MANAGEMENT

J. P. Galdina, K.V. Kravchenko, I. M. Barkova

Omsk State Technical University, Omsk, Russia

**Abstract:** Software products are reviewed for automation of educational institutions: Magellan, 1С: PROF University and Infosuite. The functionality of these systems was analyzed, rationale for the choice of software products for management automation of business processes of university was made.

**Keywords:** university management, automating the activities of educational institutions.

### 1. Введение

В силу развития тенденции к повсеместному внедрению информационных технологий и требований законодательства РФ в сфере высшего образования деятельность современных вузов сложно представить без автоматизации основных бизнес-процессов. Данная тема является актуальной, поскольку в вузе циркулирует большой поток документов, и оформление только некоторых типов документов на данный момент автоматизировано. Комплексную автоматизацию оформления документации можно по праву определить как способ управления корпоративными знаниями.

Таким образом, необходима комплексная информационная система управления вузом, которая позволила бы сократить трудозатраты сотрудников вуза на оформление документов и исключить дублирование функций исполнителей бизнес-процессов.

### 2. Результаты экспериментов

К основной деятельности вуза можно отнести, прежде всего, проведение приемной комиссии и организацию учебного процесса. Проведение приемной комиссии направлено на привлечение новых абитуриентов, которые после зачисления становятся студентами вуза. Организация учебного процесса должна

быть выстроена таким образом, чтобы максимально сохранить контингент обучающихся и повысить эффективность и качество обучения.

Рассмотрим процессы проведения приемной комиссии и организации учебного процесса на кафедре на примере Омского государственного технического университета более подробно.

Проведение приемной комиссии университета состоит из следующих этапов:

- прием документов абитуриентов;
- формирование личных дел абитуриентов;
- составление списка абитуриентов, рекомендованных к зачислению;
- анализ данных по подаче оригиналов документов;
- формирование приказов на зачисление;
- сдача личных дел в отдел кадров.

На текущий момент проведение приемной комиссии осуществляется в информационной системе «Абитуриент», которая является разработкой ОмГТУ. Данную систему тяжело обслуживать, поскольку при изменении правил приема на федеральном или вузовском уровне необходимо оперативно актуализировать функционал системы, что влечет за собой временные затраты.

Организация учебного процесса на кафедре включает:

- разработку рабочих программ дисциплин кафедры;
- проведение учебных занятий;
- распределение нагрузки;
- составление штатного расписания;
- подготовку приказов на практику;
- работу со студентами.

Некоторые перечисленные подпроцессы автоматизированы, но в разных информационных системах. Так, например, распределение учебной нагрузки, заполнение аттестационных ведомостей, а также ведение базы предприятий, предоставляющих студентам места для прохождения практики, реализовано в личном кабинете на портале ОмГТУ, а формирование и согласование приказов о направлении на практику осуществляется в системе 1С: Документооборот. Для того, чтобы повысить эффективность деятельности кафедр как основных подразделений, обеспечивающих учебный процесс, необходимо организовать их работу в единой информационной системе.

На сегодняшний день ИТ-рынок располагает большим выбором программных продуктов для автоматизации процессов управления вузом. Существуют специализированные решения для учебных заведений, которые изначально разрабатывались с учетом российского законодательства, а также специфики образовательной сферы. Ниже приведены системы автоматизации деятельности учебных заведений.

- «Магеллан» – платформа для управления учебным процессом. Состоит из семнадцати модулей, которые помогают автоматизировать работу со студентами и учебно-методическую деятельность [1].

- 1С: Университет ПРОФ – позволяет автоматизировать практически все процессы, которые связаны с учебно-методической деятельностью отделов и

деканатов, а также работу с контингентом студентов (от поступления до выпуска и трудоустройства) [2].

- Система «Infosuite. Управление образовательным учреждением» разработана на базе платформы «1С: Предприятие 8» и состоит из одиннадцати модулей. Система для управления процессами вуза [3].

В Табл. 1 подробно рассмотрен функционал вышеперечисленных программных продуктов.

**Таблица 1.** Сравнительный анализ функций ПО

Функционал	1С:Университет ПРОФ	Магеллан	Infosuite
Проведение и организация приемной компании	+	+	+
Разработка планов учебного процесса, поддержка ФГОС-3, распределение и расчет учебной нагрузки	+	+	-
Управление контингентом обучающихся	+	-	+
Формирование приказов и отчетности	+	-	+
Формирование регламентированной отчетности и таких документов, как форма ВПО-1, справки, дипломы, приложения	+	-	-
Учет трудоустройства выпускников	+	-	-
Назначение стипендий и отслеживание оплаты за обучение	+	+	+
Встроенный обмен данными с информационными системами: «GosInsp, ФИС ЕГЭ и приема», «ФРДО», ГИС, «Контингент»	+	-	-
Составление расписания учебных занятий	+	+	+
Автоматизация проведения итоговой государственной аттестации	+	-	-
Формирование регламентированной отчетности и документов	+	-	-
Управление научной деятельностью и инновациями	+	-	-
Управление послевузовским образованием и аттестацией научных кадров	+	-	-
Автоматизация деятельности диссертационных советов	+	-	-
Управление дополнительным образованием	+	-	-
Управление кампусом вуза	+	-	-
Личный кабинет для абитуриента, студента и преподавателя	+	-	+
Возможность доработки имеющегося функционала своими средствами	+	-	-

Анализируя функционал приведенных программных продуктов, следует отметить, что 1С: Университет ПРОФ обладает наиболее широким спектром возможностей в сравнении с другими системами. Еще одним преимуществом 1С:

Университет ПРОФ является интеграция с другими используемыми в вузе программными продуктами фирмы 1С.

### 3. Заключение

Таким образом, можно сделать вывод, что выбор системы «1С: Университет ПРОФ» позволит оптимизировать работу приемной комиссии и кафедр, обеспечивающих организацию учебного процесса. Внедрение данной системы приведет к сокращению временных затрат, которые возникают во время обработки большого объема документации.

### ЛИТЕРАТУРА

1. Магеллан. Система управления учебным процессом: URL: <https://magellanius.ru/> (дата обращения 15.04.2020).
2. 1С: Университет ПРОФ: URL: <https://solutions.1c.ru/catalog/university-prof> (дата обращения 16.04.2020).
3. Infosuite.Управление образовательным учреждением: URL: [http://www.tadviser.ru/index.php/Продукт:Infosuite.Управление\\_образовательным\\_учреждением](http://www.tadviser.ru/index.php/Продукт:Infosuite.Управление_образовательным_учреждением) (дата обращения 15.04.2020).

---

**Баркова** Ирина Михайловна, магистрант, barkowa.ira2018@yandex.ru

**Галдина** Юлия Павловна, магистрант, juliagaldina96@mail.ru

**Кравченко** Ксения Владимировна, старший преподаватель, каф. «Математические методы и информационные технологии в экономике», trr474747@mail.ru

УДК 004.9:65.011.56

## МОДЕЛИРОВАНИЕ БИЗНЕС-ПРОЦЕССОВ ДЛЯ СТАНДАРТИЗАЦИИ ДЕЯТЕЛЬНОСТИ КОМПАНИИ

*Е.В. Доценко, Е.Т. Гегечкори*

Омский государственный технический университет, Омск, Россия

**Аннотация.** *Моделирование бизнес-процессов для стандартизации деятельности компании на сегодняшний день является одним из эффективных инструментов оптимизации функционирования предприятия в соответствии с требованиями международной интегрированной системы менеджмента качества. Использование процессного подхода совместно с применением инструментов ARIS обеспечивает простую и качественную реализацию указанной задачи.*

**Ключевые слова:** *ARIS, графические инструменты, бизнес-процесс, информационные системы, интегрированная система менеджмента качества.*

## MODELING OF BUSINESS PROCESSES FOR STANDARDIZATION COMPANY ACTIVITIES

*E.V. Dotsenko, E.T. Gegechkori*

Omsk State Technical University, Omsk, Russia

**Abstract.** *Modeling of business processes to standardize the company's activities today is one of the effective tools for optimizing the operation of the enterprise in accordance with the requirements of an international integrated quality management system. The use of a process approach in conjunction with the use of ARIS tools provides a simple and high-quality implementation of this task.*

**Keywords:** *ARIS, graphic tools, business process, information systems, integrated quality management system.*

### 1. Введение

Ранние этапы проектов, которые направлены на реорганизацию бизнес-процессов организации и внедрение использования новейших информационных систем, часто сопровождают следующие вопросы, решение которых чрезвычайно важно именно в начале выполнения:

- 1) каких результатов с точки зрения улучшения деятельности организации можно добиться, используя технологии описания и реорганизации бизнес-процессов;
- 2) какое программное обеспечение использовать в проекте;
- 3) как моделировать процессы с использованием продукта «Х»;
- 4) как проводить анализ и выявлять проблемы при помощи продукта «Х»;
- 5) какую методологию использовать для описания процессов;
- 6) что делать дальше с полученными моделями бизнес-процессов.

На текущий момент среди аналитиков бизнеса все большую популярность приобретает инструментарий ARIS Toolset (ARIS). ARIS – архитектура интегрированных информационных систем, методология и тиражируемый программный

продукт для моделирования бизнес-процессов организаций. ARIS представляет собой векторный редактор, предназначенный для моделирования бизнес-процессов, в т.ч. для создания функциональных и поведенческих моделей информационных систем. Разработчиком методологии ARIS и одноименной среды моделирования является фирма IDS Scheer AG (Германия), в частности профессор А.В. Шеер. Модули программного продукта ARIS занимают ведущие места на мировом рынке в классе инструментов моделирования и анализа бизнес-процессов. Здесь все определено, разграничено и описано в спецификации, процесс работы формализован в набор алгоритмов [1]. Анализ деятельности предприятий и реорганизация бизнес-процессов – чрезвычайно сложная задача, требующая методической и инструментальной поддержки.

## 2. Постановка задачи

В любой организации, занимающейся выпуском стандартизированной продукции, должна быть разработана, задокументирована, внедрена и поддерживаться в рабочем состоянии интегрированная система менеджмента качества (ИСМ) в соответствии с требованиями международных стандартов, таких как, например, ISO 9001.

Принципы ИСМ, взятые за основу в процессе ее разработки и интеграции, являются результатом обобщения подходов, моделей, изложенных в вышеупомянутых международных стандартах.

Интегрированная система состоит из подсистем: менеджмента качества (СМК), менеджмента охраны здоровья и обеспечения безопасности труда (СМО-ЗиОБТ), экологического менеджмента (СЭМ). Все эти системы основаны на принципе предупреждения:

- предоставления клиенту несоответствующей продукции (ISO 9001);
- неблагоприятного воздействия на окружающую среду (ISO 14001);
- угрозы здоровью и безопасности работников (OHSAS 18001/ ISO 45001).

При разработке и интеграции ИСМ должны быть использованы следующие подходы.

– Процессный подход, рассматривающий совокупность взаимосвязанных ресурсов и видов деятельности организации, которые преобразуют входящие элементы в выходящие, образуя таким образом «модель» каждого действующего в компании процесса. На основе выделенных процессно-ориентированным образом элементов сформирована основа для интеграции действующих систем менеджмента в единую «Модель ИСМ», структурированную и управляемую как единый процесс.

– Подход, основанный на управлении рисками и возможностями, который позволяет использовать с выгодой положительные возможности и минимизировать вероятность возникновения нежелательных событий. Ключевым элементом такого подхода является принятие запланированных систематических мер для предупреждения любых потерь.



При процессном подходе в рамках ИСМ необходима стандартизация всех процессов, в том числе и процесса, который описывает разработку стандартов в рассматриваемой организации.

### 3. Теория

Модель ЕРС (процессно-событийная модель) предназначена для симуляции процессов, выполняемых в рамках одного подразделения несколькими подразделениями или конкретными сотрудниками [2].

Процессно-событийная модель:

- отражает последовательность функциональных действий (шагов) в рамках одного бизнес-процесса, которые выполняются организационными единицами, а также ограничения по времени, налагаемые на отдельные функции;
- позволяет выявлять связи между функциональной и организационной моделями.

**Модели процессов** представляют собой перечень основных и вспомогательных бизнес-процедур компании с их подробным описанием (взаимосвязи, цели, участники, и т.д.), а также со следующими пояснениями:

- событий, которые являются начальными для процесса, включая логику и условия выполнения начальных событий;
- функционала, который необходимо реализовать, с указанием участников и их ролей;
- связей «было-стало» между бизнес-процессами и иерархической структуры бизнес-процессов;
- данных на входе и выходе.

Модель процесса служит для описания алгоритма выполнения процесса последовательности функций, управляемых событиями. При этом каждая функция должна инициироваться и завершаться событием (или несколькими), иметь лицо, ответственное за ее выполнение и входную и выходную информацию. Одно событие может служить причиной выполнения одновременно нескольких функций, и, наоборот, одна функция может быть результатом наступления нескольких событий и функций. Для такого разветвленного описания процесса используются логические операторы («и», «или», «исключающие или»), а в модели ЕРС их называют *точками принятия решений* [3].

### 4. Результаты экспериментов

На примере организации, у которой реализуются процессы разработки стандартизированной продукции, требуется разработать модель типа ЕРС для процесса «Разработка стандарта» используя инструменты моделирования ARIS. Результат моделирования представлен на Рис. 1.

В данной процессной модели основное место уделяется четкой последовательности выполнения функционала. Для описания условий в модели имеются события и правила, которые могут описывать сложные алгоритмы выполнения процедур.

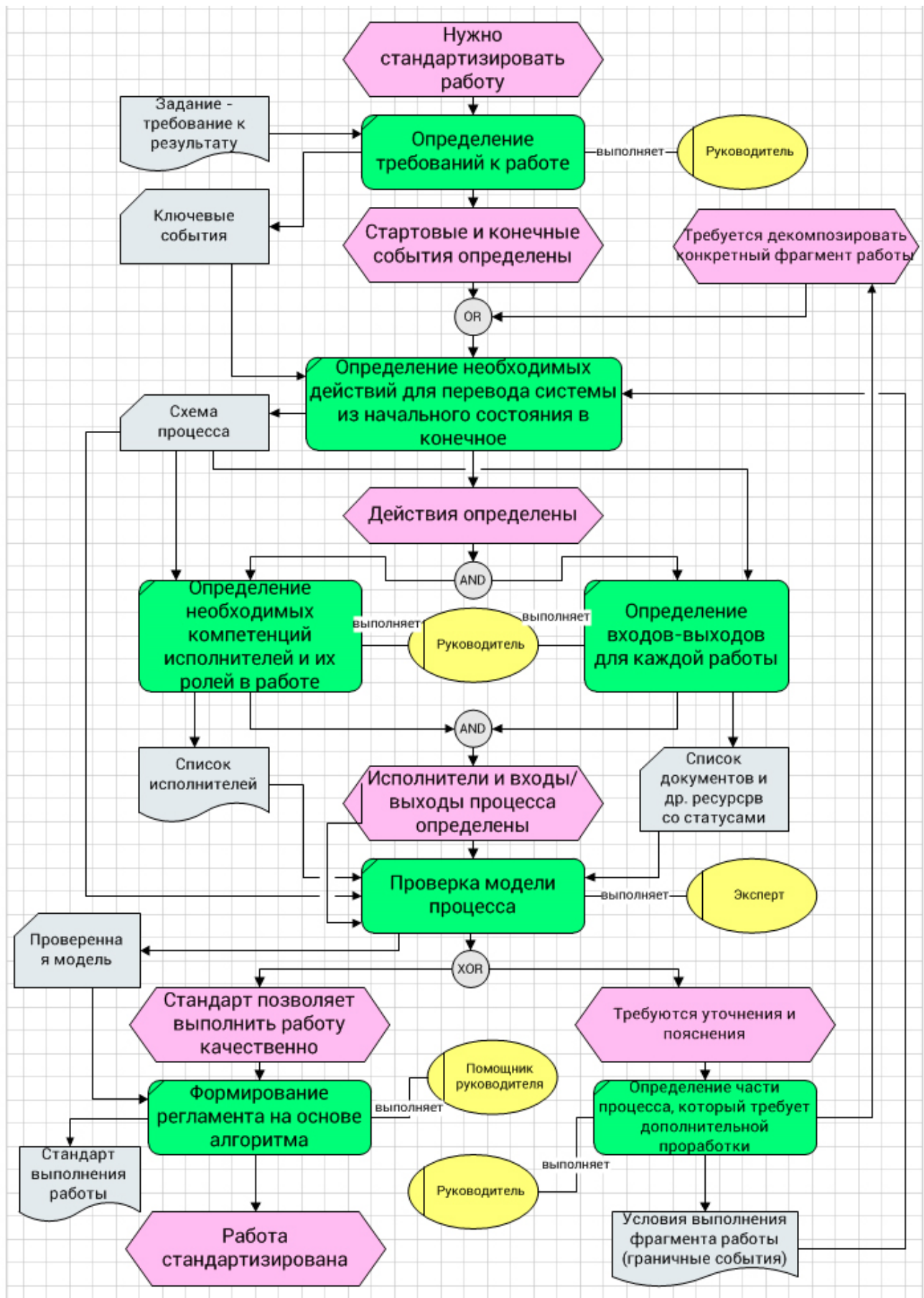


Рис. 1. Разработка стандарта предприятия в нотации EPC

Каждая модель процедуры непременно берет свое начало и заканчивается одним или несколькими событиями или интерфейсами в другие модели процедуры. Если модель начинается с интерфейса, то после него должно идти завершающее событие предшествующего процесса. Если модель процесса заканчивается интерфейсом в другой процесс, то ему должно предшествовать конечное событие текущего процесса.

Одним из основных преимуществ методологии ARIS являются эргономика и качественная, подробная визуализации разрабатываемых моделей бизнес-процессов, что позволяет данной методологии быть легкой в использовании и понятной каждому работнику самой высоко иерархичной компании, включающей в свою организационную структуру менеджеров различных уровней и простых исполнителей.

ЕРС модель используется для описания алгоритма выполнения процедур в виде последовательности функций, управляемых событиями. Каждая ЕРС модель обязательно начинается и заканчивается одним событием или их набором, так же она может заканчиваться интерфейсами в другие процедуры. Главное внимание здесь уделяется алгоритму выполнения действий – функций. Для описания условий в модели имеются события и правила, которые могут описывать сложные алгоритмы выполнения процедур [1].

## **5. Обсуждение результатов**

С помощью модели, представленной на Рис. 1, специалист, которому требуется разработать стандарт выполнения той или иной работы, сможет это сделать по упрощенному алгоритму. А специалист отдела качества и стандартизации, который исполняет и контролирует политику в области менеджмента качества организации, сможет проконтролировать выполнение стандарта организации. Подобная модель процесса является частью инструментария по стандартизации процессов работы той или иной компании.

## **6. Выводы и заключение**

ARIS в современной компании, которая выбрала своим правилом соблюдение международных стандартов качества, является удобным и достаточно простым инструментом. Методология графического представления информации составляет основу наглядного и качественного использования всех модулей этой программы. С помощью ARIS можно выводить отчетные документы, структурировать по пользовательским правилам множеству признаков аналитические отчеты о работе систем организации. Анализ рисков с использованием модуля ARIS будет организован наглядно, классифицировано, структурировано и тем самым, более работоспособно.

Оптимизация любых видов деятельности обеспечивается на основе графических средств представления информации, что обеспечивает простоту контроля и формирования более подробного представления процессов организации.

## ЛИТЕРАТУРА

1. *Морозова В.И., Врублевский К.Э.* Моделирование бизнес-процессов с использованием методологии ARIS: учебно-методическое пособие – М.: РУТ (МИИТ), 2017. – 47 с.
2. *Калянов Г.Н.* Теория и практика реорганизации бизнес-процессов. Серия «Реинжиниринг бизнеса». – М.: СИНТЕГ, 2000.
3. *Репин В., Маклаков С.* ARIS Toolset/BPwin: выбор за аналитиком. – <http://www.compress.ru/Temp/2878/index.htm>.

---

**Доценко** Елена Валерьевна, магистрант ОмГТУ гр. ПИМ-192, [iuneva.nova@gmail.com](mailto:iuneva.nova@gmail.com)

**Гегечкори** Евгений Трдатович, к.т.н., доцент, каф. «Математические методы и информационные технологии в экономике», [getnbn@inbox.ru](mailto:getnbn@inbox.ru)

## ОЦЕНКА ТОЧНОСТИ ПРИБЛИЖЕННЫХ ФОРМУЛ ДЛЯ СРЕДНЕЙ ДЛИНЫ ОЧЕРЕДИ

В. Н. Задорожний<sup>1</sup>, Ю. Г. Галич<sup>2</sup>

<sup>1</sup>Омский государственный технический университет, Омск, Россия

<sup>2</sup>Омский государственный университет путей сообщения, Омск, Россия

**Аннотация.** Выполняется сравнительный анализ приближенных формул для расчета одной из важнейших характеристик систем массового обслуживания – зависимости средней длины очереди от коэффициента загрузки. Погрешности приближенных формул определяются с помощью точных формул (где это возможно) или посредством имитационного моделирования. Наилучшие из рассматриваемых приближенных формул выявляются путем их сравнения по точности расчета искомой зависимости. Актуальность исследования определяется отсутствием точных формул, позволяющих рассчитывать эту зависимость для ряда распространенных систем массового обслуживания. Кроме того, имеющиеся для некоторых систем точные формулы порождают слишком сложные вычислительные процедуры, реализация которых требует затрат, не оправдываемых достигаемой точностью решений, излишней во многих прикладных задачах.

**Ключевые слова:** системы массового обслуживания, зависимость средней длины очереди от коэффициента загрузки, приближенные расчетные формулы, имитационное моделирование систем с очередями.

## ACCURACY ESTIMATE OF APPROXIMATE FORMULAE FOR AVERAGE QUEUE LENGTH

V. N. Zadorozhnyi<sup>1</sup>, Y. G. Galich<sup>2</sup>

<sup>1</sup>Omsk State Technical University, Omsk, Russia

<sup>2</sup>Omsk State Transport University, Omsk, Russia

**Abstract.** A comparative analysis of approximate formulas for calculating one of the most important characteristics of Queueing systems - the dependence of the average queue length on the load factor. Errors in approximate formulas are determined using exact formulas (where possible) or by simulation. The best of the considered approximate formulas are revealed by comparing them according to the accuracy of calculating the desired dependence. The relevance of the study is determined by the lack of precise formulas that allow calculating this dependence for a number of common Queueing systems. In addition, the exact formulas available for some systems give rise to too complex computational procedures, the implementation of which requires costs that are not justified by the accuracy of solutions that are unnecessary in many applications.

**Keywords:** queueing systems, dependence of the average queue length on the load factor, simulation modeling of queueing systems.

### 1. Введение

Теория массового обслуживания (ТМО) как область научных исследований возникла в связи с необходимостью анализа процессов образования очередей. Многие технические, экономические и социальные задачи связаны с системами,

предназначенными для обслуживания некоторых потоков объектов (заявок), поступающих в случайные моменты времени [1–3]. Средняя длина очереди  $L$  – это один из важнейших показателей систем массового обслуживания (СМО). Многие технико-экономические показатели эффективности функционирования СМО выражаются через среднюю длину очереди. Так, среднее время  $W$  ожидания обслуживания выражается через  $L$  и известную интенсивность  $\lambda = 1/\bar{\tau}$  входящего потока простой формулой Литтла  $W = L/\lambda = L\bar{\tau}$  [4], где  $\bar{\tau}$  – средняя длительность интервала поступления заявок. Среднее время  $U$  пребывания заявки в системе выражается через среднее время ожидания и известное среднее время  $\bar{x}$  обслуживания как сумма  $U = W + \bar{x} = L/\lambda + \bar{x}$ , а среднее число  $N$  заявок в системе – как сумма средней длины очереди и среднего числа заявок в  $m$  каналах системы:  $N = L + m\rho$ , где  $\rho = \lambda\bar{x}/m$  – коэффициент загрузки. При  $\rho < 1$ , т.е. в стационарном режиме функционирования системы класса GI|GI|m (с  $m \geq 1$  каналами), перечисленные характеристики – величины постоянные, не зависящие от времени.

При проектировании технических и организационно-технических систем, математическими моделями которых являются СМО или сети СМО, для оптимального выбора структуры и параметров систем приходится решать уравнения или неравенства, содержащие зависимости характеристик СМО от их параметров. Эти зависимости часто сводятся к зависимостям средних длин очередей от коэффициентов загрузки. Поэтому нас в данной статье интересуют приближенные формулы, позволяющие достаточно точно рассчитывать не точечные оценки средних длин очередей  $L$  при фиксированных параметрах СМО, а зависимости  $L(\rho)$  в достаточно широком диапазоне значений  $\rho$ , близком ко всей области стационарности  $0 \leq \rho < 1$ . Поэтому зависимость  $L(\rho)$  рассматривается далее как основная интересующая нас характеристика, для расчета которой подбираются подходящие приближенные формулы.

При имитационном моделировании (ИМ) СМО или сетей СМО численное решение задач оптимизации, содержащих более двух-трех варьируемых параметров, без использования градиентных методов практически невозможно. Однако непосредственное ИМ СМО и сетей СМО не позволяет эффективно использовать градиентные методы из-за наличия статистических погрешностей у имитационных оценок целевых функций (известная проблема градиентов в ИМ, [5, 6]). Поэтому одной из наиболее актуальных областей применения приближенных формул является разработка ускоренных аналитико-имитационных методов оптимизации. На первом этапе аналитико-имитационной оптимизации с помощью приближенных формул удастся быстро найти градиентными методами значения варьируемых параметров, близкие к оптимальным. Часто точность такого приближенного решения и достигаемый благодаря ему эффект бывают достаточными для непосредственного применения этого решения на практике. В противном случае полученное приближенно оптимальное решение может быть достаточно быстро – благодаря его близости к оптимальному решению – уточнить непосредственным ИМ, с помощью, например, покоординатных методов оптимизации с контролем точности и с применением ускоренного метода общих случайных чисел [6–8].

## 2. Анализ некоторых известных приближений для $L(\rho)$

**1.** Рассмотрим стационарную одноканальную СМО GI|GI|1, в которую поступает рекуррентный поток заявок с интенсивностью  $\lambda$ . Интервалы времени между заявками – независимые случайные величины (с.в.) с функцией распределения (ф.р.)  $A(t)$ , средним значением  $\bar{\tau} = 1/\lambda$  и коэффициентом вариации  $C_\tau$ . Время  $x$  обслуживания заявки в системе (также независимая с.в.) имеет произвольную ф.р.  $B(t)$  и коэффициент вариации  $C_x$ . Интенсивность обслуживания заявок  $\mu = 1/\bar{x}$ . Поскольку число каналов в системе  $m = 1$ , то ее коэффициент загрузки  $\rho = \lambda\bar{x}/m = \lambda\bar{x} = \bar{x}/\bar{\tau} = \lambda/\mu$ .

Точного аналитического выражения средней длины очереди через параметры СМО, применимого для всех СМО класса GI|GI|1, не найдено, но известны различные приближенные формулы. Например, можно указать нижнюю и верхнюю границы средней длины очереди. Доказано [4, 9], что

$$\frac{\rho^2 C_x^2}{2(1-\rho)} - \frac{\rho}{2} = L_{\text{нижн.1}} \leq L \leq L_{\text{верх.1}} = \frac{C_\tau^2 + \rho^2 C_x^2}{2(1-\rho)}. \quad (1)$$

Здесь нижняя граница получена эквивалентным преобразованием приведенной в [9] на с. 134 формулы, имеющей в [9] номер (2.60). Доказательство этой формулы в [9] не приводится и ссылок на литературу, содержащую такое доказательство, не дается. Тем не менее, многочисленные проверки, выполненные нами методом ИМ, подтверждают правильность этой формулы. В [4] доказывается другая нижняя оценка, которая хуже (меньше) использованной в (1) оценки на  $\rho/2/(1-\rho)$ . Верхняя граница – второе неравенство в (1) – известно как формула Кингмана-Келлерстрема [4].

**2.** Для случая эрланговского входящего потока (т.е. для систем  $E^k|GI|1$ ) в [9] сообщаются как доказанные следующие границы:

$$\frac{\rho^2 (C_\tau^2 + C_x^2)}{2(1-\rho)} - \frac{\rho(1-C_\tau^2)}{2} = L_{\text{нижн.2}} \leq L \leq L_{\text{верх.2}} = \frac{\rho^2 (C_\tau^2 + C_x^2)}{2(1-\rho)} + \frac{(1-\rho)(1-C_\tau^2)}{2}, \quad (2)$$

где  $C_\tau^2 = 1/k$ ,

$k$  – порядок распределения Эрланга ( $k = 1, 2, \dots$ ).

**3.** Также известна приближенная формула, предложенная для оценки средней длины очереди в системах GI|GI|1 М. А. Файнбергом [9, 10]:

$$L_{\text{прибл.}} = \frac{\rho^2 (C_\tau^2 + C_x^2)}{2(1-\rho)}. \quad (3)$$

Приближение (3) лежит как между границами (1), так и между границами (2).

Если же задающее входящий поток распределение является распределением Эрланга порядка  $k = 1$ , т.е. экспоненциальным распределением, то СМО  $E^k|GI|1$  становится системой M|G|1 с коэффициентом вариации  $C_\tau = 1$ . В этом частном

случае левая и правая части неравенств (2) совпадают, и формула (2) превращается в точную формулу Поллачека-Хинчина [11], выведенную для систем  $M|G|1$ :

$$L = \frac{\rho^2(1 + C_x^2)}{2(1 - \rho)}, \quad (4)$$

что, несомненно, является определенным достоинством верхней границы в формуле (2): эта граница достижима и становится точной для систем  $M|G|1$ , широко распространенных в практических приложениях ТМО.

Заметим, что в этом частном случае приближение Файнберга (3) также превращается в точную формулу Поллачека-Хинчина.

**4.** В то же время известно, что приближение Кингмана-Келлерстрема – верхняя граница в (1) – при  $\rho \uparrow 1$  для любых систем  $GI|GI|1$  (и, более того, для любых систем  $GI|GI|m$ ) с конечными коэффициентами вариациями  $C_\tau$  и  $C_x$  сходится к точному решению. Число  $m$  каналов в выражении этой границы учитывается определением входящего в него коэффициента загрузки  $\rho = \lambda \bar{x} / m$ . Последнее замечание позволяет поставить вопрос о том, нельзя ли распространить при  $\rho \uparrow 1$  на многоканальные системы и приближение Файнберга (2).

Для ответа на этот вопрос найдем разность между приближениями Кингмана-Келлерстрема и Файнберга:

$$\frac{C_\tau^2 + \rho^2 C_x^2}{2(1 - \rho)} - \frac{\rho^2 (C_\tau^2 + C_x^2)}{2(1 - \rho)} = \frac{C_\tau^2}{2}.$$

Из того, что разность сравниваемых приближений постоянна и не зависит от числа каналов  $m$ , и того, что при  $\rho \uparrow 1$  оба сравниваемых приближения сходятся к бесконечности, следует, что приближение Файнберга (3) можно использовать и для многоканальных систем, и что, соответственно, при  $\rho \uparrow 1$  относительная погрешность приближения Файнберга сходится к нулю.

**5.** Привлекательным приближение (3) делает еще одно соображение. Поскольку эрланговский поток является частным случаем «гамма потока», задаваемого гамма распределением интервалов поступления заявок, то можно ожидать, что приближение (3) хорошо подойдет и для систем  $\Gamma|GI|1$ . Исследование этой возможности целесообразно в связи с тем, что гамма распределение часто используется на практике для аппроксимации широкого класса распределений по эмпирическим данным. Это обусловлено тем, что гамма распределение позволяет реализовать сразу два параметра эмпирических распределений – математическое ожидание и дисперсию, и при этом не связано ограничением коэффициента вариации величинами, не превосходящими единицы (как распределение Эрланга). Гамма распределение может иметь сколь угодно большой положительный коэффициент вариации.

### 3. Задача оценки точности приближений

В СМО  $GI|GI|1$  и  $GI|GI|m$  для ряда распространенных на практике ф.р.  $A(t)$ ,



$B(t)$ , описывающих длительности интервалов поступления заявок и время их обслуживания соответственно, требуется оценить и сравнить точность нескольких известных приближенных формул для  $L(\rho)$ . В число ф.р.  $A(t)$  и  $B(t)$  входят экспоненциальное, равномерное, смещенное равномерное распределения, распределение Эрланга, гамма распределение и распределение Парето. Для расчета истинной средней длины очереди при различных значениях  $\rho$  необходимо использовать точные формулы (где это возможно) или ИМ. Оценку точности приближенных выражений  $L^*(\rho)$  характеристики  $L(\rho)$  следует произвести отдельно для различных классов СМО, поскольку в одних классах СМО могут оказаться более точными одни приближения, в других – другие.

**Примечание.** Поскольку имитационные расчеты истинной зависимости  $L(\rho)$  можно выполнить лишь для набора определенных значений  $\rho$ , можно погрешности приближений  $L^*(\rho)$  оценивать в диапазоне  $0,1 \leq \rho \leq 0,9$  с шагом по  $\rho$ , равным 0,1. И для того, чтобы итоговая оценка погрешности какого-либо приближения  $L^*(\rho)$  выражала его обобщенную по параметру  $\rho$  точность, ее можно определить как пару  $(\Delta; \delta)$  следующих двух нормированных оценок: оценки  $\Delta$  абсолютной погрешности и оценки  $\delta$  относительной погрешности:

$$\Delta = \frac{1}{9} \sum_{i=1}^9 |L^*(\rho_i) - L(\rho_i)|, \quad \delta = \frac{1}{9} \sum_{i=1}^9 \frac{|L^*(\rho_i) - L(\rho_i)|}{L(\rho_i)}, \quad (5)$$

где  $L(\rho_i)$  и  $L^*(\rho_i)$  – истинное и, соответственно, приближенное значения средней длины очереди в точке  $\rho_i = i/10$ .

Следует также иметь в виду, что если значение  $L(\rho_i)$  вычисляется посредством ИМ, то его можно использовать в (5) как истинное лишь при условии, что оно вычислено с погрешностью, модуль которой значительно меньше, чем определяемая через него погрешность  $|L^*(\rho_i) - L(\rho_i)|$  оценки  $L^*(\rho_i)$ . Это требует выполнения достаточно длинных прогонов имитационной модели СМО.

#### 4. Оценка точности приближений

##### 4.1. Оценка точности приближений для систем класса GI/M/1

Для систем GI/M/1 известна точная формула средней длины очереди [4]:

$$L = \frac{\rho\sigma}{1 - \sigma}, \quad (6)$$

где  $\sigma$  – единственный в области  $0 \leq \sigma < 1$  корень уравнения:

$$\sigma = A^*(\mu - \mu\sigma), \quad (7)$$

$A^*(s) = \int_0^\infty e^{-st} a(t) dt$  – преобразование Лапласа-Стилтьеса (ПЛС) функции  $a(t)$ ,  $a(t)$

– плотность распределения вероятностей (п.в.) интервалов поступления заявок:  $a(t) = dA(t)/dt$ . Формулы (6), (7) не позволяют получить явное аналитическое

выражение  $L$  через известные параметры СМО GI|M|1 для произвольной ф.р.  $A(t)$ , задающей входящий поток, однако позволяют для любой  $A(t)$  находить точное численное значение  $L$  и тем самым проверять точность приближений характеристики  $L(\rho)$  для систем GI|M|1.

Для многоканальной СМО GI|M|m,  $m > 1$  основным параметром также является  $\sigma$ , – единственное в области  $0 \leq \sigma < 1$  решение уравнения:

$$\sigma = A^*(m\mu - m\mu\sigma). \quad (8)$$

Для вычисления средней длины очереди используют формулу:

$$L = \frac{J\rho\sigma}{m(1-\sigma^2)}, \quad (9)$$

где значение коэффициента  $J$  вычисляется путем выполнения сложной рекуррентной процедуры, описанной в [4], включающей шаги решения уравнений и систем уравнений с интегралами и параметрами, зависящими от  $m$  и от ф.р.  $A(t)$ . Вычисление коэффициента  $J$  в формуле (9) настолько сложно, что на практике, как правило, для расчета  $L(\rho)$  в системах GI|M|m при  $m > 1$  предпочитают использовать ИМ.

Рассчитаем среднюю длину очереди в некоторых СМО класса GI|M|1 по точной и приближенным формулам, а также посредством ИМ. Интервалы времени обслуживания в этих системах имеют экспоненциальное распределение с п.в.  $a(t) = \mu e^{-\mu t}$ , среднее время обслуживания  $\bar{x} = 1/\mu$ , дисперсия  $D_x = 1/\mu^2$ , квадрат коэффициента вариации  $C_x^2 = D_x / \bar{x}^2 = 1$ . Для демонстрации способов расчета характеристики  $L(\rho)$  возьмем  $\rho = 0,6$ .

#### 4.1.1. Оценка точности приближений для системы R|M/1

В системе R|M|1 интервалы поступления заявок имеют равномерное распределение, т.е. п.в.  $a(t) = 1/h$ ,  $0 \leq t \leq h$ , где  $h$  – параметр распределения. Поэтому средняя длительность интервала поступления  $\bar{\tau} = h/2$ , дисперсия  $D_\tau = h^2/12$ , квадрат коэффициента вариации  $C_\tau^2 = D_\tau / \bar{\tau}^2 = 1/3$ , интенсивность входящего потока  $\lambda = 1/\bar{\tau} = 2/h$ . В соответствии с (7) функция  $A^*(s)$  определяется следующим образом:

$$A^*(s) = \int_0^h e^{-st} \frac{1}{h} dt = -\frac{e^{-st}}{sh} \Big|_0^h = \frac{1 - e^{-sh}}{sh}. \quad (10)$$

Пусть  $h = 20$ ,  $\mu = 1/6$  (тогда  $\rho = \lambda/\mu = (2/h) \cdot 6 = 0,6$ ). Уравнение (7) для рассматриваемой СМО с учетом (10) имеет вид:

$$\sigma = \frac{1 - e^{-(\mu - \mu\sigma)h}}{(\mu - \mu\sigma)h} = \frac{1 - e^{-\mu h(1-\sigma)}}{\mu h(1-\sigma)} \quad \text{или} \quad \sigma = \frac{1 - e^{-(10/3)(1-\sigma)}}{(10/3)(1-\sigma)}. \quad (11)$$

Решение уравнения (11) численным методом дает корень  $\sigma = 0,468403\dots$ . Тогда по формуле (6) получаем  $L = 0,528674\dots \approx 0,529$ .

По приближенным формулам (1), (2) и (3) находим три оценки  $0,150 \leq L \leq 0,867$ ;  $0,400 \leq L \leq 0,733$  и  $L_{\text{прибл.}} = 0,600$  соответственно. Обе границы (2) в данном частном случае, как и в системе  $R|M|1$ , оказались более точными, чем соответствующие границы (1).

В результате ИМ этой СМО при длине прогона 500 тыс. заявок для  $\rho$  и  $L$  получены оценки 0,600 и 0,529 соответственно. В данном частном случае имитационная оценка для  $L(\rho) = L(0,6) = 0,529$  может использоваться для сравнения точности оценок (2) и (3) в качестве истинного значения  $L(\rho)$ .

#### 4.1.2. Оценка точности приближений для системы $E^2|M|1$

В системе  $E^2|M|1$  интервалы поступления заявок имеют распределение Эрланга второго порядка, т. е. п.в.  $a(t) = \lambda^2 t e^{-\lambda t}$ , где  $\lambda$  – параметр распределения. Известно, что распределение Эрланга  $k$ -го порядка имеет с.в., равная сумме  $k$  независимых экспоненциальных с.в. с одним и тем же параметром  $\lambda$ . Поэтому в системе  $E^2|M|1$  средняя длительность интервала поступления  $\bar{\tau} = 1/\lambda + 1/\lambda = 2/\lambda$ , дисперсия  $D_{\tau} = 1/\lambda^2 + 1/\lambda^2 = 2/\lambda^2$ , квадрат коэффициента вариации  $C_{\tau}^2 = D_{\tau} / \bar{\tau}^2 = 1/2$ , интенсивность входящего потока  $\lambda = 1/\bar{\tau} = \lambda/2$ . При этом ПЛС  $A^*(s)$  определяется следующим образом:

$$A^*(s) = \int_0^{\infty} e^{-st} \lambda^2 t e^{-\lambda t} dt = \lambda^2 \int_0^{\infty} t e^{-(\lambda+s)t} dt = \lambda^2 \left( -\frac{t e^{-(\lambda+s)t}}{\lambda+s} - \frac{e^{-(\lambda+s)t}}{(\lambda+s)^2} \right) \Bigg|_0^{\infty} = \frac{\lambda^2}{(\lambda+s)^2}.$$

При такой ПЛС  $A^*(s)$ , полагая  $\lambda = 2/5$ ,  $\mu = 1/3$ , решаем (7) численным методом и находим  $\sigma = 0,495841\dots$ . Поскольку  $\lambda = \lambda/2 = 1/5$ , имеем  $\rho = \lambda/\mu = 0,6$  и по формуле (6)  $L = 0,590101\dots$ . Меняя значения  $\mu$  в формуле (7), рассчитываем зависимость  $L$  от  $\rho$  (второй столбец Табл. 1) с 6-ю точными знаками после запятой.

По приближенным формулам (1), (2) и (3) имеем  $0,150 \leq L \leq 1,075$ ;  $0,525 \leq L \leq 0,775$  и  $L_{\text{прибл.}} = 0,675$  соответственно. Меняя значения  $\rho$  в этих формулах получаем зависимости  $L_{\text{нижн.1}}$ ,  $L_{\text{верх.1}}$ ,  $L_{\text{нижн.2}}$ ,  $L_{\text{верх.2}}$ ,  $L_{\text{прибл.}}$  от  $\rho$ , скорректированные в области их отрицательных значений (левые части неравенств (1) и (2) при малых значениях  $\rho$  отрицательны, поэтому в Табл. 1 приведены скорректированные значения  $L'_{\text{нижн.1}} = \max\{0; L_{\text{нижн.1}}\}$  и  $L'_{\text{нижн.2}} = \max\{0; L_{\text{нижн.2}}\}$ .

Имитационная модель СМО  $E^2|M|1$  на языке GPSS имеет вид:

GENERATE	(Gamma (1, 0, 2.5, 2))	;Генерация заявок входящего потока
QUEUE	1	;Регистрация входа заявок в очередь
SEIZE	1	;Заявка занимает канал
DEPART	1	;Выход заявки из очереди
ADVANCE	(Exponential (1, 0, 3))	;Обслуживание заявки
RELEASE	1	;Заявка освобождает канал
TERMINATE		;Заявка покидает систему
** Таймер модели **		
GENERATE	50000000	;Воспроизводимое время СМО
TERMINATE	1	;Транзакт останавливает процесс ИМ
START	1	

Здесь время обслуживания заявки задано в блоке ADVANCE как экспоненциально распределенная с.в. со средним значением 3. Интервал поступления заявок определен в блоке GENERATE как с.в., имеющая распределение Эрланга второго порядка со средним значением 5, которое реализовано посредством библиотечного генератора с.в., имеющих гамма-распределение. В имитационном эксперименте система наблюдалась в течение  $5 \cdot 10^7$  ед. времени, за которое через нее проходит в среднем  $10^7$  заявок. В результате для  $L$  получена имитационная оценка 0,591 (Табл. 1).

**Таблица 1.** Расчет  $L(\rho)$  и соответствующих оценок для СМО  $E^2|M|1$

$\rho$	$L$	$L_{\text{им}}$	$L'_{\text{нижн.1}}$	$L_{\text{верх.1}}$	$L'_{\text{нижн.2}}$	$L_{\text{верх.2}}$	$L_{\text{прибл.}}$
0,1	0,003006	0,003	0,000	0,283	0,000	0,233	0,008
0,2	0,020696	0,021	0,000	0,338	0,000	0,238	0,038
0,3	0,064984	0,065	0,000	0,421	0,021	0,271	0,096
0,4	0,151956	0,152	0,000	0,550	0,100	0,350	0,200
0,5	0,309017	0,308	0,000	0,750	0,250	0,500	0,375
0,6	0,590101	0,591	0,150	1,075	0,525	0,775	0,675
0,7	1,120437	1,119	0,467	1,650	1,050	1,300	1,225
0,8	2,275184	2,293	1,200	2,850	2,200	2,450	2,400
0,9	5,929461	5,878	3,600	6,550	5,850	6,100	6,075

Меняя параметры блока ADVANCE, получаем зависимость  $L_{\text{им}}(\rho)$  (третий столбец в Табл. 1). Заметим, что относительные погрешности имитационных оценок  $L$  не превосходят 1-2%. В целом согласованность рассчитанных по точным формулам значений  $L$  и их имитационных оценок  $L_{\text{им}}$  подтверждает правильность выполненных вычислений.

#### 4.1.3. Оценка точности приближений для систем $E^k/M/1$

В системах  $E^k|M|1$  интервалы поступления заявок имеют распределение Эрланга  $k$ -го порядка с параметром  $\dot{\lambda}$ . Средняя длительность интервала поступления  $\bar{\tau} = k/\dot{\lambda}$ , дисперсия  $D_{\tau} = k/\dot{\lambda}^2$ , квадрат коэффициента вариации  $C_{\tau}^2 = D_{\tau}/\bar{\tau}^2 = 1/k$ , интенсивность входящего потока  $\lambda = 1/\bar{\tau} = \dot{\lambda}/k$ .

При этом функция  $A^*(s)$  как ПЛС суммы  $k$  независимых одинаково распределенных с.в. равна  $k$ -й степени ПЛС слагаемых:

$$A^*(s) = \frac{\dot{\lambda}^k}{(\dot{\lambda} + s)^k}.$$

Пусть  $k = 3$ . Тогда корнем уравнения (7) при  $\dot{\lambda} = 1/5$ ,  $\mu = 1/9$ , ( $\rho = 0,6$ ) является  $\sigma = 0,448549\dots$ , и по формуле (6) получаем  $L = 0,488039\dots$ . По приближенным формулам (1), (2) и (3) получаем оценки  $0,150 \leq L \leq 0,867$ ;  $0,400 \leq L \leq 0,733$  и  $L_{\text{прибл.}} = 0,600$  соответственно. Вновь обе границы приближения (2) более точны, чем границы приближения (1).

#### 4.1.4. Предварительные выводы по классу систем $GI|M|1$

1. Приближения (1)–(3) в разных системах класса  $GI|M|1$  различаются только параметром  $C_\tau^2$ . Поэтому если параметр  $C_\tau^2$  у нескольких разных систем рассматриваемого класса совпадает, то совпадают и используемые для них приближения (1)–(3). Например, характеристика  $L(\rho)$  будет описываться одними и теми же приближениями у системы  $R|M|1$  и у системы  $E^3|M|1$ . Но зависимости  $L(\rho)$  у этих двух систем разные. Например, как рассчитано выше, в первой из них  $L(0,6) = 0,528674$ , во второй  $L(0,6) = 0,488039$ . Поэтому рассматриваемые приближения (1)–(3) с точки зрения погрешностей (5) должны подходить к одной из них больше, чем к другой.

2. Интервал (2) для некоторых систем  $E^k|M|1$  имеет более точные границы, чем интервал (1).

Действительно, показатель погрешности  $(\Delta; \delta)$ , вычисленный по формулам (5) для системы  $E^2|M|1$  по данным Табл. 1, составляет для границы  $L'_{\text{нижн.1}}$  пару  $(0,561; 0,799)$ , а для границы  $L'_{\text{нижн.2}}$  – пару  $(0,052; 0,381)$ , которая заметно лучше, так как погрешности 0,052 и 0,381 меньше соответствующих погрешностей 0,561 и 0,799. Точно так же и верхняя граница  $L_{\text{верх.2}}$  имеет показатель погрешности  $(\Delta; \delta) = (0,195; 10,3)$  заметно лучший, чем граница  $L_{\text{верх.1}}$ , у которой этот показатель составляет  $(0,445; 13,3)$ . Преимущество границ (2) перед границами (1) для системы  $E^2|M|1$  обусловлено тем, что нижняя граница в (2) на всем диапазоне  $0,1 \leq \rho \leq 0,9$  лучше (выше) нижней границы в (1), и, одновременно, верхняя граница в (2) лучше (ниже) верхней границы в (1) (см. Табл. 1). Заметим, что обе сравниваемые границы доказаны, т.е. верны.

3. Однако такое соотношение границ (2) и (1) имеет место не для всех  $k$ . Если в системе  $E^k|M|1$  порядок  $k$  распределения Эрланга возрастает, то верхняя граница в (2) при некоторых  $\rho$  становится хуже верхней границы в (1).

Так, вычитая из верхней границы (1) верхнюю границу (2) при  $C_\tau^2 = 1/k$ , находим, что разница

$$\begin{aligned} L_{\text{верх.1}} - L_{\text{верх.2}} &= \frac{C_\tau^2 + \rho^2 C_x^2}{2(1-\rho)} - \frac{\rho^2 (C_\tau^2 + C_x^2)}{2(1-\rho)} - \frac{(1-\rho)(1-C_\tau^2)}{2} = \\ &= C_\tau^2 - \frac{(1-\rho)}{2} = \frac{1}{k} - \frac{(1-\rho)}{2} \end{aligned}$$

положительна лишь при  $\frac{k-2}{k} < \rho < 1$ . Таков диапазон значений  $\rho$ , в котором оценка  $L_{\text{верх.2}}$  точнее оценки  $L_{\text{верх.1}}$ . С ростом порядка распределения Эрланга этот диапазон сужается и его размер, равный  $1 - \frac{k-2}{k} = \frac{2}{k}$ , с ростом  $k$  сходится к нулю. Что касается нижней границы, то в (2) она при любых  $\rho$  не хуже, чем в (1), поскольку превышает нижнюю границу в (1) на неотрицательную величину

$$L_{\text{нижн.2}} - L_{\text{нижн.1}} = \frac{\rho^2 (C_\tau^2 + C_x^2)}{2(1-\rho)} - \frac{\rho(1-C_\tau^2)}{2} - \frac{\rho^2 C_x^2}{2(1-\rho)} + \frac{\rho}{2} = \rho C_\tau^2 / 2 / (1-\rho).$$

4. Ухудшение с ростом  $k$  верхней границы  $L_{\text{верх.2}}$  позволяет предположить, что с уменьшением  $k$  она улучшается, и вместе с этим улучшается приближение (3). Наибольший интерес представляет диапазон значений  $k < 1$ , реализуемый уже не распределением Эрланга, а обобщающим его гамма распределением. При вещественных положительных  $k < 1$  гамма распределение имеет коэффициент вариации  $C^2 > 1$ .

#### 4.2. Оценка точности приближений для систем класса GI/GI/1

1. В системе RT|R|1 интервалы поступления заявок имеют смещенное равномерное распределение RT. Через RT мы обозначаем равномерное распределение, заданное на интервале от половины своего среднего до полутора средних. Основные его параметры – средняя длительность интервала поступления  $\bar{\tau}$ , дисперсия  $D_{\tau} = \bar{\tau}^2 / 12$  и квадрат коэффициента вариации  $C_{\tau}^2 = D_{\tau} / \bar{\tau}^2 = 1/12$ . Интервалы времени обслуживания имеют равномерное распределение R на интервале от 0 до двух средних. При среднем  $\bar{x}$  дисперсия  $D_x = \bar{x}^2 / 3$ , квадрат коэффициента вариации  $C_x^2 = D_x / \bar{x}^2 = 1/3$ .

В системе E<sup>3</sup>|Pa|1 интервалы поступления заявок имеют распределение Эрланга третьего порядка, основные параметры которого приведены в разделе 4.1.3. Интервалы времени обслуживания имеют распределение Парето Pa( $K, \alpha$ ) с п.в.  $a(t) = \alpha K^{\alpha} / t^{\alpha+1}$ , где  $\alpha$  – параметр формы,  $K$  – граничный параметр, задающий минимальное значение с.в.  $x$ . Среднее время обслуживания  $\bar{x} = \alpha K / (\alpha - 1)$ , дисперсия  $D_x = \alpha K^2 / ((\alpha - 1)^2 (\alpha - 2))$ , квадрат коэффициента вариации  $C_x^2 = D_x / \bar{x}^2 = 1 / (\alpha(\alpha - 2))$ ,  $\alpha > 2$ . Рассмотрим распределение Pa(1; 2,5), для которого  $C_x^2 = 1 / (2,5(2,5 - 2)) = 0,8$ .

Расчет зависимостей  $L_{\text{им}}, L'_{\text{нижн.1}}, L_{\text{верх.1}}, L'_{\text{нижн.2}}, L_{\text{верх.2}}, L_{\text{прибл.}}$  при различных значениях  $\rho$  в системах RT|R|1 и E<sup>3</sup>|Pa|1 выполняется аналогично тому, как это делается в разделе 4.1.2. На рис. 1 а) представлены (в двух масштабах) графики рассматриваемых зависимостей для СМО RT|R|1, а на рис. 1 б) – для СМО E<sup>3</sup>|Pa|1. На графиках видно, что сравниваемые верхние оценки  $L_{\text{верх.1}}$  и  $L_{\text{верх.2}}$ , как и было показано в п. 3 раздела 4.1.4, имеют в системе с эрланговским потоком разную предпочтительность на разных интервалах коэффициента загрузки: при малых  $\rho$  предпочтительна оценка  $L_{\text{верх.1}}$ , при больших  $\rho$  – оценка  $L_{\text{верх.2}}$ . Причем это проявляется не только в системе E<sup>3</sup>|Pa|1 с эрланговским потоком, но и в системе RT|R|1. Нижняя оценка  $L'_{\text{нижн.2}}$  везде не хуже нижней оценки  $L'_{\text{нижн.1}}$ .

Рис. 1 позволяет также видеть, что приближение (3), график которого помечен круглыми маркерами, в целом выглядит лучше, чем другие приближения к истинной зависимости  $L(\rho)$ , представленной результатами ИМ (пунктирная кривая).

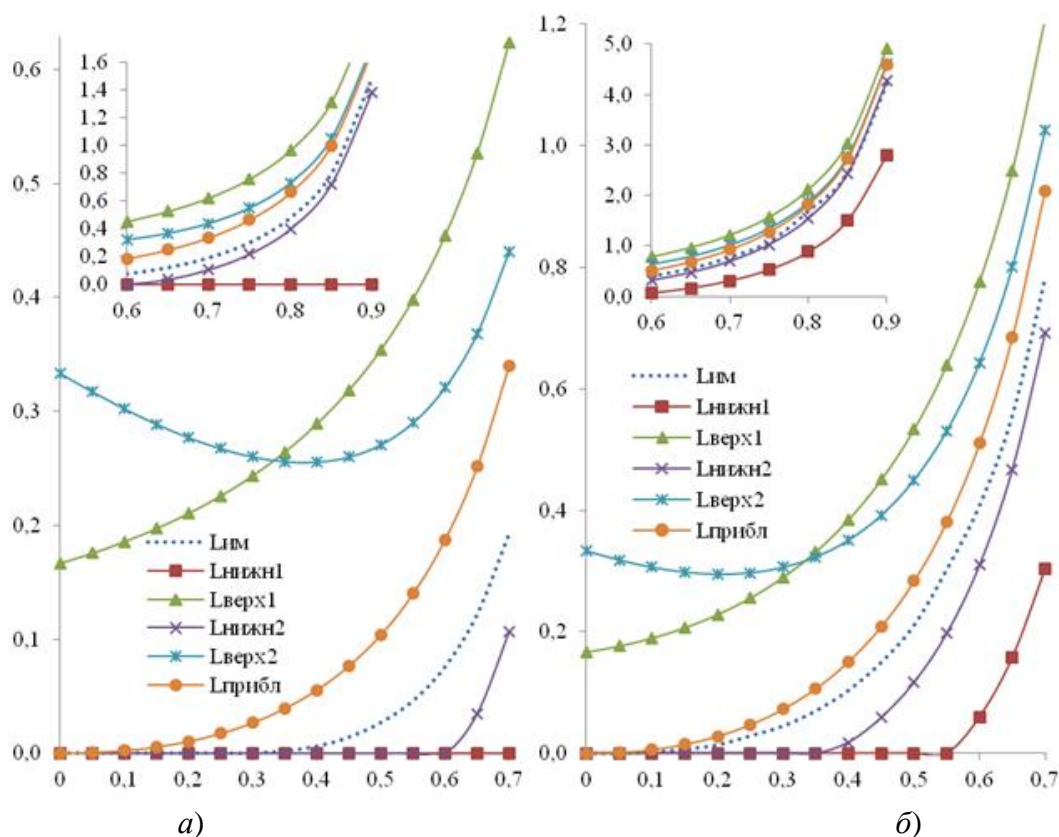


Рис. 1. Графики приближений средней длины очереди в СМО RT|R|1 (слева) и E³|Pa|1

2. Проверим экспериментально гипотезу, обоснованную в п.4 раздела 4.1.4, о том, что приближение (3) может иметь хорошую точность, если входящий поток задается гамма распределением, определяющим коэффициент вариации  $C_\tau^2 > 1$ . В Табл. 2 представлены результаты ИМ системы Г|M|1 с коэффициентом вариации  $C_\tau^2 = 2$ , полученные прогоном 5 млн. заявок при каждом конкретном коэффициенте загрузки. Текст имитационной модели имеет следующий вид.

```

GENERATE (Gamma(1,0,1/0.5,0.5)) ;Гамма поток, интенсивность =1
QUEUE 1 ;Вход в очередь
SEIZE 1 ;Занятие канала
DEPART 1 ;Выход из очереди
ADVANCE (Exponential(1,0,0.1)) ;Экспоненц-е обслуживание
RELEASE 1 ;Освобождение канала
TERMINATE ;Уход из системы

*
* Таймер модели
*

GENERATE 5000000 ;Момент завершения - 1 млн. ед. времени
TERMINATE 1 ;Завершение прогона
START 1

```

Коэффициент загрузки варьируется путем изменения соответствующего параметра генератора экспоненциальной с.в. в блоке ADVANCE.

Одновременно с расчетом погрешности приближения (3) в Табл. 2 рассчитываются приближение (3), приближение  $L_{\text{верх.1}}$  Кингмана-Келлерстрема и их погрешности  $\Delta_\Phi$  и  $\Delta_K$  соответственно.

**Таблица 2.** Сравнение точности двух приближений путем ИМ системы  $\Gamma|M|1$ 

$\rho$	$L_{\text{им}}$	Формула (3)	$L_{\text{верх.1}}$	$\Delta_{\text{ф}}$	$\Delta_{\text{к}}$
0,1	0,0033	0,017	1,116667	0,013	1,113
0,25	0,182	0,125	1,375	0,057	1,193
0,5	0,889	0,750	2,25	0,139	1,361
0,75	3,592	3,375	5,125	0,217	1,533
0,9	12,22	12,15	14,05	0,070	1,830

Данные Табл. 2 позволяют сделать вывод, что в данном частном случае проверяемая гипотеза подтвердилась. Полученные при высоком коэффициенте вариации  $C_{\tau}^2$  входящего гамма потока результаты расчета по приближенной формуле (3) оказались не только приемлемыми, но и более точными, чем по формуле Кингмана-Келлерстрема.

#### 4.3. Оценка точности приближений для систем класса $GI/GI/m$

Для многоканальных систем мы имеем известное приближение Кингмана-Келлерстрема – это верхняя граница в (1) и, возможно, приближение Файнберга (3), возможность распространения которого на многоканальные системы мы обосновали в п. 4 раздела 2. Теперь подтвердим эту возможность экспериментально и заодно сравним точность приближения (3) и приближения Кингмана-Келлерстрема.

Эксперименты выполним для системы  $\Gamma/M/3$  с имитационной моделью, отличающейся от приведенной выше модели системы  $\Gamma/M/1$  только использованием многоканального устройства вместо одноканального. Результаты моделирования и расчета проверяемых приближений и их погрешностей приведены в Табл. 3.

**Таблица 3.** Сравнение точности двух приближений путем ИМ системы  $\Gamma|M|3$ 

$\rho$	$L_{\text{им}}$	Формула (3)	$L_{\text{верх.1}}$	$\Delta_{\text{ф}}$	$\Delta_{\text{к}}$
0,1	0,004	0,017	1,116667	0,013	1,113
0,25	0,055	0,125	1,375	0,070	1,320
0,5	1,500	0,75	2,25	0,750	0,750
0,75	0,750	3,375	5,125	2,625	4,375
0,9	2,7	12,15	14,05	9,450	11,350

Как видим, приближение Файнберга (3) действительно можно распространить на многоканальные системы, причем в данном частном случае его погрешности  $\Delta_{\text{ф}}$  оказались заметно меньшими, чем погрешности  $\Delta_{\text{к}}$  приближения Кингмана-Келлерстрема (лишь при  $\rho = 0,5$  эти две погрешности совпали).

### Выводы

Описать процесс работы СМО, когда потоки событий не простейшие, в виде явных аналитических формул удастся лишь в редких случаях. Однако для принятия разумного решения по организации работы СМО часто не требуется знания точных значений всех ее характеристик, а достаточно лишь приближенно рассчитать среднюю длину  $L$  очереди и, если нужно, ряд других характеристик системы, определяемых через  $L$ .



Выполненный в статье анализ известных приближений для  $L$  привел к обнаружению ряда замечательных качеств приближения Файнберга, редко упоминаемого в литературе по ТМО (мы нашли только одно упоминание – в книге Е.С. Вентцель [10]). Проведенное аналитическое исследование этого приближения и последующие экспериментальные проверки показали, что оно может быть успешно распространено на системы с гамма распределением интервала поступления заявок, что значительно расширяет возможности применения данного приближения на практике. Более того, его оказалось возможным распространить и на многоканальные системы с входящим гамма потоком. И, наконец, в выполненных экспериментальных проверках это приближение оказалось более точным, чем известное приближение Кингмана-Келлестрема.

Все это позволяет, с одной стороны, рекомендовать приближение Файнберга для более широкого применения, в том числе при имитационном моделировании для решения задач сглаживания его результатов. С другой стороны, это ставит задачу более полного исследования данного приближения и разработки на его основе модификаций, пригодных для численного либо аналитико-имитационного решения актуальных задач оптимизации СМО и сетей СМО.

## ЛИТЕРАТУРА

1. Хейт Ф. Математическая теория транспортных потоков / пер. с англ. под ред. И.Н. Коваленко. М. : Мир, 1966. – 288 с.
2. Вишневский В. М. Теоретические основы проектирования компьютерных сетей. М. : Техносфера, 2003. – 512 с.
3. Фомин Г. П. Системы и модели массового обслуживания в коммерческой деятельности: учеб. пособие. М. : Финансы и статистика, 2000. – 144 с.
4. Клейнрок Л. Вычислительные системы с очередями / пер. с англ. под ред. Б. С. Цыбакова. М. : Мир, 1979. – 600 с.
5. Рыжиков Ю. И. Имитационное моделирование. Теория и технологии. – СПб.: КОРОНА принт; М. : Альтекс-А, 2004. – 384 с.
6. Задорожный В. Н. Оптимизация однородных немарковских сетей массового обслуживания // Проблемы управления, 2009. № 6. – С. 68-75.
7. Zadorozhnyi V. N., Zakharenkova T. R. Optimization of Channel Distribution over Nodes in Networks with Fractal Traffic // Dynamics of Systems, Mechanisms and Machines: conference proceeding, 15-17 Nov. 2016, Omsk State Technical University. – IEEE Inc., 2017. – P. 1-7.
8. Задорожный В. Н. Полуаналитические методы оптимизации транспортных сетей // Вестник кибернетики. 2018. № 4 (32). – С. 16-28.
9. Платонов Г. А., Файнберг М. А., Штильман М. С. Поезда, пассажиры и... математика. М. : Транспорт, – 1977. 239 с.
10. Вентцель Е. С. Исследование операций: задачи, принципы, методология: учеб. пособие для студентов вузов. М. : Дрофа, 2006. – 206 с.
11. Хинчин А. Я. Работы по математической теории массового обслуживания, М. : Физматгиз, 1963. – 236 с.

---

**Задорожный** Владимир Николаевич, д.т.н., профессор ОмГТУ, zwn2015@yandex.ru  
**Галич** Юлия Геннадьевна, аспирант ОмГУПС, galichyulia@list.ru

УДК 519.2:004.421.5:004.7

## РАСЧЕТ И МОДЕЛИРОВАНИЕ ОДНОКАНАЛЬНЫХ СИСТЕМ С АБСОЛЮТНЫМИ ПРИОРИТЕТАМИ

В. Н. Задорожный<sup>1</sup>, Т. Р. Захаренкова<sup>1</sup>, М. П. Маркова<sup>2</sup>

<sup>1</sup>Омский государственный технический университет, Омск, Россия

<sup>2</sup>Омский государственный университет путей сообщения, Омск, Россия

**Аннотация.** В статье приводится пример решения задачи нахождения среднего времени ожидания в системах массового обслуживания с абсолютными приоритетами и дообслуживанием. Особенностью рассматриваемой задачи является способ назначения абсолютных приоритетов заявкам входящего потока. Задача решена аналитически и с помощью имитационного моделирования. Приведены рекомендации по использованию имитационного моделирования для исследования систем с абсолютными приоритетами.

**Ключевые слова:** системы массового обслуживания, среднее время ожидания, назначение абсолютных приоритетов.

## CALCULATION AND SIMULATION SINGLE CHANNEL QUEUEING SYSTEMS WITH PREEMPTIVE PRIORITIES

V. N. Zadorozhnyi<sup>1</sup>, T. R. Zakharenkova<sup>1</sup>, M.P. Markova<sup>2</sup>

<sup>1</sup>Omsk State Technical University, Omsk, Russia

<sup>2</sup>Omsk State Transport University, Omsk, Russia

**Abstract.** Article considers the example of calculation the average waiting time in queueing systems with preemptive priorities. The feature of the problem under consideration is the technique of priorities appointment for input flow. The problem is solved analytically and by means of simulation modeling. Recommendations for correct using of simulation modeling for in queueing systems with preemptive priorities are given.

**Keywords:** queueing systems, average waiting time, preemptive priorities appointment.

### 1. Введение

Одной из характеристик систем массового обслуживания (СМО) является дисциплина обслуживания, определяющая очередность обслуживания заявок. Заявки могут быть обслужены в беспriorитетном порядке, например, в порядке поступления, либо в порядке, определяемом их приоритетами. Самыми популярными дисциплинами приоритетного обслуживания являются дисциплины с абсолютными приоритетами и дисциплины с относительными приоритетами. В системах с относительными приоритетами заявка с наиболее высоким приоритетом, попадающая в очередь, ставится в начало очереди и дожидается окончания обслуживания заявки, занимающей канал, независимо от приоритета этой обслуживаемой заявки. В системах с абсолютными приоритетами обслуживание менее приоритетной заявки немедленно прерывается, когда в СМО поступает заявка с более высоким приоритетом, и начинается обслуживание этой более приоритетной заявки. Заявки, имеющие одина-

ковый приоритет, обслуживаются в порядке поступления. Дисциплины с абсолютными приоритетами используются, например, в мультипрограммных вычислительных системах при возникновении очередей программ к процессору. При этом после выполнения поступившей более приоритетной программы прерванная программа продолжает выполняться с точки прерывания, т.е. ранее выполненная по прерванной программе работа не теряется. Этот вариант дисциплины абсолютных приоритетов называется дисциплиной абсолютных приоритетов с дообслуживанием. Именно такая дисциплина абсолютных приоритетов рассматривается далее в данной статье.

## 2. Постановка задачи

Рассмотрим СМО вида М/Ра/1 с входящим пуассоновским потоком с интенсивностью  $\Lambda$  и бесконечной буферной емкостью. Время обслуживания  $x$  имеет функцию распределения (ф. р.) Парето (Ра) следующего вида:

$$F(t) = 1 - \left( \frac{K}{t} \right)^\alpha, \quad t \geq K, \quad (1)$$

где  $\alpha$  – параметр формы,  $K$  – параметр масштаба и наименьшее значение случайной величины (с. в) с распределением Парето.

Рассмотрим простой пример разбиения множества заявок входящего пуассоновского потока на приоритетные классы в зависимости от времени  $x$  обслуживания заявок. В этом примере  $N = 1000$  приоритетных классов определяются следующим образом:

$$\begin{aligned} &\text{Если } K \leq x < K + 1, \text{ то } Pr = 999, \text{ номер класса } k = 1, \\ &\text{Если } K + 1 \leq x < K + 2, \text{ то } Pr = 998, k = 2, \\ &\dots \\ &\text{Если } K + 998 \leq x < K + 999, \text{ то } Pr = 1, k = 999, \\ &\text{Если } K + 999 \leq x < \infty, \text{ то } Pr = 0, k = 1000, \end{aligned} \quad (2)$$

где  $Pr$  – приоритет заявки, имеющей время обслуживания  $x$ . В общем случае правило (2) разбивает ось возможных значений времени обслуживания на конечное или бесконечное число промежутков равной длины. При конечном числе равновеликих промежутков имеется один (последний) полубесконечный промежуток. При уменьшении времени  $x$  обслуживания заявки ее приоритет либо не изменяется (если  $x$  остается в «своем» промежутке), либо возрастает. В примере (2)  $Pr = 999$  – самый высокий приоритет. Такой способ задания приоритетных классов введен и исследован в [1]. При оптимальном выборе длины и числа равных промежутков правило (2) позволяет кардинально сокращать среднее время ожидания заявок при бесконечной дисперсии времени обслуживания, которая характерна для обмена файлами в условиях фрактального трафика.

Правила вида (2) формируют приоритетные классы в ходе поступления заявок. В традиционных задачах оптимального назначения приоритетов конечное число входящих потоков (классов) заявок определяется условиями задачи произвольно и статично: конечное число входящих потоков и их параметры перечисляются в условиях задачи, и приоритеты нужно распределить оптимально по

этим заданным входящим потокам. Вопросы оптимального распределения заявок по приоритетным классам при этом не ставятся и не решаются.

Рассчитаем среднее время ожидания  $W$  в системе с абсолютными приоритетами, формируемыми по правилу (2).

### 3. Теория

Аналитическое выражение среднего времени ожидания  $W$  в СМО с абсолютными приоритетами известно для СМО вида  $M/G/1$ , когда второй момент ф.р.  $G$  является конечным [2]. В работе [1] разработан численный метод, позволяющий рассчитывать среднее время ожидания  $W$  в СМО с бесконечным вторым моментом  $b^{(2)} = \infty$  ф.р.  $G$ . Для исследования СМО с отличным от пуассоновского входящим потоком приходится использовать имитационное моделирование.

Для того чтобы рассчитать среднее время ожидания в СМО с абсолютными приоритетами, необходимо рассчитать средние времена ожидания для заявок каждого приоритетного класса  $k$ . В работе [2, с. 150] приводится формула для среднего времени  $U_k$  пребывания в СМО заявок  $k$ -го приоритетного класса:

$$U_k = \frac{b_k}{1 - \sigma_{k-1}} + \frac{\sum_{i=1}^k \lambda_i b_i^{(2)}}{2(1 - \sigma_k)(\sigma_{k-1})}, \quad k = 1, 2, \dots, \quad (3)$$

где  $b_k$  – среднее время обслуживания заявок  $k$ -го приоритетного класса,  $b_i^{(2)}$  – второй момент времени обслуживания заявок  $i$ -го приоритетного класса,

$\lambda_i$  – интенсивность поступления заявок  $i$ -го приоритетного класса,  $\sigma_k = \sum_{i=1}^k \rho_k$  –

сумма коэффициентов загрузки системы заявками с 1-го по  $k$ -ый приоритетные классы,  $\rho_k = \lambda_k b_k$ ,  $\sigma_{k-1} = \sigma_k - \rho_k$ . Различие индексов у суммы в (3) и у соответствующей суммы в формуле (3.39) из [2, с. 150] объясняется тем, что в [2] чем больше номер класса  $k$ , тем выше приоритет  $Pr$ . В нашей работе наоборот, чем больше  $k$ , тем ниже приоритет  $Pr$ . По этой же причине различаются индексы в определении параметра  $\sigma_k$  в [2] и в нашей статье.

Теперь, зная среднее время (3) пребывания в СМО заявок  $k$ -го приоритетного класса, можно вычислить и среднее время ожидания заявок  $k$ -го приоритетного класса:

$$w_k = U_k - b_k, \quad k = 1, 2, \dots \quad (4)$$

Тогда общее среднее время ожидания будет выражено суммой:

$$W = \sum_{k=1}^{1000} p_k w_k, \quad (5)$$

где  $p_k$  – вероятность попадания времени  $x$  в промежуток  $[K + k - 1; K + k)$ .

Принимая во внимание то, что время обслуживания имеет распределение Парето (1), найдем точки  $t_k = K + k$  и входящие в (3) параметры СМО:

$$\begin{aligned}
 p_k &= K^\alpha (t_{k-1}^{-\alpha} - t_k^{-\alpha}), & \sigma_k &= \rho(1 - K^{\alpha-1} t_k^{1-\alpha}), \\
 b_k &= \frac{\alpha}{\alpha-1} \cdot \frac{t_{k-1}^{1-\alpha} - t_k^{1-\alpha}}{t_{k-1}^{-\alpha} - t_k^{-\alpha}}, & b_k^{(2)} &= \frac{\alpha}{\alpha-2} \cdot \frac{t_{k-1}^{2-\alpha} - t_k^{2-\alpha}}{t_{k-1}^{-\alpha} - t_k^{-\alpha}}.
 \end{aligned} \quad (6)$$

С помощью этих формул вычислим точно среднее время ожидания  $W$  в рассматриваемой СМО М/Па/1 с абсолютными приоритетами и сравним его с результатами имитационного моделирования этой СМО.

#### 4. Расчет и моделирование системы

Пусть интенсивность входящего потока  $\Lambda = 1$  и параметры распределения Парето  $\alpha = 2,5$ ,  $K = 0,48$ . При разбиении входящего потока на 1000 приоритетных классов (2) находим по расчетным формулам (3), (6), что среднее время ожидания  $W = 1,297$ .

Сравним полученное точное решение с результатами имитационного моделирования системы. Ее имитационную модель построим с помощью инструментов «Библиотеки моделирования процессов» системы имитационного моделирования Anylogic [4]. Диаграмма состояний изучаемой модели СМО с дисциплиной абсолютных приоритетов с дообслуживанием представлена на Рис. 1.

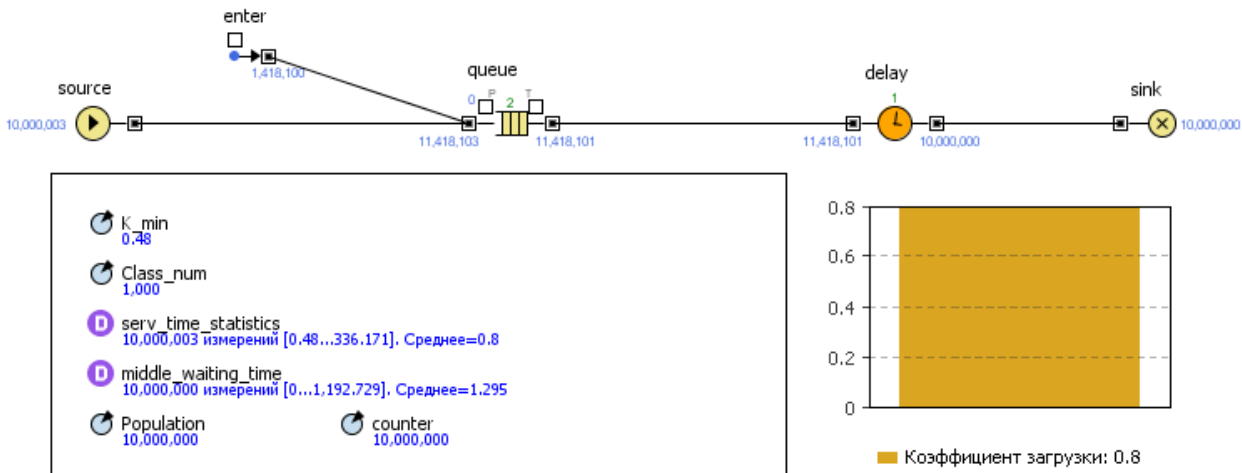


Рис. 1. Результаты моделирования СМО М/Па/1 с абсолютными приоритетами при  $N = 1000$  приоритетных классов и длине прогона  $n = 10$  млн. заявок

На Рис. 1 представлены и результаты имитационного моделирования СМО. В ходе моделирования было сгенерировано 10 млн. заявок (см. блок *source*) входящего потока.

Для вычисления статистики по среднему времени ожидания заявок в СМО (*middle\_waiting\_time*) был создан класс, расширяющий возможности базового класса *Entity*. В новый класс были добавлены атрибуты для фиксации времени нахождения заявки в различных блоках имитационной модели и для определения приоритета заявки.

При моделировании в Anylogic 6 мы обнаружили, что Anylogic не свободен от недостатка, обнаруженного при моделировании приоритетных СМО в GPSS в работе [3]. При прерывании обслуживания заявки другой заявкой с более высоким приоритетом прерванная заявка ставится в конец очереди заявок с таким же приоритетом, тогда как в соответствии с дисциплиной абсолютных приоритетов

должна попадать в начало этой очереди. Для исправления данного недостатка при построении имитационной модели мы использовали способ с четными и нечетными приоритетами, предложенный в [3] для моделирования СМО с абсолютными приоритетами на GPSS [5].

В блоке *enter* реализовано правильное помещение прерванной заявки в свою приоритетную очередь. Правильное помещение заявки в свою приоритетную очередь потребовало создания новой заявки, идентичной прерванной заявке, снятия прерванной заявки с обслуживания и удаления ее из имитационной модели в целом.

При прогоне 10 млн. заявок для  $W$  получена имитационная оценка  $W_{им} = 1,295$ , и уже при 15 млн. заявок оценка приняла значение  $W_{им} = 1,297$ .

## 5. Заключение

В статье решена задача нахождения среднего времени ожидания в системе М/Ра/1 аналитически и с помощью имитационного моделирования. Обнаружены проблемы имитационного моделирования систем с абсолютными приоритетами и дообслуживанием в среде Anylogic, обусловленные неправильным определением стандартными средствами Anylogic места в очереди для прерванных заявок. Применен и испытан предложенный в [3] метод для решения такой же проблемы в GPSS. Метод основан на повышении приоритета прерванной заявки на «полуровня» и обеспечивается изначальным назначением только четных приоритетов. Полученные в статье результаты могут быть использованы для моделирования систем GI/GI/1 с абсолютными приоритетами в среде Anylogic. Построенная в статье имитационная модель может применяться в качестве примера для построения аналогичных моделей других приоритетных систем.

## ЛИТЕРАТУРА

1. Zadorozhnyi V. N., Zakharenkova T. R. Methods to Reduce Loss Probability in Systems with Infinite Service Time Dispersion. In: Dudin A., Nazarov A., Moiseev A. (eds) Information Technologies and Mathematical Modelling. Queueing Theory and Applications. ITMM 2019. Communications in Computer and Information Science, vol 1109. Springer, Cham, – 2019. – P. 296–311.
2. Клейнрок Л. Вычислительные системы с очередями / Москва: изд. Мир, 1979. – 586 с.
3. Соснин, В. В. Моделирование дисциплины обслуживания с абсолютными приоритетами в GPSS World / В. В. Соснин // Имитационное моделирование. Теория и практика: сборник докладов третьей всероссийской научно-практической конференции ИММОД-2007. – Санкт-Петербург, 2007. – Том 1. – С. 224–229.
4. Карпов Ю. Г. Имитационное моделирование систем. Введение в моделирование с AnyLogic 5. – СПб: БХВ-Петербург, 2006. – 400 с.
5. GPSS World reference manual. Minuteman Software, Fifth Edition, Holly Springs, 2009.

**Задорожный Владимир Николаевич**, д.т.н., профессор, каф. «Математические методы и информационные технологии в экономике», zwn2015@yandex.ru

**Захаренкова Татьяна Романовна**, ассистент, каф. «Математические методы и информационные технологии в экономике», zakharenkovatiana@gmail.com

**Маркова Марина Петровна**, аспирант 2 курса ОмГУПС, markovamp@mail.ru

## ОБОБЩЕНИЕ МЕТОДА РАВНОВЕРОЯТНОГО ВЫБОРА ОСТОВНЫХ ДЕРЕВЬЕВ

*В. Н. Задорожный, М. Н. Юдина*

Омский государственный технический университет, Омск, Россия

**Аннотация.** Развивается метод случайного выбора остовных деревьев, предложенный и реализованный авторами в предыдущих публикациях для анализа подграфов на четырех вершинах. Решается задача получения несмещенных и эффективных статистических оценок для анализа подграфов на любом заданном числе вершин.

**Ключевые слова:** подграфы, случайный выбор подграфов, сетевые мотивы, графлеты.

## GENERALIZATION OF METHOD OF SPANNING TREES SAMPLING

*V. N. Zadorozhnyi, M. N. Yudina*

Omsk State Technical University, Omsk, Russia

**Annotation.** The method of spanning trees sampling is developed. This method was proposed and implemented in our previous publications if subgraphs on four vertices was considered. The problem of obtaining unbiased and effective statistical estimates for the frequency of occurrence of subgraphs at any given number of vertices is solved.

**Keywords:** subgraphs, subgraph sampling, network motifs, graphlets.

### 1. Введение

Задача расчета частот встречаемости связных подграфов возникла в рамках Науки о сетях (Network Science). Рассматриваются только связные подграфы. Подграфы, относящиеся к одному классу изоморфизма, считаются идентичными, т.е. считаются одним и тем же подграфом, встречающимся в графе несколько раз. Как правило, рассчитываются частоты встречаемости всех *разных* подграфов фиксированного размера (т.е. разных подграфов с одним и тем же числом вершин в подграфах). Частоты, с которыми разные подграфы встречаются внутри графа исследуемой сети, являются важной характеристикой топологии сети и лежат в основе таких понятий как сетевые мотивы (network motifs [1]) и распределение частот встречаемости подграфов (graphlet degree distribution [2]). Различают два вида частоты встречаемости подграфа. Абсолютная частота встречаемости – это число вхождений данного подграфа в граф. Каждое вхождение определяется конкретным набором вершин и всех ребер графа (дуг, если граф ориентированный), соединяющих пары вершин этого набора. Относительная частота встречаемости подграфа – это отношение его абсолютной частоты к сумме абсолютных частот всех разных подграфов.

В последние годы предложены различные методы для подсчета частот встречаемости подграфов, в том числе статистические методы расчета [3]. Среди статистических методов следует выделить методы случайного выбора пути [4–

7]. Методы случайного выбора пути не используют эвристических допущений в отличие от таких статистических методов, как методы обобщения графов [2, 8]. Кроме того, методы случайного выбора пути позволяют оценивать как относительные, так и абсолютные частоты встречаемости подграфов в отличие от методов на основе случайного блуждания [9, 10], которые позволяют получать статистические оценки только для относительных частот. Наконец, методы случайного выбора пути являются корректной реализацией метода Монте-Карло и позволяют получать состоятельные, несмещенные, эффективные статистические оценки частот встречаемости подграфов в отличие, например, от методов перечисления с отсечением выборки [11, 12].

В пионерской работе [4] по разработке методов случайного выбора пути предлагается использовать случайный равновероятный выбор пути из двух ребер и трех вершин («3-путь») и оценивать долю «треугольников» в графе. Зная общее число всех «3-путей» возможно оценить и общее число каждого из двух неориентированных связных подграфов на трех вершинах.

В работах [5] и [6] предлагается использовать случайный выбор пути на трех ребрах («4-путь») с последующим подсчетом доли подграфов для пяти из шести возможных классов изоморфизма. Число подграфов, относящихся к шестому классу изоморфизма, можно вычислить через частоты встречаемости других подграфов и общее число «4-путей».

В работе [7] предлагается реализация метода случайного пути для подграфов на пяти вершинах.

Общий недостаток предложенных методов случайного выбора пути заключается в том, что они могут использоваться только для работы с простыми неориентированными графами. В то же время многие сети, исследуемые в Науке о сетях (генные сети, сети ссылок веб-страниц, некоторые социальные сети), представляются смешанными или ориентированными простыми графами.

Этот недостаток удалось устранить в работе [13], в которой предлагается метод случайного выбора остовных деревьев (каркасов) для расчета частот встречаемости подграфов на четырех вершинах. В работах [13, 14] предложены алгоритмы и математические соотношения, позволяющие получать методом случайного выбора остовных деревьев состоятельные, несмещенные и эффективные оценки для частот встречаемости подграфов на четырех вершинах.

В настоящей статье выполняется дальнейшая разработка метода случайного выбора остовных деревьев (каркасов) [13], предлагается и исследуется его версия, представляющая собой, по-существу, новый метод, который позволяет с высокой скоростью и точностью рассчитывать корректные оценки частот встречаемости подграфов, заданных на любом фиксированном числе вершин.

## 2. Метод случайного выбора остовных деревьев

В методе случайного выбора остовных деревьев предлагается выбирать не пути, а остовные деревья различных подграфов. Остовным деревом (или каркасом) связного подграфа на  $k$  вершинах называется вложенное в него дерево, содержащее все  $k$  вершин и  $k - 1$  ребер подграфа. При этом направления дуг не учитываются.



Использование случайного выбора остовных деревьев позволяет работать как с неориентированными, так и с ориентированными и смешанными графами.

В методе случайного выбора остовных деревьев выборочное пространство определяется как множество реализаций (экземпляров) остовного дерева в заданном большом графе. Оценка  $n_i$  для частоты встречаемости конкретного связного подграфа ( $i$ -го из числа всех возможных на  $k$  вершинах) вычисляется по случайной выборке  $N^+$  экземпляров остовного дерева, вложенного в этот подграф. Выборка осуществляется таким образом, чтобы все экземпляры рассматриваемого остовного дерева выбирались равновероятно [13]. Для каждого экземпляра остовного дерева в выборке путем его дополнения всеми ребрами графа, соединяющими пары вершин этого экземпляра, определяется подграф, в который данный экземпляр дерева вложен, и подсчитывается число  $C_i$  обнаружений подграфа  $i$  по всей выборке. Зная точное общее число  $N_i$  экземпляров рассматриваемого остовного дерева в графе, можно вычислить следующую статистическую оценку для частот встречаемости подграфа  $i$  в большом графе:

$$n_i = \frac{C_i}{N_i^+} N_i / \text{coef}_i,$$

где поправочный коэффициент  $\text{coef}_i$  равен числу разных способов, которыми рассматриваемое остовное дерево может быть вложено в подграф  $i$ .

На Рис. 1 представлен алгоритм вычисления статистических оценок для частот всех подграфов, обнаруживаемых при анализе равновероятной случайной выборки  $N^+$  экземпляров одного и того же остовного дерева.

#### НАЧАЛО

**Ввод:** граф  $G(V,E)$ , остовное дерево, массив поправочных коэффициентов  $\text{coef}[i]$ ,  $i \in X$ , где  $X$  – номера различных подграфов фиксированного размера

**Вывод:** статистические оценки частот  $n[i]$  встречаемости подграфов,  $i \in X$

1. Для всех  $i \in X$  выполнить  $n[i] \leftarrow 0$ ;  $C[i] \leftarrow 0$ ;
2. Рассчитать общее число  $N$  остовных подграфов в графе;
3. Для  $j = 1, \dots, N^+$  выполнить
  - 3.1. Выбрать из графа равновероятно экземпляр остовного дерева и определить множество вершин  $V'$ , входящих в этот экземпляр
  - 3.2.  $\text{Num} \leftarrow \text{GetISO}[G[V']]$  // определить номер  $\text{Num} \in X$  подграфа, лежащего на  $V'$
  - 3.3.  $C[\text{Num}] = C[\text{Num}] + 1$ ;
4. Для всех  $i \in X$  выполнить  $n[i] = N \cdot C[i] / N^+ / \text{coef}[i]$ .

#### КОНЕЦ

Рис. 1. Алгоритм расчета статистических оценок для частот встречаемости подграфов

Применение описанного алгоритма достаточно легко обосновывается для случая, когда в подграф входит только одно остовное дерево. Если же в подграф вложено нескольких различных остовных деревьев (Рис. 2), то расчет оценки для частоты его встречаемости несколько усложняется. В этом случае применение описанного выше алгоритма к выборкам, полученных для разных остовных деревьев, дает несколько различных статистических оценок частоты встречаемости одного и того же подграфа.

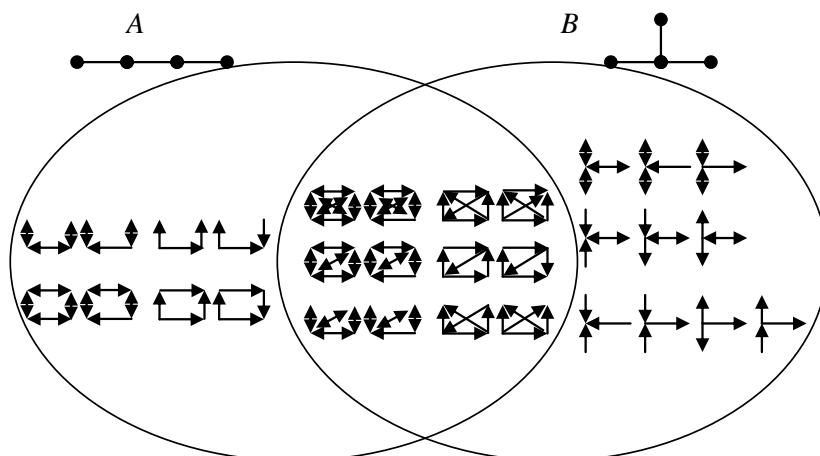


Рис. 2. Диаграмма Венна охвата остовными деревьями подграфов, относящихся к различным классам изоморфизма,  $A$  – множество подграфов, которые находятся на остовном дереве «4-путь»,  $B$  – множество подграфов, которые находятся на остовном дереве «4-ветвление»

У связных подграфов на пяти вершинах существуют три различных остовных дерева: «5-путь», «5-ветвление» («крест») и «Т-дерево». Некоторые 5-подграфы также содержат два разных остовных дерева (Рис. 3).

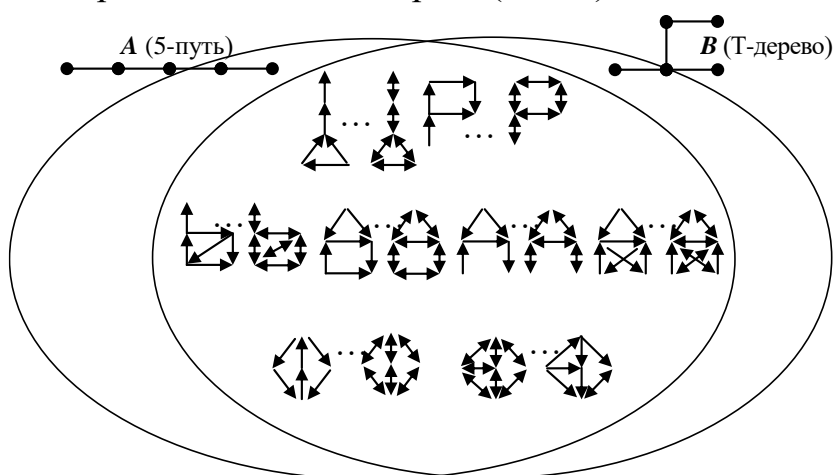


Рис. 3. Подграфы на пяти вершинах, для которых могут быть получены две различные статистические оценки

Если для некоторого подграфа в результате двух различных статистических экспериментов получены две различные оценки  $n^{[1]}$  и  $n^{[2]}$ , то итоговую оценку частоты встречаемости этого подграфа целесообразно определить как линейную комбинацию этих оценок [7, 13, 14]:

$$n = \lambda n^{[1]} + (1 - \lambda) n^{[2]}, \quad 0 < \lambda < 1. \quad (1)$$

Оценка  $n$  как функция случайных величин  $n^{[1]}$  и  $n^{[2]}$  также является случайной величиной. Значение коэффициента  $\lambda$  определяется для (1) из условия минимизации среднеквадратичного отклонения (или коэффициента вариации) итоговой оценки  $n$ , являющихся статистическими аналогами абсолютной погрешности (или относительной погрешности) оценки  $n$  соответственно. Случайные вели-

чины  $n^{[1]}$ ,  $n^{[2]}$  (и, следовательно,  $n$ ) имеют асимптотически нормальное распределение (счетчики  $C_i$  встречаемости имеют распределение Бернулли). При этом, как это показано в работах [7, 14], минимум дисперсии оценки  $n$  достигается при

$$\lambda = \frac{D(n^{[2]})}{D(n^{[2]}) + D(n^{[1]})} = \frac{D_2}{D_2 + D_1} = \frac{D_1^{-1}}{D_1^{-1} + D_2^{-1}}. \quad (2)$$

В статистических экспериментах дисперсии в (2) заменяются их оценками, и, соответственно, значение  $\lambda$  определяется приближенно.

Однако при расчете статистических оценок для частот встречаемости подграфов на пяти вершинах могут быть получены уже три различные статистические оценки (Рис. 4) для частот встречаемости подграфов на шести вершинах – шесть оценок, на семи – одиннадцать и т.д. (по числу различных остовных деревьев в связных графах на заданном числе вершин).

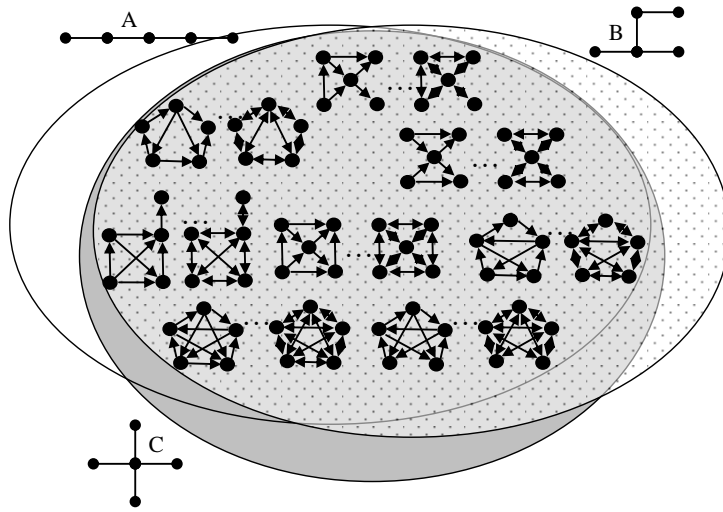


Рис. 4. Ориентированные подграфы на пяти вершинах, для которых могут быть получены три различные статистические оценки

Поэтому необходимо рассмотреть общий случай, когда при использовании метода случайного выбора остовных деревьев в процессе расчета частот встречаемости подграфов получаются  $r$  различных статистических оценок.

### 3. Свертки оценок при $r$ различных остовных деревьях

Итоговую статистическую оценку абсолютных частот встречаемости какого либо подграфа в случае вхождения в него  $r$  различных остовных деревьев представим в виде линейной комбинации  $r$  полученных оценок:

$$n = \lambda_1 n^{[1]} + \lambda_2 n^{[2]} + \dots + \lambda_r n^{[r]}, \text{ где } \lambda_1 + \lambda_2 + \dots + \lambda_r = 1 \quad (3)$$

Найдем такие коэффициенты  $\lambda_1, \lambda_2, \dots, \lambda_r$ , которые минимизируют дисперсию  $D[n]$  итоговой оценки  $n$ , т.е. обеспечивают получение эффективной оценки. Заметим, что математическое ожидание оценки  $n$

$$\begin{aligned} M[n] &= M[\lambda_1 n^{[1]} + \lambda_2 n^{[2]} + \dots + \lambda_r n^{[r]}] = \lambda_1 M[n^{[1]}] + \lambda_2 M[n^{[2]}] + \dots + \lambda_r M[n^{[r]}] = \\ &= \lambda_1 M_1 + \lambda_2 M_2 + \dots + \lambda_r M_r. \end{aligned}$$

Экстремальную задачу

$$D[n] = D[\lambda_1 n^{[1]} + \lambda_2 n^{[2]} + \dots + \lambda_r n^{[r]}] = \lambda_1^2 D_1 + \lambda_2^2 D_2 + \dots + \lambda_r^2 D_r \rightarrow \min, \\ \lambda_1 + \lambda_2 + \dots + \lambda_r = 1$$

решим методом множителей Лагранжа.

Составляя функцию Лагранжа с множителем  $\lambda$

$$L = \lambda_1^2 D_1 + \lambda_2^2 D_2 + \dots + \lambda_r^2 D_r + \lambda(\lambda_1 + \lambda_2 + \dots + \lambda_r - 1),$$

приравнивая ее частные производные по  $\lambda_i$  к нулю, и добавляя к  $r$  полученным равенствам уравнение связи  $\lambda_1 + \lambda_2 + \dots + \lambda_r = 1$ , получаем следующую систему уравнений:

$$\begin{cases} L_1 = 2D_1 \cdot \lambda_1 + \lambda = 0, \\ L_2 = 2D_2 \cdot \lambda_2 + \lambda = 0, \\ \dots \\ L_r = 2D_r \cdot \lambda_r + \lambda = 0, \\ \lambda_1 + \lambda_2 + \dots + \lambda_r = 1. \end{cases}$$

Ее решение

$$\lambda_1 = \frac{D_1^{-1}}{D_1^{-1} + \dots + D_r^{-1}}, \dots, \lambda_r = \frac{D_r^{-1}}{D_r^{-1} + \dots + D_r^{-1}}. \quad (4)$$

дает значения  $\lambda_1, \lambda_2, \dots, \lambda_r$ , доставляющие минимум дисперсии  $D[n]$ , поскольку дифференциал второго порядка функции Лагранжа  $L$  везде больше нуля, за исключением точек, в которых  $d\lambda_1, d\lambda_2, \dots, d\lambda_r$  одновременно равны нулю:

$$\begin{aligned} d^2 L &= L''_{\lambda_1 \lambda_1} (d\lambda_1)^2 + L''_{\lambda_2 \lambda_2} (d\lambda_2)^2 + \dots + L''_{\lambda_r \lambda_r} (d\lambda_r)^2 + \\ &+ 2L''_{\lambda_1 \lambda_2} d\lambda_1 d\lambda_2 + 2L''_{\lambda_1 \lambda_3} d\lambda_1 d\lambda_3 + \dots + 2L''_{\lambda_{r-1} \lambda_r} d\lambda_{r-1} d\lambda_r = \\ &= 2D_1 (d\lambda_1)^2 + 2D_2 (d\lambda_2)^2 + \dots + 2D_r (d\lambda_r)^2 > 0. \end{aligned}$$

Заметим, что решение (2) является частным случаем решения (4).

#### 4. Заключение

В представленной статье развивается метод случайного выбора остовных деревьев в большом графе. Основа предложенного метода заключается в равновероятном выборе реализаций (экземпляров) остовного дерева на заданном числе вершин. Измеряя в таком эксперименте относительные частоты встречаемости подграфов, лежащих на вершинах выбранных экземпляров остовного дерева, и используя известное точное общее число экземпляров этого дерева в графе мы рассчитываем статистические оценки частот встречаемости разных подграфов. Если на заданном числе вершин существует несколько разных (относящихся к различным классам изоморфизма) остовных деревьев, то проводятся статистические испытания по каждому из этих деревьев. При этом может быть

получено несколько (частных) статистических оценок для частоты встречаемости одного и того же подграфа, полученных по разным выборкам. В этом случае итоговые оценки частот встречаемости подграфов определяются как линейные комбинации соответствующих частных оценок. Коэффициенты линейных комбинаций рассчитываются по найденным в статье формулам, обеспечивающим минимум дисперсии каждой итоговой оценки. Выведенные аналогичным способом формулы, обеспечивающие минимум коэффициентов вариации итоговых оценок (не приведенные в статье в связи с ограничениями ее объема), также являются частью описанного метода. Тем самым в описанном методе реализуются не только корректные (несмещенные) оценки частот встречаемости подграфов, но и минимизируются абсолютные и/или относительные погрешности этих оценок.

## ЛИТЕРАТУРА

1. Milo R., et al. Network motifs: Simple building blocks of complex networks // *Science*, 2002. – V. 594, № 298. – P. 824–827.
2. Pržulj N. et al. Efficient estimation of graphlet frequency distributions in protein–protein interaction networks // *Bioinformatics*, 2006. – V. 22, №8. – P. 974–980.
3. Ribeiro P., et al. Survey on Subgraph Counting: Concepts, Algorithms and Applications to Network Motifs and Graphlets – arXiv preprint arXiv:1910.13011, 2019.
4. Seshadhri C., et al. Triadic measures on graphs: The power of wedge sampling. In *Proceedings of the SIAM International Conference on Data Mining*. SIAM, 2013. – P. 10–18.
5. Jha M., et al. Path sampling: A fast and provable method for estimating 4-vertex subgraph counts // In *Proceedings of the 24th International Conference on World WideWeb*, 2015. – P. 495–505.
6. Yudin E. B., Zadorozhnyi V. N. Statistical Approach to Calculation of Number of Network Motifs / E. B. Yudin, V. N. Zadorozhnyi // *Proceedings – International Siberian conference on control and communications, SIBCON*. – Omsk, 2015. – P. 1–4.
7. Wang P. et al. MOSS-5: A fast method of approximating counts of 5-node graphlets in large graphs. *IEEE Transactions on Knowledge and Data Engineering*, 2018. – V. 30, №1. P. 73–86.
8. Rahman M., et al. Graft: An efficient graphlet counting method for large graph analysis // *IEEE Transactions on Knowledge and Data Engineering*, 2014. – V. 26, №10. – P. 2466–2478.
9. Bhuiyan M. A., et al. Guise: Uniform sampling of graphlets for large graph analysis // *Proceedings of the 12th International Conference on Data Mining (ICDM)*, 2012. – P. 91–100.
10. Yang C., et al. SSRW: A Scalable Algorithm for Estimating Graphlet Statistics Based on Random Walk // In *Proceedings of the International Conference on Database Systems for Advanced Applications*, Springer, 2018. – P. 272–288.
11. Wernicke S., Rasche F. FANMOD: a tool for fast network motif detection // *Bioinformatics*, 2006. – V. 22, № 9. P. 1152–1153.
12. Paredes P., Ribeiro P. Rand-FaSE: fast approximate subgraph census // *Social Network Analysis and Mining*, 2015. – V. 5, № 1. – P. 17.
13. Задорожный В. Н., Юдина М. Н. Расчет сетевых мотивов комбинированным методом случайной выборки каркасов // *Вестник технологического университета*, 2019. – Т.22, № 11. – С. 122–127.
14. Юдина М. Н. Анализ несмещенности и эффективности оценок частот встречаемости сетевых мотивов в статистических методах расчета // *Вестник кибернетики*, 2019. – Т.36, № 4. – С. 6–18.

---

**Задорожный** Владимир Николаевич, д.т.н., профессор кафедры ММиИТЭ, zwn2015@yandex.ru  
**Юдина** Мария Николаевна, ассистент кафедры ММиИТЭ, mg-and-all@mail.ru

УДК 519.168: 004.421

## ПОСТРОЕНИЕ ГРАФОВЫХ МОДЕЛЕЙ СЕТИ ЦИТИРОВАНИЯ НАУЧНЫХ ПУБЛИКАЦИЙ

И. Г. Ольгина<sup>1</sup>, И. В. Пронин<sup>2</sup>, А. Н. Абдрахманов<sup>2</sup>

<sup>1</sup>Омский государственный технический университет, г. Омск, Россия

<sup>2</sup>Омский государственный университет путей сообщения, г. Омск, Россия

**Аннотация.** Предметом исследования являются графовые модели сетей цитирования научных публикаций. Разрабатываются метод и алгоритм построения ориентированного графа сети цитирования. Вершины такого графа соответствуют публикациям, а дуги – ссылкам на публикации (цитированиям). Приводится описание реализованной в процессе исследования программы – поискового робота. Представляются результаты построения модели сети в виде орграфа на основе данных о цитировании, собранных с помощью поискового робота. Информация о текущем состоянии цитирования извлечена из базы данных Научной электронной библиотеки eLIBRARY.RU.

**Ключевые слова:** сеть цитирования, ориентированные графы, алгоритм построения модели, поисковый робот.

## BUILDING GRAPH MODELS OF THE SCIENTIFIC PUBLICATION CITATION NETWORK

I. G. Olgina<sup>1</sup>, I. V. Pronin<sup>2</sup>, A. N. Abdrahmanov<sup>2</sup>

<sup>1</sup>Omsk State Technical University, Omsk, Russia

<sup>2</sup>Omsk State Transport University, Omsk, Russia

**Abstract.** The subject of the study is graph models of scientific publication citation networks. An algorithm that implements a method for constructing an oriented graph of the citation network is presented. The vertices of this graph correspond to publications, and the arcs – links to publications (citations). The description of the search robot program implemented during the research is given. The results of constructing a network model in the form of a digraph based on citation data collected by a search robot are presented. Information about the current state of citation is extracted from the database of the Scientific electronic library eLIBRARY.RU.

**Keywords:** citation network, directed graphs, algorithm of the model, search robot.

### 1. Введение

Развитие технологий использования и пополнения источников научной информации, рост в геометрической прогрессии общего объема публикаций обуславливают актуальность задач анализа взаимосвязей научных публикаций. В настоящее время большое число ученых мира занимается определением ценности публикаций [1, 2], вычислением показателей популярности журналов, оценкой вклада авторов и организаций в науку [3], изучением тенденций развития предметных областей. Для наукометрического анализа публикаций разрабатыва-

ются и исследуются новые дополнительные показатели: альтметрики, вебометрики и информетрики и другие [4]. Постоянно расширяется применение наукометрического анализа публикаций, основанного на анализе сетей цитирования.

Наука о сетях (Network Science) представляет собой научный подход, который использует анализ взаимосвязей для изучения и интерпретации сложных сетевых систем, включая представление их в виде графовых моделей. В процессе выявления взаимосвязи публикаций образуются сети цитирования [5]. Узлами сети цитирования являются научные публикации. Связь, направленная от узла  $a$  к узлу  $b$  показывает, что публикация  $a$  ссылается на публикацию  $b$  (цитирует ее).

Первоначальной задачей в изучении структурной информации о сети цитирования является сбор данных. От качества и полноты собранных данных зависит, насколько адекватно будет отражать модель структуру сети. Следовательно, адекватность модели зависит от того, насколько точный и глубокий анализ связей между публикациями можно будет осуществить. Сбор данных целесообразно выполнять в автоматизированном режиме с использованием поисковых роботов, или «веб-пауков». В данной статье разрабатывается алгоритм сбора данных о цитировании в интернете. Целью работы является разработка методов и моделей для исследования сетей цитирования научных публикаций.

## 2. Основные понятия

Математической моделью сетей цитирования являются ориентированные графы. Узлами сети являются научные публикации, связями – коммуникации между ними, реализуемые путем цитирования. Цитироваться могут как научные статьи, так и другие источники информации (книги, журналы, брошюры). На Рис. 1 в качестве примера представлен фрагмент графа сети цитирования.

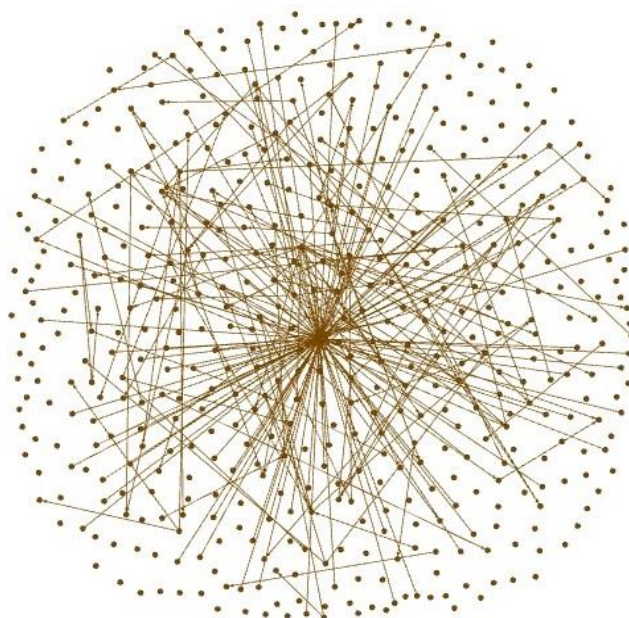


Рис. 1. Фрагмент графа сети цитирования Научной электронной библиотеки eLIBRARY.RU

Сеть цитирования научных публикаций представляется в виде ориентированного графа  $G = (V, E)$ , где  $V$  – множество вершин графа,  $E$  – множество дуг графа.

Граф представляется в виде списка дуг  $(v_i, v_j) \in E$ , где на первом месте указывается вершина, из которой дуга исходит – начальная вершина дуги, а на втором – конечная вершина, в которую дуга заходит. Вершина  $v_i$  соответствует цитирующей публикации, вершина  $v_j$  – цитируемой. Для публикаций, находящихся в отношении цитирования, должно выполняться следующее условие: никакая публикация не должна цитировать саму себя.

Следовательно, для всякой дуги  $(v_i, v_j)$  должно выполняться неравенство  $v_i \neq v_j$  (граф сети цитирования не должен иметь петель).

В статье [6] рассмотрен другой способ представления графа – описание графа матрицей смежности  $M = ||m_{ij}||$ , и учтено еще одно условие для отношения цитирования: если публикация  $a$  цитирует публикацию  $b$ , то публикация  $b$  не может цитировать публикацию  $a$ . В настоящей статье данное условие не принимается во внимание, так как в ходе экспериментов по сбору данных о цитировании иногда встречаются случаи, когда две статьи обе одновременно цитируют одна другую, находясь в одном и том же выпуске журнала или сборника статей. Мы считаем, что информация обо всех существующих цитированиях должна входить в модель.

### 3. Алгоритм получения сети цитирования научных публикаций

Основной задачей выявления структурной информации о сети цитирования является сбор данных. Сбор данных необходимо выполнять в автоматическом режиме с использованием компьютерных программ – поисковых роботов. Пример реализации алгоритма получения сети ссылок веб-страниц на основе поиска в ширину и применения программных средств для сбора данных приводится в статьях [7, 8].

На Рис. 2 и Рис. 3 представлен алгоритм сбора данных о цитировании научных публикаций, который позволяет построить граф сети цитирования. В статье рассматривается случай для ориентированного графа. Приведенный алгоритм можно также использовать при построении неориентированного графа. Для этого следует все дуги между вершинами ориентированного графа заменить ребрами между соответствующими вершинами.

Построение сети начинается с выбора научной публикации  $v$  в интернете либо в исследуемой базе данных, где должен быть произведен сбор данных о научных публикациях. В ходе извлечения сети необходимо реализовать возможность при дальнейшем анализе полученных данных осуществлять выборку по ключевым словам либо исследуемым тематикам публикаций. Поэтому в алгоритме предусмотрено создание следующих двумерных массивов: массива вершин  $V$ ; массива  $K$ , хранящего информацию о ключевых словах публикаций и массива  $U$ , содержащего код УДК для определения тематики публикаций и исследовательского направления работы. Для построения ориентированного графа сети цитирования формируется список дуг  $E$ . Данные массивы создаются в блоке



3. После этого в блоке 4 вершина  $v$  добавляется в массив  $V$ , в блоке 5 происходит добавление ключевых слов в массив  $K$  и УДК в массив  $U$ .

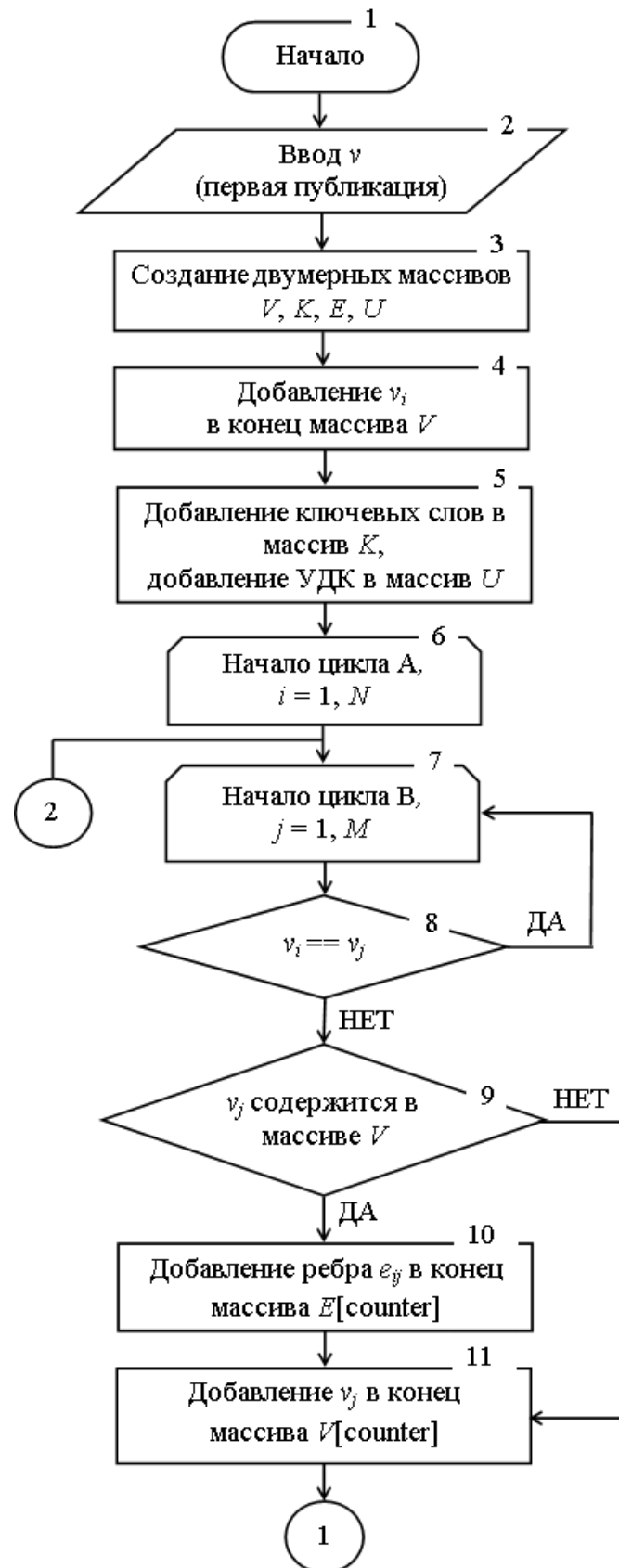


Рис. 2. Первая часть схемы алгоритма сбора данных для построения сети цитирования научных публикаций

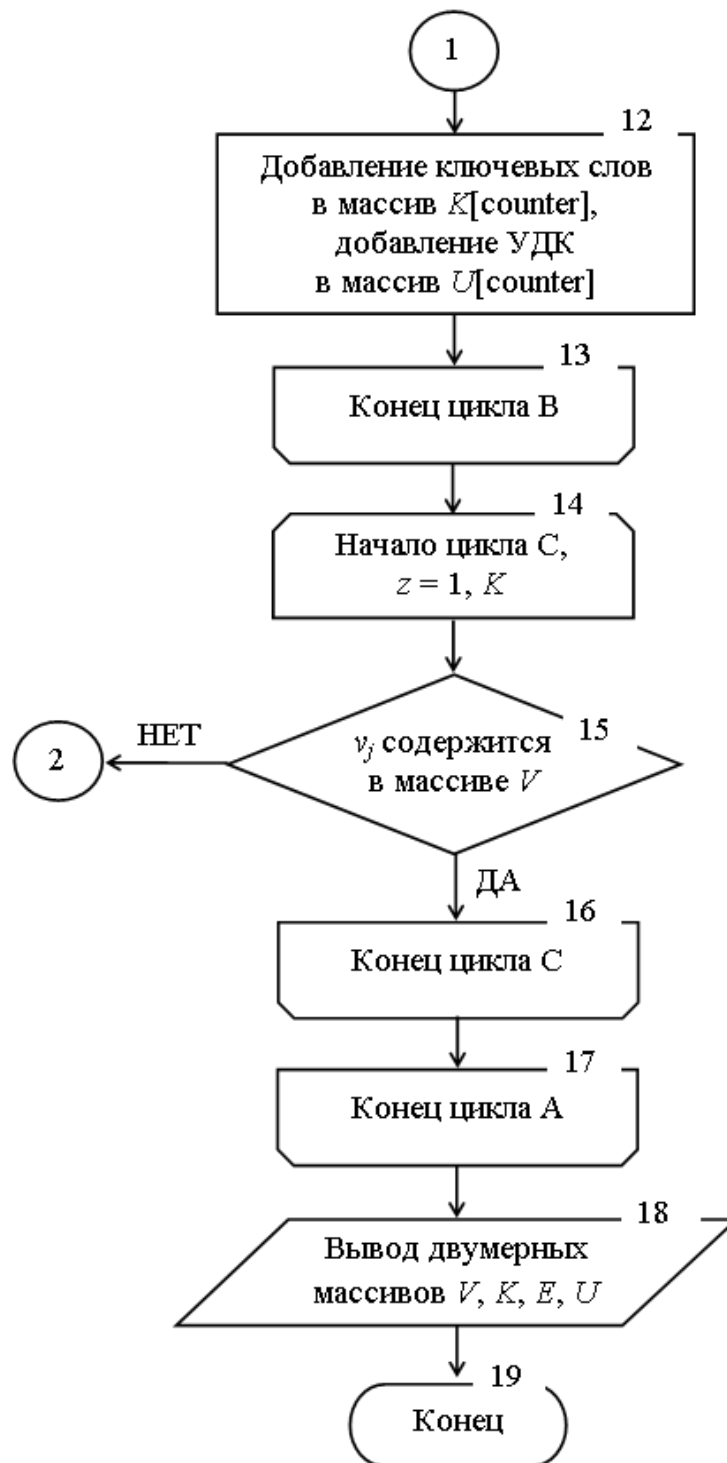


Рис. 3. Вторая часть схемы алгоритма сбора данных для построения сети цитирования научных публикаций

Алгоритм содержит один внешний цикл А (блоки 6-17) и два вложенных в него цикла В (блоки 7-13) и С (блоки 14-16). Цикл А производит перебор вершин  $v_i$  в массиве  $V$ . Цикл В осуществляет перебор по всем публикациям  $v_j$ , которые цитирует текущая публикация  $v_i$ , и добавляет их поочередно в конец массива  $V$ . В блоке 8 проверяется условие самоцитирования статей, в блоке 9 – проверяется наличие вершины  $v_j$  в массиве  $V$ . После этого в блоках 10-12 происходит заполнение созданных массивов. Цикл С перебирает публикации, которые цитируют

данную публикацию  $v_i$ , в блоке 15 проверяется наличие таких публикаций в массиве  $V$ . На каждом шаге по внешнему циклу  $A$  внутренний цикл  $B$  выполняется  $M$  раз, цикл  $C - K$  раз. Число  $M$  равно количеству публикаций, на которые ссылается публикация  $v_i$ , число  $K$  – количеству публикаций, которые ссылаются на публикацию  $v_i$ . Количество повторений  $N$  внешнего цикла  $A$  равно числу найденных научных публикаций в интернете на момент сбора данных либо может ограничиваться количеством публикаций, содержащихся только в одной исследуемой базе данных.

Процесс сбора данных выполняется циклично, пока не достигнуто условие постановления, заданное пользователем. При реализации программы, использующей представленный алгоритм сбора данных, предусмотрен ввод ограничения на количество найденных публикаций, если не требуется сбор данных обо всех существующих публикациях.

#### 4. Результаты экспериментов

Для реализации представленного на Рис. 2 и Рис. 3 алгоритма была разработана программа на высокоуровневом языке программирования Python с использованием библиотек BeautifulSoup и Selenium WebDriver. Для проведения экспериментов была выбрана база данных Научной электронной библиотеки eLIBRARY.RU. В результате работы программы формируется массив дуг, отображающий пары номеров вершин для связанных между собой публикаций. На основе данного массива построен ориентированный граф сети цитирования, вершинами которого являются научные публикации исследуемой базы данных. На Рис. 4 и Рис. 5 представлены подграфы графа сети цитирования Научной электронной библиотеки eLIBRARY.RU [9]. Информация о текущем состоянии цитирования извлечена 07.04.2020 г. Первый подграф состоит из 17 вершин, второй – из 10 вершин.

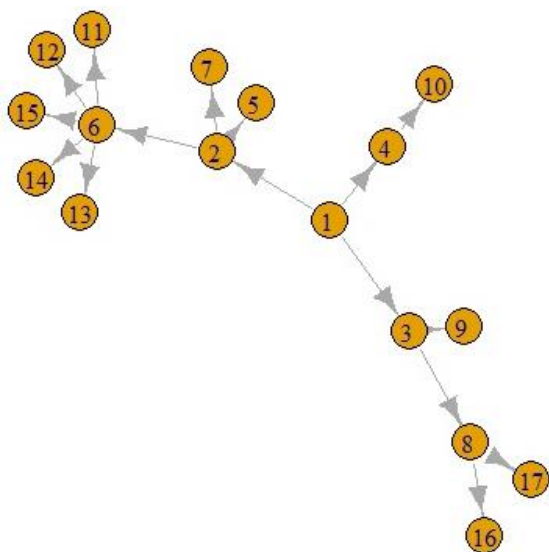


Рис. 4. Подграф сети цитирования Научной электронной библиотеки eLIBRARY.RU, состоящий из 17 вершин

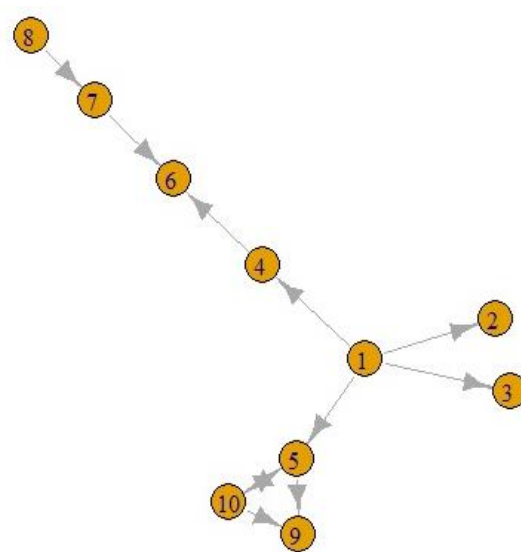


Рис. 5. Подграф сети цитирования Научной электронной библиотеки eLIBRARY.RU, состоящий из 10 вершин

Для ориентированных графов, моделирующих сети цитирования, предложен метод определения важности узлов сети с учетом релевантных мер центральности и профиля исследования [10]. Разработанный в [10] подход применим для решения задач оценки и ранжирования публикаций с учетом тематического направления публикаций, например, как дополнительный к предложенным в статье [11] вариантам.

Заметим, что на Рис. 5 представлен случай взаимного цитирования, когда две публикации одновременно цитируют друг друга. В Табл. 1 содержится информация об этих публикациях, соответствующих вершинам подграфа № 5 и № 10. В разработанной программе реализована возможность записи и хранения в таблице Microsoft Excel библиографических записей о цитируемых публикациях и изданиях.

**Таблица 1.** Связь номеров вершин с библиографическими данными публикаций

№	Название публикации	Авторы	УДК	Ключевые слова	Издание	Том, Вып., №	Год, стр.
5	Характерные особенности MCU-FR	Гуревич М.И., Калугин М.А., Олейник Д.С., Шкаровский Д.А.	621.03 9.17	уравнение переноса, программа MCU-FR, метод Монте-Карло, прецизионные расчеты	Российский научный центр "Курчатовский институт" (Москва)	№ 5	2016, 17-21
10	Тестирование программы MCU-FR применительно к расчетам критичности быстрых реакторов	Алексеев Н.И., Калугин М.А., Кулаков А.С., Олейник Д.С., Шкаровский Д.А.	621.03 0.5	быстрый ядерный реактор, бенчмарк эксперименты, прецизионный расчет, метод Монте-Карло, программа MCU-FR, банк данных MDBFR60	Российский научный центр "Курчатовский институт" (Москва)	№ 5	2016, 22-26

## 5. Заключение

На основе предложенного в статье алгоритма разработана программа для сбора данных о цитировании. Разработанная программа позволяет: получать сеть цитирования научных публикаций в виде ориентированного графа; останавливать (ставить на паузу) и продолжать сбор данных в любой момент времени без потери данных; получать библиографические данные любой публикации по соответствующему ей номеру вершины графа, а также список изданий, на которые ссылаются авторы.

В дальнейшем планируется разработка пользовательского приложения, обеспечивающего хранение и обработку информации, полученной на основе анализа сети цитирования научных публикаций.

Путем построения и анализа сети цитирования авторами решается ряд смежных прикладных задач, включая:

- автоматическое построение рефератов,
- подбор источников информации для научных исследований,
- подготовка научных обзоров литературы и др.

## ЛИТЕРАТУРА

1. Zhang Y., Wang M., Gottwalt F., Saberi M., Chang E. Ranking scientific articles based on bibliometric networks with a weighting scheme // *Journal of Informetrics*. 2019. V. 13, Iss. 2. P. 616-634.
2. Masaki Eto. Extended co-citation search: Graph-based document retrieval on a co-citation network containing citation context information // *Information Processing & Management*. 2019. V. 56, Iss. 6. P. 102046.
3. Гордукалова Г.Ф. Индекс цитирования в науке: цели использования, основные разновидности и ограничения // *Вестник Санкт-Петербургского государственного университета культуры и искусств*. 2014. № 2. С. 54-57.
4. Вахрушев М. В. Альтметрики, вебометрики и информетрики как взаимодополняющие направления в современной библиометрии // *Научные и технические библиотеки*. 2019. № 8. С. 67-76.
5. Шарабчиев Ю.Т. Проблемы "картографирования" научных направлений и выявления активных "точек роста" // *Медицинские новости*. 1996. № 10. С. 44-52.
6. Бредихин С.В., Ляпунов В.М., Щербакова Н.Г., Юргенсон А.Н. // *Проблемы информатики*. 2016. № 1. С. 39-57.
7. Пахомов И.А., Юдин Е.Б. Алгоритм получения сети ссылок веб-страниц на основе поиска в ширину // *Динамика систем, механизмов и машин*. 2016. № 3. С. 35-38.
8. Pakhomov I.A., Yudin E.B. Detection of structural information about web network // *Information technology and control automation*. Omsk. 2016. P. 240 – 244.
9. Научная электронная библиотека eLIBRARY.RU. URL: <https://elibrary.ru/defaultx.asp> (дата обращения 07.04.20).
10. Ольгина И.Г. Метод определения важных узлов сети с учетом релевантных мер центральности // *Сборник материалов VII Международной научной конференции, посвященной памяти С.С. Ефимова*. Омск. 2019. С. 144-146.
11. Аристова Н.И., Ицкович Э.Л. Предложения по оценке статей и ранжированию российских журналов, публикующих результаты прикладных научных разработок в области автоматизации промышленных предприятий // *Автоматизация в промышленности*. 2020. №1. С. 9-13.

---

**Ольгина** Инна Геннадьевна, ассистент кафедры «Математические методы и информационные технологии в экономике», [inna\\_olgina@mail.ru](mailto:inna_olgina@mail.ru)

**Пронин** Игорь Валерьевич, аспирант кафедры «Автоматика и системы управления», [pronin\\_igor7@rambler.ru](mailto:pronin_igor7@rambler.ru)

**Абдрахманов** Амангельды Нурзадаевич, аспирант кафедры «Автоматика и системы управления», [galiokpar@mail.ru](mailto:galiokpar@mail.ru)

УДК 004.896

## СИСТЕМА ПРЕДСКАЗАТЕЛЬНОГО ТЕХНИЧЕСКОГО ОБСЛУЖИВАНИЯ НА ОСНОВЕ ГЛУБОКИХ НЕЙРОННЫХ СЕТЕЙ

*Сай Ван Квонг, М. В. Щербаков*

Волгоградский государственный технический университет, Волгоград, Россия

**Аннотация.** В данной статье предлагается система предсказательного технического обслуживания, основанная на объединении сверточной нейронной сети (Convolutional Neural Network, CNN) для автоматического извлечения признаков и снижения размерности больших массивов многомерных сенсорных данных и рекуррентной нейронной сети долгой краткосрочной памяти (Long short-term memory, LSTM) для темпорального моделирования долговременных зависимостей между временными шагами, что позволило повысить точность прогнозирования отказов сложных многообъектных систем и автоматизировать принятие решений по техническому обслуживанию сложных многообъектных систем.

**Ключевые слова:** предсказательное техническое обслуживание, остаточный ресурс, глубокие нейронные сети, CNN, LSTM.

## PREDICTIVE MAINTENANCE SYSTEM BASED ON DEEP NEURAL NETWORKS

*Sai Van Kwong, M.V. Shcherbakov**Volgograd state technical university, Volgograd, Russia*

**Аннотация.** This paper proposes a predictive maintenance system based on the combination of convolutional Neural Network (CNN) for automatic feature extraction and dimensionality reduction of large arrays of multidimensional sensory data and Long Short Term Memory network (LSTM) for temporal modeling of long-term dependencies between time steps, this made it possible to increase the accuracy of predicting failures of complex multi-object systems and automate decision-making on maintenance of complex multi-object systems.

**Ключевые слова:** predictive maintenance, remaining useful life, deep neural networks, CNN, LSTM.

### 1. Введение

В современных экономических условиях проблема обеспечения надежности, эффективности и безопасности сложных промышленных систем является критической, поскольку отказы подобных систем могут сопровождаться огромными экономическими потерями, а иногда и привести к неопределимым потерям, связанным с гибелью людей. Это сказывается на повышении себестоимости производства продукции и снижает конкурентоспособность предприятий, работающих в различных предметных областях [1].

Обслуживание таких систем требует высоких материальных затрат, и все же не исключает возможного возникновения сбоев. Это свидетельствует о том, что задача обеспечения надежности сложных технических систем еще далека от своего решения.

В связи с этим в данной работе предлагается проактивная система предсказательного технического обслуживания, основанная на использовании инновационных технологий искусственного интеллекта (Artificial Intelligence, AI). В основном речь идет о создании предсказательных моделей для предотвращения отказов оборудования. Предложенная система основывается на объединении сверточной нейронной сети (*Convolutional Neural Network, CNN*) для автоматического извлечения признаков и снижения размерности больших массивов многомерных сенсорных данных и рекуррентной нейронной сети долгой краткосрочной памяти (*Long short-term memory, LSTM*) для темпериального моделирования долговременных зависимостей между временными шагами.

## 2. Предложенная система

На рис. 1 представлена предложенная система, состоящая из двух обученных гибридных нейронных сетей. Первая – это классифицирующая сеть, которая предсказывает принадлежность состояния системы к двум классам (нормальное состояние и деградирующее состояние). Вторая сеть является регрессионной и позволяет предсказывать величину остаточного ресурса (*Remaining useful life, RUL*) оборудования. Такая система была разработана потому, что регрессионная сеть дает более точные прогнозы, когда оборудование находится ближе к моменту отказа (деградация).

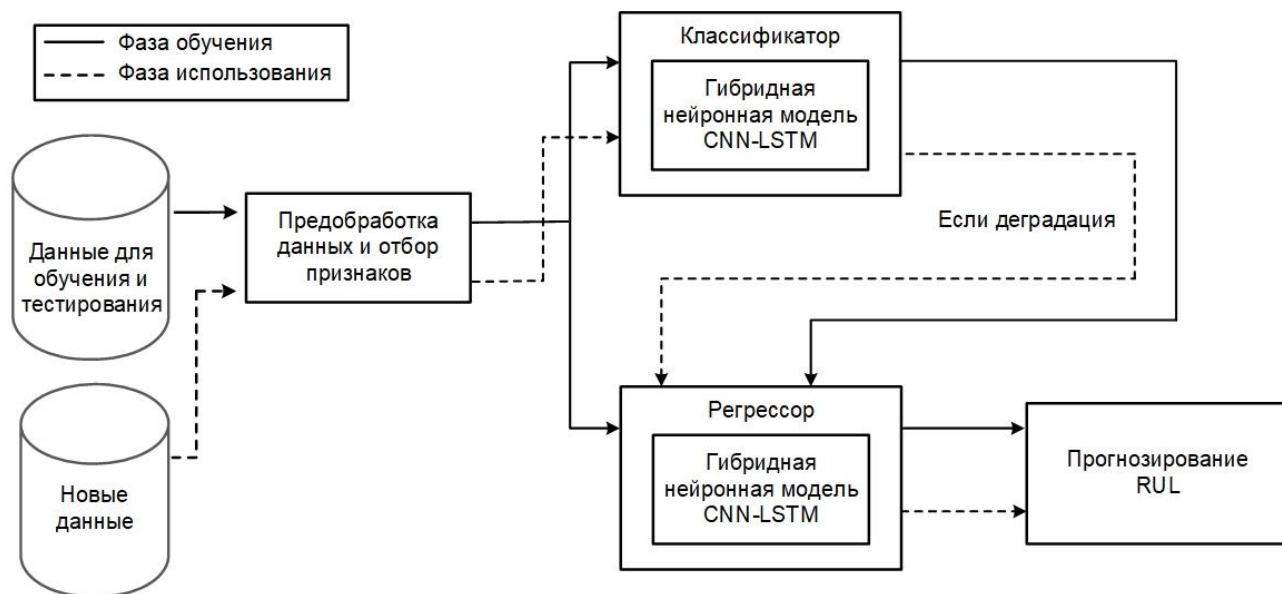


Рис. 1. Предложенная гибридная нейросетевая модель для предсказательного технического обслуживания

## 3. Результаты экспериментов

Разработанный метод был протестирован на наборе данных FD001 из широко известной базы данных C-MAPPS Dataset (NASA) [2], которая содержит информацию об отказах турбовентиляторных двигателей. Данный набор данных

состоит из обучающей выборки (100 двигателей) и тестовой выборки (100 двигателей). На основе обучающей выборки формируются и обучаются нейросетевые модели. Полученные модели затем используются для тестирования на тестовой выборке.

Точность классификации для предложенной гибридной модели на тестовых двигателях показана в Табл. 1.

**Таблица 1.** Точность классификатора (класс 0 – нормальное состояние оборудования, класс 1 – деградирующее состояние)

Класс	Точность
Класс 0	99%
Класс 1	98%

Заключительный этап в рабочем процессе предложенного метода заключается в прогнозировании величины остаточного ресурса двигателей следующим образом: если приведенная выше классификация относится к классу 1 (деградирующее состояние), то величина остаточного ресурса прогнозируется с использованием обученной регрессионной сети (алгоритм 1). Для измерения производительности регрессора используется средняя квадратичная ошибка (Root Mean Square Error, RMSE), которая рассчитывается по формуле:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (e_i)^2} = \sqrt{\text{mean}(e_i)^2},$$

где  $e_i = RUL_i - \hat{RUL}_i$ ,  $RUL_i$  и  $\hat{RUL}_i$  – фактическое и прогнозное значения  $RUL$  для  $i$ -ого двигателя соответственно.

---

#### Алгоритм 1: Окончательная ошибка модели

---

**Функция:** model\_error(x)

прогнозный класс = классификатор(x)

**if** фактический класс != 0 и прогнозный класс != 0 **then**

    Прогнозный RUL = регрессор(x)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (e_i)^2}$$

**end if**

**return** RMSE

**end функция**

---

Величина RMSE для двигателей в тестовом наборе составляет 7,98.

## 5. Выводы и заключение

Результаты вычислительных экспериментов на основе открытых данных NASA, показали, что гибридная нейросетевая модель CNN-LSTM дает высокую



точность. Полученные результаты и алгоритмы могут быть использованы в системах превентивного обслуживания оборудования с целью высоконадежной идентификации стадии его деградации и прогнозирования отказов сложных многообъектных систем.

*Работа выполнена при поддержке РФФИ, проект № 19-47-340010\_p\_a.*

## ЛИТЕРАТУРА

1. Сай Ван Квонг. Глубокие нейронные сети для предсказательного технического обслуживания // Моделирование, оптимизация и информационные технологии (МОИТ): Сетевой научный журнал. – 2019. – Т.7, № 4(27). – 11 с. – DOI: 10.26102/2310-6018/2019.27.4.011.
2. Bektas O., Jones J. A., Sankararaman S., Roychoudhury I., Goebel K. (2018). Reconstructing secondary test database from PHM08 challenge data set. Data in brief, 21, 2464-2469.

---

**Сай Ван Квонг**, аспирант, svcuonghvktqs@gmail.com

**Щербakov** Максим Владимирович, д.т.н., профессор кафедры «Системы автоматизированного проектирования и поискового конструирования», maxim.shcherbakov@vstu.ru

УДК 519.245:004.421.5

## ПОКАЗАТЕЛЬ СТАТИСТИЧЕСКОЙ ЗНАЧИМОСТИ ПОДГРАФОВ $R'$ В ЗАДАЧЕ ВЫЯВЛЕНИЯ СЕТЕВЫХ МОТИВОВ

М. Н. Юдина

Омский государственный технический университет, Омск, Россия

**Аннотация.** *Выполняется анализ показателей статистической значимости подграфов при решении задачи выявления сетевых мотивов в графах больших сетей. Предлагается новый обобщающий показатель значимости подграфов. Показатель предназначен для выявления сетевых мотивов, когда для расчета частот встречаемости подграфов используются статистические методы. При использовании предлагаемого показателя тот факт, что подграф в исследуемом графе сети встречается чаще, чем в рандомизированных версиях этого графа (признак сетевого мотива), устанавливается статистически достоверно.*

**Ключевые слова:** *сетевые мотивы, анализ больших сетей, показатели статистической значимости подграфов, Z-Score, показатель  $R$ , статистический расчет частот встречаемости подграфов.*

## A MEASURE $R'$ TO ESTIMATE STATISTICAL SIGNIFICANCES OF SUBGRAPHS WHILE NETWORK MOTIFS DISCOVERING

M. N. Yudina

Omsk State Technical University, Omsk, Russia

**Abstract.** *An analysis of subgraph significance statistical measures while network motifs discovering the problem is given. A new significance measure  $R'$  is proposed. The proposed measure  $R'$  is generalization of the well-known significance measure  $R$ . The  $R'$  can be useful while network motifs discovering does by statistical methods to calculate subgraphs frequencies. The  $R'$  allows us to statistically reliably establish that the subgraph is more frequent in the graph than in randomized versions of the graph.*

**Keywords:** *network motifs, network analysis, statistical significance of a subgraph, Z-Score and  $R$  statistical measures, statistical methods of subgraphs frequencies calculation.*

### 1. Введение

Изучением общих свойств больших сетей занимается новая отрасль науки – «Наука о сетях» (Network Science). Причем, если 20 лет назад большое внимание уделяли структурным характеристикам сети в целом (безмасштабности сетей [1], эффекту тесного мира, правилу шести рукопожатий [2]), то сейчас достаточно много внимания уделяют локальным структурным взаимодействиям. Ведь они могут сказать о сети не меньше.

Одним из подходов, разработанным в «Науке о сетях», является выявление в графах сетей так называемых сетевых мотивов. Сетевые мотивы на  $k$ -вершинах ( $k$ -мотивы) выявляются при сравнении частот встречаемости связанных  $k$ -

вершинных подграфов, относящихся к различным классам изоморфизма, в исследуемом графе сети и в его рандомизированных версиях. Анализ больших сетей на основе выявления сетевых мотивов привел к открытию новых белковых взаимодействий [3], новых способов диагностирования рака [4], новых подходов к управлению обменом информации в технических сетях [5], и к выявлению «скрытых особенностей» социальных сетей.

Преобладание тех или иных сетевых мотивов является особенностью всякой большой сети, отличающей ее от других сетей, в первую очередь, от сетей другой природы. Так, выявление FULLY CONNECTED TRIAD-мотива (полный подграф на трех вершинах) является признаком графов информационных сетей интернета, для графов молекулярных сетей клетки характерны FEED-FORWARD LOOP- (петля обратной связи) и BI-FAN-мотивы (Рис.1), в графах трофических сетей обнаруживаются THREE-CHAIN- и BI-PARALLEL-мотивы и т.д.

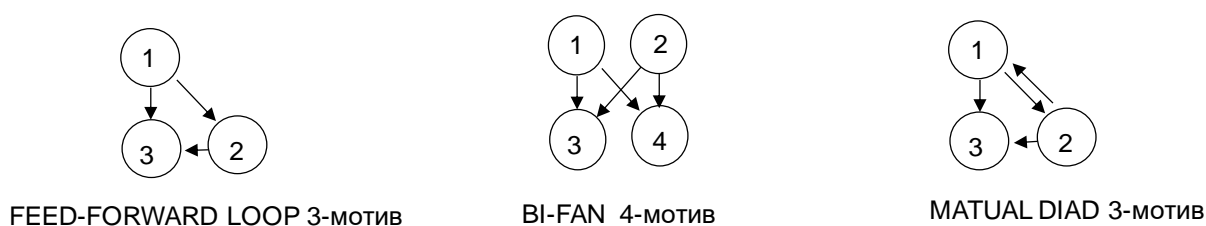


Рис. 1. Примеры сетевых мотивов

Задачу нахождения  $k$ -мотивов в графе  $G$  можно разбить на следующие три подзадачи.

1. Получение множества  $G_R$  рандомизированных версий графа  $G$ .
2. Подсчет в графе  $G$  и в каждом графе множества  $G_R$  частот встречаемости подграфов, относящихся к различным классам изоморфизма. При этом решается задача определения классов изоморфизма для найденных подграфов на  $k$  вершинах.
3. Анализ сетевых мотивов на основе их встречаемости в графе  $G$  и в графах множества  $G_R$ .

Третью подзадачу решают путем использования различных показателей статистической значимости подграфов. Традиционно [6, 7] сетевые мотивы исследователи больших сетей определяют исходя из показателей статистической значимости подграфов  $R$  и  $Z$ -score.

Показатель статистической значимости  $R$  определяется как отношение абсолютной частоты встречаемости  $n_G$  исследуемого подграфа в графе сети к среднему значению абсолютной частоты встречаемости этого же подграфа  $M[n_{GR}]$  в рандомизированных версиях графа:

$$R = n_G / M[n_{GR}]. \quad (1)$$

Стандартизированная оценка ( $Z$ -score) определяется по следующей формуле:

$$Z\text{-score} = \frac{n_G - M[n_{GR}]}{\sigma[n_{GR}]}, \quad (2)$$

где  $M[n_{GR}]$  и  $\sigma[n_{GR}]$  – математическое ожидание и стандартное отклонение абсолютной частоты встречаемости подграфа в рандомизированных версиях графа.

Сетевыми мотивами считаются такие подграфы, которые относятся к классам изоморфизма с показателем  $Z\text{-score}$ , превышающем некоторое пороговое значение. Сетевыми мотивами называются подграфы, для которых  $Z\text{-score} > 2$  и  $R > 1.1$  [7].

## 2. Постановка задачи

Выявление сетевых мотивов достаточно трудоемкий процесс, причем самой трудоемкой его частью является расчет встречаемости подграфов. Точный расчет встречаемости подграфов был обоснован при зарождении этого подхода в 2002 году [6], когда изучались, например, графы белковых сетей, содержащие от нескольких сотен до нескольких тысяч узлов. Сейчас исследуемые графы сетей белковых взаимодействий содержат сотни тысяч узлов, точный расчет частот встречаемости подграфов оказывается невозможным. Также выявление сетевых мотивов используется при анализе сетей интернета, которые по своим размерам часто превосходят сети белковых взаимодействий. Так, в Табл. 1 представлено время расчета частот встречаемости в некоторых графах сетей интернета. Данные о графах Google Plus, Twitter, Email Enron, Email\_AU-All получены из базы графов стенфордского университета (<http://snap.stanford.edu/data/>). Граф VKOmgту описан в работе [7]. Отсутствие данных в соответствующих ячейках Табл. 1 означает, что при использовании программы получить результат не удалось за пять дней работы. Использовалась самая известная программа, причем подходящая для расчетов именно больших сетей – программа Fanmod, а также библиотека igraph для среды матричных вычислений R.

**Таблица 1.** Время расчета (в секундах) частот встречаемости подграфов в графах сетей

Граф	Библиотека igraph для среды R		FANMOD	
	Подграфы на трех вершинах	Подграфы на четырех вершинах	Подграфы на трех вершинах	Подграфы на четырех вершинах
Google Plus	39442	–	–	–
Twitter	76	35217	1076	–
Email Enron	23	10882,83	27	11646
Email_AU-All	40	38508	136	272158
VKOmgту	0,22	19	1	93

Увеличение размеров графов исследуемых сетей привело к развитию статистических методов расчета, позволяющих получать статистические оценки для частот встречаемости подграфов (см. Табл. 1). Следует заметить, что большинство статистических методов не позволяют исследовать смешанные графы, а разработаны для проведения расчетов только в неориентированных графах. Кроме того, многие методы не позволяют получать оценки абсолютных частот встречаемости подграфов, а только оценки их относительных частот встречаемости.

**Таблица 2.** Методы статистического расчета частот встречаемости подграфов

Название метода	Ссылка на файл для запуска	Год создания	Ограничение на размер $k$ подграфов	Расчет относительных / абсолютных ( $C_i^A / n_i^A$ ) частот встречаемости	Возможность расчета в смешанных графах
ESA [8]	[9]	2004	Нет	$C_i^A$	Да
RAND-ESU [10]	[11]	2005	Нет	$n_i^A$	Да
GUISE [12]	[13]	2012	5	$C_i^A$	Нет
Wedge Sampling [14]	[15]	2013	3	$n_i^A$	Нет
GRAFT [16]	[17]	2014	5	$C_i^A$	Нет
$k$ -profile sparsifier [18]	[19]	2016	4	$C_i^A$	Нет
MOSS-5 [20]	[21]	2018	5	$n_i^A$	Нет
MFS [22]	[23]	2019	3, 4	$n_i^A$	Да

Заметим, что использование в формулах (1) и (2) относительных частот встречаемости подграфов вместо абсолютных частот встречаемости некорректно (хотя и используется без объяснения во многих программных продуктах, таких как FANMOD, MFinder). В этом случае вычисляется другая величина. Причина значительного отличия этих величин в том, что суммарное число всех подграфов в исследуемом графе обычно сильно отличается от суммарного числа всех подграфов в рандомизированных версиях этого графа. Для большинства исследуемых нами сетей в рандомизированном графе число подграфов увеличивается, как правило, на порядок.

С другой стороны, даже для методов, позволяющих получать оценки абсолютных частот встречаемости подграфов, подстановка статистических оценок в формулу (1) и (2) также является приближенной операцией. Ведь статистические методы имеют определенную погрешность.

Возникает необходимость разработки таких показателей статистической значимости, которые позволят выявлять сетевые мотивы с использованием статистических методов для расчета частот встречаемости подграфов.

### 3. Показатель $R'$

Предложим обобщение показателя  $R$ , который можно применять при получении статистических оценок для частот встречаемости подграфов. Возможность это сделать дает метод MFS [22] или метод RAND-ESU, реализованный в известных программах FANMOD [11] и igraph [24]. Одна из главных особенностей этих методов заключается в возможности получать не только точечные, но и интервальные оценки для частот встречаемости подграфов, как в исследуемом графе, так и в рандомизированной его версии.

Заметим, что при получении статистической оценки в исследуемом и рандомизированном графе возможно три варианта (Рис.2), при которых значение точечной оценки для частоты встречаемости подграфа в исследуемом графе превышает значение оценки для частоты встречаемости в рандомизированном графе (что соответствует понятию сетевого мотива):

1) точечная оценка  $n_R$  для частоты встречаемости подграфа в рандомизированном графе не попадает в доверительный интервал  $n_G \pm 3\sigma_G$  для частоты встречаемости в исследуемом графе, но доверительные интервалы  $n_G \pm 3\sigma_G$  и  $n_R \pm 3\sigma_R$  имеют общие точки;

2) доверительные интервалы для частоты встречаемости подграфа в рандомизированном графе  $n_R \pm 3\sigma_R$  и в исследуемом графе  $n_G \pm 3\sigma_G$  не имеют общих точек;

3) точечная оценка для частоты встречаемости подграфа в рандомизированном графе  $n_R$  попадает в доверительный интервал  $n_G \pm 3\sigma_G$  для частоты встречаемости в исследуемом графе, но доверительные интервалы  $n_R \pm 3\sigma_R$  и  $n_G \pm 3\sigma_G$  имеют общие точки.

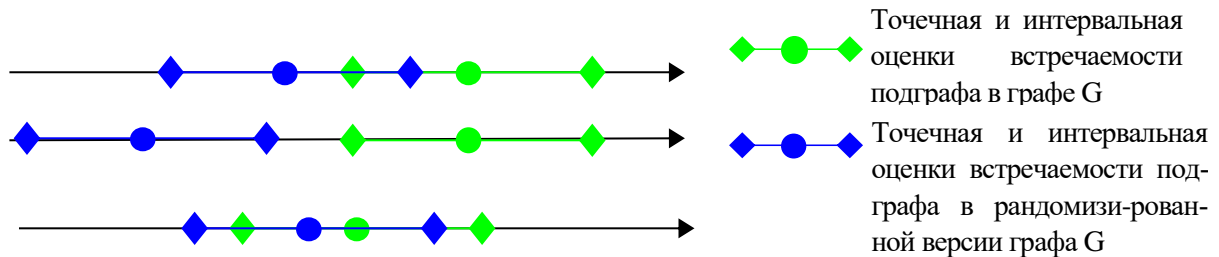


Рис. 2. Иллюстрация возможных соотношений статистических оценок для частот встречаемости подграфов в исследуемом графе и в его рандомизированной версии

Для вводимого показателя выберем наиболее строгое отношение между получаемыми статистическими оценками для частот встречаемости подграфа в графе сети и в его рандомизированной версии (случай 2). Определим показатель значимости  $R'$  подграфа следующей формулой:

$$R' = M[n_G - 3\sigma_G] / M[n_R + 3\sigma_R] \quad (3)$$

где  $n_G$ ,  $n_R$  – оценки абсолютных частот встречаемости подграфа, полученные при расчете в графе и в его рандомизированной версии соответственно;  $\sigma_G$ ,  $\sigma_R$  – соответствующие среднеквадратичные отклонения, характеризующие точность полученных оценок. Предложенная оценка обобщает известный показатель  $R$  и показывает, что число подграфов в исследуемом графе «статистически достоверно» превышает это число в его рандомизированных версиях.

При выявлении сетевых мотивов в больших графах можно использовать расчеты лишь одной оценки для частоты встречаемости подграфов в графе G и одной в его рандомизированной версии, тогда выражение (3) запишется как

$$R' = (n_G - 3\sigma_G) / (n_R + 3\sigma_R).$$

#### 4. Результаты экспериментов

В Табл. 3 приведены результаты исследования сетевых мотивов в ориентированном графе GenReg, расчет  $R'$  выполнен в программе MFSView. Граф GenReg описывает регуляторную сеть генов, задействованных в появлении рака коры надпочечников у человека [25]. Использовался один расчет частот встречаемости подграфов по графу GenReg и 100 расчетов по его рандомизированным версиям. Показатели статистической значимости подграфов, соответствующие сетевым мотивам, выделены серым цветом.

**Таблица 3.** Результаты исследования сетевых мотивов в графе GenReg

$id$	$n_G$	$M[n_R]$	$R$	$R'$	$Z\text{-Score}$
3	9,1750E+08	1,0711E+09	0,8566	0,7944	-0,2125
7	2,4150E+09	2,5841E+09	0,9346	0,8934	0,1104
12	2,4994E+09	2,8199E+09	0,8864	0,8426	-0,0296
13	6,9143E+09	8,8490E+09	0,7814	0,7619	-0,4433
14	2,0649E+08	9,1315E+07	2,2613	1,9320	4,9079
17	2,1207E+08	8,8373E+07	2,3997	2,0491	5,5049
19	4,1816E+07	3,9063E+07	1,0705	0,6887	0,6199
20	8,9347E+06	1,2634E+06	7,0722	2,6975	36,8311
24	4,4803E+09	4,7995E+09	0,9335	0,9055	0,0953
29	1,2684E+09	2,2750E+09	0,5575	0,5221	-1,2077
35	1,4025E+08	5,1712E+07	2,7122	2,2189	6,1270
41	5,2174E+09	6,9174E+09	0,7542	0,7315	-0,5620
42	1,3556E+08	7,0856E+07	1,9132	1,5896	3,5735
47	3,7075E+08	1,7072E+08	2,1717	1,9336	4,6475
52	3,4448E+07	3,8433E+07	0,8963	0,5603	-0,3249
58	1,0032E+07	1,4154E+06	7,0874	2,8405	26,2518
66	3,1451E+07	1,8959E+07	1,6589	0,9470	1,9731
67	1,3481E+07	2,0614E+06	6,5397	2,9901	43,4739
76	6,8123E+09	7,2798E+09	0,9358	0,9154	0,0830
77	6,5371E+08	2,4802E+08	2,6357	2,4031	5,7933
78	1,2740E+07	1,1991E+06	10,6245	4,2876	38,4280
80	1,5866E+07	2,1128E+06	7,5095	3,5385	32,9139
82	2,2976E+07	2,5780E+06	8,9125	4,5779	44,1505
83	6,2056E+06	5,4260E+04	114,3679	14,5793	31,4693

В Табл. 4 представлены результаты расчетов статистических оценок для частот встречаемости подграфов во фрагменте социальной сети Вконтакте [7]. Как можно видеть из Табл. 4 при использовании «стандартных значений»  $R > 1,1$ ,

$R' > 1,1$  выявляются те же сетевые мотивы, что и при использовании  $Z\text{-score} > 2$ , это подграфы с идентификаторами 142, 203, 205, 217.

**Таблица 4** – Результаты выявления сетевых мотивов в графе VKOmgTu

id	$n_G$	$3\sigma[n_G]$	$M[n_R]$	$M[3\sigma[n_R]]$	$R$	$R'$	$Z\text{-Score}$
94	3,6431E+07	3,7112E+04	3,9414E+07	1,7724E+04	0,9243	0,9202	-20,112
125	1,9361E+07	4,0556E+04	2,0889E+07	1,8763E+04	0,9268	0,9185	-7,6054
142	3,2168E+06	2,1799E+04	7,5442E+05	1,0295E+04	4,2640	4,0130	128,26
203	9,2010E+04	5,0490E+03	3,5203E+04	2,8163E+03	2,6137	1,7608	8,9843
205	2,1147E+05	5,8759E+03	8,7283E+03	1,1230E+03	24,2282	16,0236	750,97
217	3,1995E+04	2,2920E+03	5,6415E+01	1,1283E+02	567,1389	63,6079	60424

## 5. Обсуждение результатов

Полученные результаты показывают, что предложенный показатель значимости  $R'$  обобщает известный показатель значимости  $R$  и может быть использован при выявлении сетевых мотивов при применении статистических методов для расчета частот встречаемости подграфов. Особенность предложенного показателя  $R'$  заключается в том, что он позволяет статистически достоверно выявлять сетевые мотивы даже по двум статистическим расчетам частот встречаемости подграфов: один – в исследуемом графе сети, другой – в его рандомизированной версии.

## 6. Выводы и заключение

В статье рассмотрены известные статистические показатели значимости подграфов  $R$  и  $Z\text{-Score}$ , используемые при выявлении в графе сети так называемых сетевых мотивов. Показатели значимости  $R$  и  $Z\text{-Score}$ , по определению, используют результаты вычислений абсолютных частот встречаемости подграфов в исследуемом графе и в его рандомизированных версиях. Однако в ряде программ разработки применяют эти показатели, «подменяя» абсолютные частоты встречаемости относительными частотами встречаемости. В других случаях точные значения частот встречаемости «подменяются» их статистическими оценками. При этом сами показатели статистической значимости не развиваются, а определяются различные смыслы использования существующих показателей, которые даже для специалистов могут остаться «скрытыми».

В данной работе исследуется изменение смысла показателя значимости подграфа  $R$ , если для расчета частот встречаемости вместо точного расчета используются статистические методы. Предложенный показатель значимости  $R'$  обобщает известный показатель значимости  $R$  и позволяет статистически достоверно выявлять сетевые мотивы при использовании статистических методов расчета частот встречаемости подграфов.



## ЛИТЕРАТУРА

1. Barabasi A.L., Albert R. Emergence of scaling in random networks. *Science*. 1999. Vol. 286. P. 509–512.
2. Newman, M.E.J. The Structure and Function of Complex Networks. *SIAM Review*. 2003. Vol. 45. №2. P. 167–256.
3. Albert I., Albert R. Conserved network motifs allow protein–protein interaction prediction *Bioinformatics*. 2004. Vol. 20. №18. P. 3346–3352.
4. Chen, L., et al. Identification of breast cancer patients based on human signaling network motifs. *Scientific Reports*. 2004. Vol. 3368. № 3. P.1–17.
5. Zeng T., Semiar O., Saad W. Spatio-temporal motifs for optimized vehicle-to-vehicle (v2v) communications. In: 2018 International Conference on Computing, Networking and Communications (ICNC). March 2018. Pp. 789–794.
6. Milo R. Network motifs: Simple building blocks of complex networks / R. Milo [et al.] // *Science*. –2002. – V. 594, № 298. – P. 824–827.
7. Юдина М. Н. Узлы в социальных сетях: меры центральности и роль в сетевых процессах. Омский научный вестник. Сер. Приборы, машины и технологии. 2016. Т. 148, №4. С. 161–165.
8. Kashtan N. et al. Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. *Bioinformatics*. 2004. V. 20, №11. P. 1746–1758.
9. Alon U. Network Motif Software. Режим доступа: <https://www.weizmann.ac.il/mcb/UriAlon/download/network-motif-software/>.
10. Wernicke S., Rasche F. FANMOD: a tool for fast network motif detection // *Bioinformatics*. 2006. V. 22, №9. P. 1152–1153.
11. Wernicke S. FANMOD: a tool for fast network motif detection. – Режим доступа: <http://theinf1.informatik.uni-jena.de/motifs/>.
12. Bhuiyan M.A., Rahman M, Hasan M.A. Guise: Uniform sampling of graphlets for large graph analysis. *Proceedings of the 12th International Conference on Data Mining (ICDM)*. 2012. P. 91–100.
13. Bhuiyan M.A., Rahman M, Hasan M.A. GUISE: Uniform Sampling of Graphlets for Large Graph Analysis – Режим доступа: <https://github.com/DMGroup-IUPUI/GUISE-Source>.
14. Seshadhri C., Pinar A., Kolda T.G. Triadic measures on graphs: The power of wedge sampling. In *Proceedings of the SIAM International Conference on Data Mining*. SIAM. 2013. P. 10–18.
15. Kolda T., Pinar A., Seshadhri C. Triadic Measures on Graphs: The Power of Wedge Sampling. Режим доступа: <http://www.sandia.gov/~tgkolda/feastpack/>.
16. Rahman M., Bhuiyan M.A., Hasan M.A. Graft: An efficient graphlet counting method for large graph analysis. *IEEE Transactions on Knowledge and Data Engineering*. 2014. V. 26, №10. – P. 2466–2478.
17. Rahman M., Bhuiyan M.A., Hasan M.A. GRAFT: an approximate graphlet counting algorithm for large graph analysis. Режим доступа: <https://github.com/DMGroup-IUPUI/GRAFT-Source>.
18. Elenberg E.R., et al. Distributed estimation of graph 4-profiles. In the *Proceedings of the 25th International Conference on World WideWeb*. 2016. P. 483–493.
19. Elenberg E.R. GraphLab PowerGraph implementation of 4-profile counting. Режим доступа: <https://github.com/eelenberg/4-profiles>.
20. Wang P. et al. MOSS-5: A fast method of approximating counts of 5-node graphlets in large graphs. *IEEE Transactions on Knowledge and Data Engineering*. 2018. V. 30, №1. P. 73–86.
21. Wang P. MOSS-5: Fast Method of Approximating Counts of 5-Node Graphlets in Large Graphs. Режим доступа: <http://nskeylab.xjtu.edu.cn/dataset/phwang/code/mosscode.zip>.
22. Yudina M.N., Zadorozhnyi V.N., Yudin E.B. Mixed Random Sampling of Frames method for counting number of motifs // *Journal of Physics: Conference Series*. – 2019. – V. 1260. – P. 022013
23. Yudina M.N., MFSView. Режим доступа: <https://github.com/MNYudina/MFSView>.
24. igraph – Network analysis software Режим доступа: <https://igraph.org/>
25. Le S, Riva A, Tran D. A high-performance pipeline for genome-wide network reconstruction from gene expression data [Электронный ресурс] / S. Le, A. Riva, D. Tran // *Proceedings of NetBio SIG, ISMB – Orlando, FL, USA, 2016*. – Режим доступа: <https://doi.org/10.7490/f1000research.1115059.1>.

## Авторский указатель: А-Я








<b>Абдрахманов</b> Амангельды Нурзадаевич, аспирант кафедры АиСУ ОмГУПС	<b>56, 119</b>
<b>Бадрызлов</b> Владимир Александрович, ст. пр. кафедры ММиИТЭ ОмГТУ	<b>61</b>
<b>Баркова</b> Ирина Михайловна, магистрант ОмГТУ	<b>71, 83</b>
<b>Бахмутский</b> Юрий Андреевич, заведующий кафедрой ММиИТЭ ОмГТУ	<b>71</b>
<b>Берестинов</b> Амиржан Ахметжанович, студент ЮУрГУ	<b>76</b>
<b>Будзинская</b> Елизавета Олеговна, студент ОмГТУ	<b>80</b>
<b>Галдина</b> Юлия Павловна, магистрант ОмГТУ	<b>71, 83</b>
<b>Галич</b> Юлия Геннадьевна, аспирант ОмГУПС	<b>94</b>
<b>Гегечкори</b> Евгений Трдатович, к.т.н., доцент кафедры ММиИТЭ ОмГТУ	<b>88</b>
<b>Доценко</b> Елена Валерьевна, магистрант ОмГТУ	<b>88</b>
<b>Задорожный</b> Владимир Николаевич, д.т.н., профессор кафедры ММиИТЭ ОмГТУ, профессор кафедры АиСУ ОмГУПС	<b>7, 27, 94, 107, 112</b>
<b>Захаренкова</b> Татьяна Романовна, ассистент кафедры ММиИТЭ ОмГТУ	<b>7, 107</b>
<b>Землянов</b> Антон Алексеевич, студент КГТУ	<b>50</b>
<b>Калиберда</b> Елена Анатольевна, к.т.н., доцент кафедры ММиИТЭ ОмГТУ	<b>80</b>
<b>Кравченко</b> Ксения Владимировна, ст. пр. кафедры ММиИТЭ ОмГТУ	<b>71, 83</b>
<b>Кулганатов</b> Аскар Зайдакбаевич, студент ЮУрГУ	<b>76</b>
<b>Маркова</b> Марина Петровна, аспирант ОмГУПС	<b>107</b>
<b>Наумов</b> Владимир Аркадьевич, д.т.н., профессор кафедры ВРП КГТУ	<b>50</b>
<b>Новиков</b> Артем Андреевич, студент ОмГТУ	<b>80</b>
<b>Ольгина</b> Инна Геннадьевна, аспирантка кафедры ММиИТЭ ОмГТУ	<b>119</b>
<b>Омельченко</b> Сергей Дмитриевич, студент ЮУрГУ	<b>76</b>
<b>Пагано</b> Микеле, профессор University of Pisa, Italy	<b>7</b>
<b>Пронин</b> Игорь Валерьевич, аспирант кафедры АиСУ ОмГУПС	<b>119</b>
<b>Сай Ван Квонг</b> , аспирант ВолгГТУ	<b>127</b>
<b>Щербаков</b> Максим Владимирович, д.т.н., профессор кафедры САПРиПК ВолгГТУ	<b>127</b>
<b>Юдин</b> Евгений Борисович, к.т.н., доцент кафедры ММиИТЭ ОмГТУ	<b>27</b>
<b>Юдина</b> Мария Николаевна, ассистент кафедры ММиИТЭ ОмГТУ	<b>27, 112, 131</b>

## НИЛ ИМСАИТ

Научно-исследовательская лаборатория Имитационного моделирования,  
системного анализа и информационных технологий

[http://](http://imsait.github.io/) <https://imsait.github.io/>

при кафедре ММиИТЭ ОмГТУ оказывает следующие услуги

	<b>Имитационное моделирование и оптимизация</b> Разрабатываем на языках GPSS, AnyLogic, C++ и т.д. имитационные модели систем с очередями, технологических процессов, мероприятий в сфере обслуживания и т.д.
	<b>Обучение</b> Обучаем имитационному моделированию и системному анализу, разрабатываем математические модели сложных систем и процессов, выполняем их оптимизацию.
	<b>Исследование и проектирование больших сетей</b> Проводим мониторинг социальных сетей, сетей финансовых операций, разрабатываем оптимальные стратегии сетевого маркетинга, исследуем телекоммуникационные сети.
	<b>Моделирование и оптимизация транспортных сетей</b> Выполняем многоуровневое распределенное аналитико-имитационное моделирование автотранспортных сетей как единого целого с учетом стохастической динамики транспортных потоков. Оптимизируем транспортные сети и алгоритмы управления транспортными потоками.
	<b>Дискретная оптимизация</b> Оптимизируем компоновку оборудования, раскрой материалов, маршруты перевозок, решаем задачи на графах.
	<b>NVIDIA CUDA</b> Используя графические процессоры, многократно ускоряем свои программы и программы заказчиков, предназначенные для научного, промышленного или прикладного применения.
	<b>Распределенные вычисления</b> Разрабатываем параллельные алгоритмы решения сложных задач и программы для вычислительного кластера

### Наши партнеры:

- Группа компаний **ИНКОМ**, г. Томск;
- Национальный исследовательский Томский политехнический университет **ТПУ**, Институт кибернетики;
- Санкт-Петербургский институт информатики и автоматизации РАН **СПИИРАН**;
- Сургутский Государственный Университет **СурГУ**;
- ООО «Элина-компьютер», г. Казань;
- ЗАО «Автоматика-Д», г. Омск;
- МУ г. Омска «Управление информационно-коммуникационных технологий»;
- Некоммерческое партнерство «Национальное общество имитационного моделирования», г. Санкт-Петербург;
- Санкт-Петербургский государственный электротехнический университет **СПбГЭТУ** им. В.И. Ульянова («ЛЭТИ»).
- Пизанский университет (University of Pisa), Италия.

*Научное издание*

СИСТЕМЫ УПРАВЛЕНИЯ,  
ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ  
И МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Материалы  
II Всероссийской научно-практической конференции  
с международным участием  
(Омск, 19–20 мая 2020 года)

В двух томах

Том I

Общая научная редакция *В. Н. Задорожного*

Компьютерная верстка *Т. Р. Захаренковой*

Ответственность за содержание материалов несут авторы

Печатается в авторской редакции

Для дизайна обложки использованы материалы  
из открытых интернет-источников

Подписано в печать 20.07.20. Формат 60×84<sup>1</sup>/<sub>16</sub>. Бумага офсетная.

Отпечатано на дупликаторе. Усл. печ. л. 8,75. Уч.-изд. л. 8,75.

Тираж 100 экз. Заказ 329.

---

Издательство ОмГТУ. 644050, г. Омск, пр. Мира, 11; т. 23-02-12.  
Типография ОмГТУ.