УДК 519.217.2, 519.614.2

Вокруг степенного закона распределения компонент вектора PageRank. Часть 1. Численные методы поиска вектора PageRank*

А.В. Гасников[1,2], Е.В. Гасникова[1], П.Е. Двуреченский[2,3], А.А.М. Мохаммед[1], Е.О. Черноусова[1]

[1]Московский физико-технический институт, Институтский пер., 9, Долгопрудный, Московская обл., 141700

[2]Институт проблем передачи информации им. А.А. Харкевича Российской академии наук, Большой Каретный пер., 19, строение 1, Москва, 127051

[3]Институт прикладного анализа и стохастики им. К. Вейерштрасса, Моренштрассе, 39, Берлин, Германия, 10117

E-mails: gasnikov.av@mipt.ru (Гасников А.В.), egasnikova@yandex.ru (Гасникова Е.В.), pavel.dvurechensky@wias-berlin.de (Двуреченский П.Е.), a.nafea@science.sohag.edu.eg (Мохаммед А.А.М.), lena-ezhova@rmbler.ru (Черноусова Е.О.)

Гасников А.В., Гасникова Е.В., Двуреченский П.Е., Мохаммед А.А.М., Черноусова Е.О. Вокруг степенного закона распределения компонент вектора PageRank. Часть 1. Численные методы поиска вектора PageRank // Сиб. журн. вычисл. математики / РАН. Сиб. отд-ние. — Новосибирск, 2017. — Т. 20, № 4. — С. 359–378.

В этой, первой из двух, части работы рассматривается задача поиска вектора рангов веб-страниц, также известная как задача поиска вектора PageRank и Google problem. Обсуждается связь этой задачи и эргодической теоремы, дается описание различных численных методов решения этой задачи и используемых в них теоретических конструкций, таких как Markov chain Monte Carlo, равновесие макросистемы.

DOI: 10.15372/SJNM20170402

Ключевые слова: *марковская цепь, эргодическая теорема, мультиномиальное распределение, концентрация меры, оценка максимального правдоподобия, Google problem, градиентный спуск, автоматическое дифференцирование, степенной закон распределения.*

Gasnikov A., Gasnikova E., Dvurechensky P., Mohammed A., Chernousova E. About the power law of the PageRank vector distribution. Part 1. Numerical methods for finding the PageRank vector // Siberian J. Num. Math. / Sib. Branch of Russ. Acad. of Sci. — Novosibirsk, 2017. — Vol. 20, № 4. — P. 359–378.

In Part 1 of this paper, we consider the web-pages ranking problem also known as the problem of finding the PageRank vector or Google problem. We discuss the connection of this problem with the ergodic theorem and describe different numerical methods to solve this problem together with their theoretical background, such as Markov Chain Monte Carlo and equilibrium in a macrosystem.

Keywords: *Markov chain, ergodic theorem, multinomial distribution, measure concentration, maximum likelihood estimate, Google problem, gradient descent, automatic differentiation, power law distribution.*

*Статья написана по материалам курсов, прочитанных А.В. Гасниковым в ЛШСМ 2012, 2013 гг. (Ратмино, Дубна) и в июльской смене 2016 г. лагеря Сириус (Сочи). Исследование П.Е. Двуреченского в пункте 4 поддержано грантом Президента РФ (МК-1806.2017.9). Исследование А.В. Гасникова и П.Е. Двуреченского в пункте 5 выполнено в ИППИ РАН за счет гранта Российского научного фонда (проект № 14-50-00150).

Посвящается памяти Виталия Дмитриевича Арнольда.

1. Введение

Данная работа, состоящая из двух частей, посвящена нескольким фундаментальным вопросам, связанным с эргодической теоремой для марковских процессов, с методами Монте-Карло, явлением концентрации меры, понятием равновесия макросистемы, основной теоремой математической статистики (теорема Фишера) о свойствах оценки максимального правдоподобия, ролью степенных законов распределения, машинным обучением, невыпуклой оптимизацией, автоматическим дифференцированием и рядом смежных вопросов. Все темы и вопросы обсуждаются на одном примере — Google problem, также известной как задача поиска вектора PageRank. Это позволяет увидеть, как математические результаты, входящие, по нашему мнению, в "золотой фонд" современной математики, соотносятся друг с другом. С одной стороны, постановка задачи поиска вектора PageRank мотивирована практикой. Умение эффективно и многократно находить вектор PageRank, например, позволяет поисковым системам ранжировать web-страницы в ответ на пользовательский запрос [33, 41, 43, 47, 52, 59]. Схожие постановки задач возникают в моделях поиска консенсуса [1]. Отметим также книгу [40], которая близка по содержанию к настоящей работе. С другой стороны, задача поиска вектора PageRank является непростой с точки зрения вычислительной математики. Задача сводится к поиску собственного вектора матрицы, т. е. по сути к решению системы линейных уравнений. Сложность, однако, заключается в том, что размерность этого вектора равна числу веб-страниц и может достигать миллиарда. Заметная часть изложенных в статье результатов была специально адаптирована для максимально широкой аудитории. В связи с этим важную роль играют ссылки на литературу, которая позволяет восстановить опущенные в статье детали.

В этой, первой из двух, частей предложена оригинальная интерпретация вектора PageRank, согласно которой осуществляется ранжирование web-страниц. Важно отметить, что эта интерпретация позволяет построить параллельный алгоритм для поиска вектора PageRank. В этой части также описаны другие численные методы поиска этого вектора. Во второй части проведена проверка степенного закона убывания компонент вектора PageRank для web-графа, построенного по наиболее популярной сейчас модели роста Интернета — модели Бакли–Остгуса [33]. Также обсуждаются вычислительные аспекты ранжирования web-страниц на практике. Очень ценным при этом оказался опыт совместной работы с сотрудниками компании Яндекс [41].

2. Google problem и эргодическая теорема

В 1998 г. Ларри Пейджем и Сергеем Брином был предложен специальный способ ранжирования web-страниц, который и лег в основу поисковой системы Google [52, 43]. Далее мы опишем этот способ, предварив описание необходимыми сведениями из теории дискретных марковских цепей [9, 24, 38, 44, 46, 50, 54–57, 64].

Рассмотрим ориентированный взвешенный граф (см. рисунок 1). Граф имеет n вершин. Каждой паре вершин соответствует некоторый вес $p_{ij} \geq 0$. Ребро, выходящее из вершины i в вершину j, имеет вес $p_{ij} > 0$. Если из вершины i в вершину j ребра нет, то полагаем $p_{ij} = 0$. Число p_{ij} интерпретируется как вероятность перейти из вершины i в вершину j. Набор чисел $\{p_{ij}\}_{i,j=1,1}^{n,n}$ удобно будет записать в виде матрицы $P = \|p_{ij}\|_{i,j=1,1}^{n,n}$

(квадратной таблицы, на пересечении i-й строки и j-го столбца которой стоит p_{ij}). Поскольку распределение вероятностей должно быть нормировано на единицу, то для любой вершины i имеет место равенство $\sum_{j=1}^{n} p_{ij} = 1$, и матрица P является стохастической по строкам.

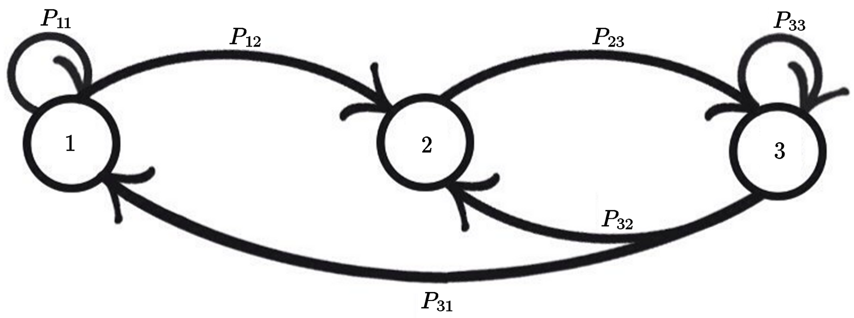

Рис. 1. Пример web-графа

На граф можно посмотреть как на город, вершинами которого являются различные районы города, а ребра — дороги (вообще говоря, с односторонним движением). Предположим, что в городе имеется "Красная площадь" — такой район, в который можно попасть по дорогам из любого другого. Оказывается, в этом предположении верен следующий результат (***эргодическая теорема для марковских процессов***): *если пустить блуждать человека по городу в течение длительного времени так, что человек будет случайно перемещаться из района в район согласно весам ребер графа, то доли* $\{\nu_k\}_{k=1}^{n}$ *времени, которые человек провел в разных районах, будут удовлетворять следующему уравнению*: $\sum_{i=1}^{n} \nu_i p_{ij} = \nu_j$ *или в векторном виде* $\nu^\top P = \nu^\top$, *имеющему единственное решение, удовлетворяющее* $\sum_{i=1}^{n} \nu_i = 1$. В связи с последним представлением говорят, что[1] $\nu = (\nu_1, \dots, \nu_n)^\top$ является левым собственным вектором матрицы P. Вектор ν также называют инвариантным или стационарным распределением вероятности в марковском процессе. В теории неотрицательных матриц [29], используемой в математической экономике, вектор $\nu^\top$ также называют *вектором Фробениуса–Перрона*. Предположение о "Красной площади" обеспечивает единственность ν. Если это предположение не верно, то вектор ν существует, но он не единственен и зависит от района, из которого стартовал человек. В общем случае город распадается на отдельные несвязанные между собой существенные кластеры[2] из районов, связанных между собой внутри кластера, и отдельного кластера несущественных районов. Последний определяется тем, что из любого его района можно попасть в один из существенных кластеров, но попасть обратно в несущественный кластер из существенных кластеров невозможно. Из такого несущественного кластера человек в конечном итоге попадет в один из существенных кластеров, откуда уже не выйдет. Рисунок 2 поясняет общий случай.

[1]Здесь и далее, если A — матрица из m строк и n столбцов с элементами A_{ij}, стоящими на пересечении i-й строки и j-го столбца, то по определению $A^\top$ — матрица из n строк и m столбцов с элементами A_{ji}, стоящими на пересечении i-й строки и j-го столбца. В частности, для рассматриваемого здесь случая $m = 1$.

[2]То есть из каждого такого кластера в любой другой кластер дорог нет.

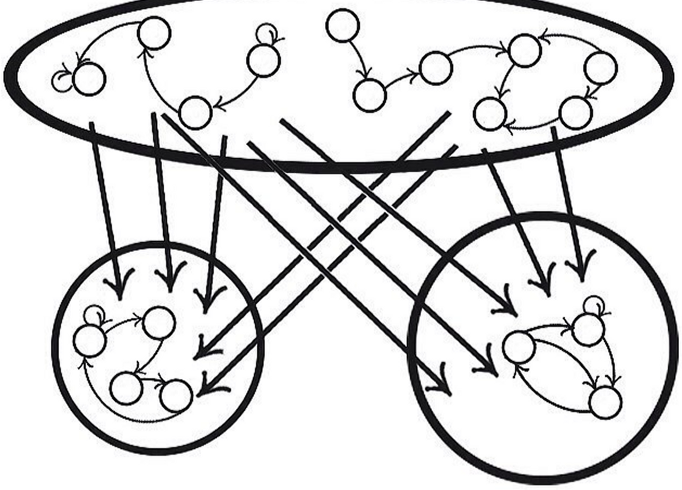

Рис. 2. Граф однородной дискретной марковской цепи в общем случае

Чтобы понять, откуда получается приведенная выше в вольной трактовке эргодическая теорема, получим с помощью формулы полной вероятности уравнение Колмогорова–Чэмпена. Это уравнение является основным при описании *дискретных однородных марковских цепей*. Обозначим через $p_i(t)$ — вероятность того, что человек находится в момент времени t в районе i. Тогда по формуле полной вероятности для любого $j = 1, \ldots, n$:

$$p_j(t+1) = \sum_{i=1}^{n} \underbrace{P\begin{pmatrix}\text{человек в момент времени } t\\ \text{находился в районе номер } i\end{pmatrix}}_{p_i(t)} \underbrace{P\begin{pmatrix}\text{человек перешел в район } j\\ \text{при условии, что был в районе } i\end{pmatrix}}_{p_{ij}}.$$

Или в матричном виде

$$p^\top(t+1) = p^\top(t)\, P. \tag{1}$$

Последнее равенство и называют ***уравнением Колмогорова–Чэмпена***.

Предположим теперь, что существует предел $\lim_{t\to\infty} p(t) = \nu$. Какому уравнению должен удовлетворять вектор ν? Переходя к пределу в обеих частях равенства (1) и учитывая, что $\lim_{t\to\infty}\left(p^\top(t)\, P\right) = \lim_{t\to\infty}\left(p^\top(t)\right) P$, получим уже известное нам соотношение

$$\nu^\top P = \nu^\top. \tag{2}$$

Именно из такого соотношения и было предложено искать вектор ранжирования web-страниц, называемый *вектором PageRank*, в модели Брина–Пейджа. Отличие этой модели от описанной выше только в интерпретации. Теперь вершины графа — это web-страницы, ребра графа — это гиперссылки. Под *Google problem* будем понимать задачу поиска вектора ν в этой модели.

Заметим, что в общем случае эргодическая теорема — это похожий по формулировке факт для общих динамических систем, в которых динамика, в отличие от (1), существенно нелинейная, равно как и получаемый при этом аналог соотношения (2). Грубо, эту теорему можно сформулировать так: *подобно* (2) *ищется инвариантная мера, т. е. такое распределение, которое не изменяется при действии оператора перехода* *динамической системы* (*в рассмотренном частном случае оператор задается матрицей* P *согласно* (1)), *если это распределение единственно*, *то доля времени*, *которую динамическая система проведет в том или ином состоянии, описывается этой инвариантной мерой*. С примерами применения этой теоремы любознательный читатель мог встречаться в теории чисел: теорема Вейля [11], теорема Гаусса–Кузьмина [3], а также в курсе термодинамики [23] (эргодическая гипотеза Лоренца). На самом деле, связь между эргодической теорией дискретных однородных цепей Маркова и эргодической теорией динамических систем намного глубже. В основе этой связи лежит *конструкция Улама* [6] (см. также [35]). К сожалению, все это существенно выходит за рамки даже институтских курсов, поэтому здесь мы лишь ограничимся ссылкой на анимацию, которая, на наш взгляд, хорошо поясняет написанное [65].

Отметим, что приведенные выше рассуждения справедливы в предположении существования предела $\lim_{t\to\infty} p(t) = \nu$. Казалось бы, что предположения о “Красной площади” будет достаточно и тут. Однако, как показывает простейший пример (рис. 3), в котором $n = 2$, $p_{11} = p_{22} = 0$, $p_{12} = p_{21} = 1$, хотя вектор $\nu = (1/2, 1/2)^{\top}$ существует и единственен, предел $\lim_{t\to\infty} p(t)$ не существует, поскольку с ростом t будет происходить периодическое чередование нулей и единиц в каждой компоненте вектора $p(t)$.

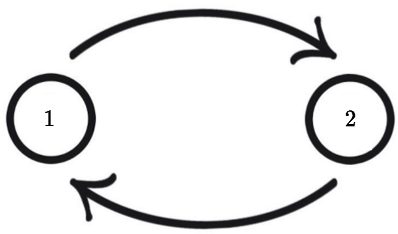

Рис. 3. Периодическая марковская цепь с периодом 2

Оказывается, что если выполняется условие “непериодичности”, то предел $\lim_{t\to\infty} p(t) = \nu$ действительно существует. Более того, эти два условия (существования “Красной площади” и “непериодичности”) являются не только достаточными, но и необходимыми для существования предела. Опишем, в чем заключается условие “непериодичности”. Из “Красной площади” выходит много различных маршрутов, которые в конце снова приводят на “Красную площадь”. Условие “непериодичности” означает, что наибольший общий делитель последовательности длин всевозможных маршрутов (начинающихся и заканчивающихся на “Красной площади”) равен 1. Уточним, что длина маршрута равна числу ребер, вошедших в маршрут. В типичных web-графах оба отмеченных условия выполняются, поэтому в дальнейшем мы считаем эти два условия выполненными.

3. Стандартные численные подходы к решению Google problem

В предыдущем пункте мы не только описали Google problem, но и неявно описали один из основных численных методов поиска вектора PageRank [52]. С одной стороны, этот вектор является решением уравнения (2). С другой стороны, он является пределом $\nu = \lim_{t\to\infty} p(t)$, где $p(t)$ рекуррентно рассчитывается согласно (1). Значит, (1) можно использовать для численного расчета $p(t)$ как приближения к вектору PageRank. Перейдем к анализу скорости сходимости такого метода.

Назовем *спектральной щелью* матрицы P наименьшее такое число $\alpha = 1 - |\beta| > 0$, где (вообще говоря, комплексное) число β удовлетворяет условию $|\beta| < 1$ и существует такой вектор η, что $\eta^\top P = \beta\eta^\top$. Хотя бы одно такое β существует. Другими словами, α — расстояние между максимальным собственным значением матрицы P (для стохастических матриц оно всегда равно 1) и следующим по величине модуля. Отсюда и название — спектральная щель (спектральный зазор — от английского spectral gap). Оказывается, что имеет место следующий результат:

$$\|p\,(t) - \nu\|_1 \overset{\text{def}}{=} \sum_{k=1}^{n} |p_k\,(t) - \nu_k| \le C \exp\big(-\alpha t/\tilde{C}\big). \tag{3}$$

Здесь и далее константы C, $\tilde{C}$ будут обозначать некоторые (каждый раз свои) универсальные константы, которые зависят от некоторых дополнительных деталей постановки и, как правило, ограничены числом 10.

Заметим, что часто факт сходимости процесса (1) к решению (2) также называют эргодической теоремой (см., например, [24]). В предыдущем пункте мы видели, что для этого есть основания. Однако, на наш взгляд, сходимость процесса (1) больше связана с принципом сжимающих отображений. Если под пространством понимать всевозможные лучи неотрицательного ортанта, а под метрикой на этом пространстве, в которой матрица P “сжимает”, понимать метрику Биркгофа–Гильберта [26], то можно показать, что сходимость процесса (1) к решению (2) следует из принципа сжимающих отображений. К сожалению, детали здесь не совсем элементарны, поэтому мы ограничимся только следующей аналогией. Формула (3) отражает геометрическую скорость сходимости, характерную для принципа сжимающих отображений, а коэффициент α как раз и характеризует степень сжатия, осуществляемого матрицей P. Геометрически себе это можно представлять (правда, не очень строго) как сжатие с коэффициентом, не меньшим $1 - \alpha$, к инвариантному направлению, задаваемому вектором ν.

С одной стороны, оценка (3) всего лишь сводит задачу оценки скорости сходимости метода (1) к задаче оценки спектральной щели α. С другой стороны, последняя задача в ряде случаев может быть эффективно решена. В частности, к эффективным инструментам оценки α относятся изопериметрическое неравенство Чигера и неравенство Пуанкаре [24, 38, 46, 54, 64], которое также можно понимать как неравенство концентрации меры [64] (см. далее). Имеются и другие способы оценки α (см., например, [44, 55–57, 60]), однако все эти способы далеко выходят за рамки даже институтских курсов. Поэтому здесь мы ограничимся простыми, но важными в контексте рассматриваемых приложений, случаями.

Предположим сначала, что для рассматриваемого web-графа существует такая web-страница, на которую есть ссылка из любой web-страницы в том числе из самой себя (усиленный аналог предположения о наличии “Красной площади”, поскольку теперь на “Красную площадь” из любого района есть прямые дороги), более того, предположим, что на каждой такой ссылке стоит вероятность, не меньшая чем γ. Для такого web-графа имеет место неравенство $\alpha \ge \gamma$ [58].

Предположим далее, что в модели блуждания по web-графу имеется “телепортация”: с вероятностью $1 - \delta$ человек действует как в исходной модели, а с вероятностью δ “забывает про все правила” и случайно равновероятно выбирает среди n вершин одну, в которую и переходит. Тогда, если ввести квадратную матрицу E размера n на n, состоящую из одинаковых элементов $1/n$, уравнение (2) примет вид

$$p^\top\,(t+1) = p^\top\,(t)\,\big((1-\delta)\,P + \delta E\big). \tag{4}$$

В таком случае вектор PageRank необходимо будет искать из уравнения

$$\nu^\top = \nu^\top \big((1-\delta) P + \delta E \big). \tag{5}$$

При $0 < \delta < 1$ уравнение (5) гарантированно имеет единственное в классе распределений вероятностей решение. Более того, для спектральной щели матрицы $(1-\delta)P + \delta E$ имеет место оценка $\alpha \geq \delta$ [52]. На практике для вычисления вектора PageRank обычно используют уравнение (5) с $\delta = 0.15$ [33, 52].

Поскольку матрица P для реальных web-графов обычно сильно разреженная (большая часть элементов равна нулю), то использовать формулу (4) в таком виде не рационально: одна итерация будет стоить $3n^2$ арифметических операций типа сложения и умножения чисел. Однако формулу (4) можно переписать следующим образом:

$$p^\top(t+1) = (1-\delta)\, p^\top(t)\, P + \delta\, (1/n, \ldots, 1/n). \tag{6}$$

Расчет по этой формуле требует по порядку лишь $2sn$ арифметических операций, где s — среднее число ненулевых элементов в строке матрицы P. Для Интернета $n \approx 10^{10}$, а $s \ll 10^4$.

Выше был описан исторически самый первый алгоритм, использовавшийся для расчета вектора PageRank. Он получил название метода простой итерации (МПИ). Как было отмечено, МПИ эффективен, если α не очень близко к нулю. В частности, как в модели с телепортацией. Действительно, современный ноутбук в состоянии выполнять до 10^{10} арифметических операций в секунду. С учетом программной реализации и ряда других ограничений этот порядок на практике обычно уменьшается до 10^8. Из оценки (3) видно, что сходимость, например при $\alpha \geq 0.15$, очень быстрая. Хорошая точность получается уже после нескольких десятков итераций. С учетом того, что одна итерация требует по порядку $2sn \leq 10^{13}$ арифметических операций, описанный метод позволяет за день найти вектор PageRank для всего Интернета. К сожалению, на самом деле все не так просто. Проблема в памяти, необходимой для матрицы P. Разумеется, P необходимо загружать не как матрицу из n^2 элементов, а в виде так называемых списков смежностей по строкам. Однако это все равно не решает проблемы. Быстрая память компьютера — кэш-память разных уровней. Кэш-память процессора совсем маленькая, ее совершенно недостаточно. Более медленная — оперативная память. Тем не менее, если удалось бы выгрузить матрицу P в такую память, то производительность программы соответствовала бы оценке, приведенной выше. Обычно оперативной памяти в современном персональном компьютере не более нескольких десятков Гбайт ($\sim 10^{10}$ байт), чего, очевидно, недостаточно. Следующая память — жесткий диск. Обращение программы к этой памяти, по сути, останавливает нормальную работу программы. Как только кончаются ресурсы оперативной памяти, мы видим, что программа либо не работает совсем, либо начинает работать очень медленно. Отмеченную проблему можно решать, увеличивая оперативную память либо используя распределенную память. Мы не будем здесь на этом останавливаться. Укажем лишь, что во избежание указанных проблем с памятью все численные эксперименты в статье проводятся для размеров графов $n \leq 10^5$.

Но что делать, если α оказалось достаточно малым или мы не можем должным образом оценить снизу α, чтобы гарантировать быструю сходимость МПИ? В таком случае полезным оказывается следующий результат [30] (Поляк–Тремба, 2012), не предполагающий, кстати, выполнения условия “непериодичности”:

$$\left\| P^\top \bar{p}_T - \bar{p}_T \right\|_1 \leq \frac{C}{T}, \quad \bar{p}_T \stackrel{\text{def}}{=} \frac{1}{T} \sum_{t=1}^{T} p(t). \tag{7}$$

Эта оценка уже никак не зависит от α. Параллельное процедуре (1) суммирование получающихся векторов позволяет вычислять вектор $\bar{p}_T$ за время (по порядку величины), равное времени расчета $p\,(T)$. К сожалению, для наших целей, т. е. исследования степенного закона убывания компонент вектора PageRank, метод Поляка–Трембы не подходит. Причина связана с видом оценки (7). Имеется принципиальная разница с оценкой (3) в том, что в оценке (3) мы можем гарантировать близость найденного вектора $p\,(t)$ к вектору PageRank ν, обеспечив малость $\|p\,(t)-\nu\|_1$. Что касается соотношения (7), то в типичных случаях из $\left\|P^\top \bar{p}_T - \bar{p}_T\right\| \approx \varepsilon$ можно при больших n лишь получить, что $\|\bar{p}_T - \nu\| \approx C\varepsilon/\alpha$, т. е. опять возникает "нехорошая" зависимость в оценке от α. Для симметричной матрицы P и 2-нормы этот результат был получен Красносельским и Крейном в 1952 г. [25].

К сожалению, и другие способы поиска вектора PageRank, которые на первый взгляд не используют в своих оценках α, на деле оказываются методами, выдающими такой вектор $\tilde{p}_T$, что $\left\|P^\top \tilde{p}_T - \tilde{p}_T\right\| \approx \varepsilon$ в некоторой норме (обычно это 1-норма, 2-норма и бесконечная норма; в следующем пункте мы поясним, что имеется в виду под этими нормами), что приводит к той же сложности, что и в методе Поляка–Трембы. Тем не менее, хочется отметить большой прогресс, достигнутый за последнее время, во многом благодаря разработкам Б.Т. Поляка, А.С. Немировского, Ю.Е. Нестерова (см., например, [14] и цитированную там литературу), в создании эффективных численных методов решения задач выпуклой оптимизации вида

$$\left\|P^\top p - p\right\|_l^b \to \min_{p\geq 0:\ \sum\limits_{k=1}^{n} p_k = 1}, \quad b = 1, 2;\ \ l = 1, 2, \infty.$$

Эти наработки оказываются полезными, поскольку вектор PageRank можно также понимать как решение такой задачи. Действительно, всегда имеет место неравенство $\left\|P^\top p - p\right\| \geq 0$ и только на ν имеет место равенство $\left\|\, P^\top \nu - \nu\right\| = 0$.

Отметим, в частности, метод условного градиента из работы [2], который позволяет получить вектор $\tilde{p}_T$, удовлетворяющий $\left\|P^\top \tilde{p}_T - \tilde{p}_T\right\|_2 \leq \varepsilon$ за $C\left(n + s^2\varepsilon^{-2}\ln n\right)$ арифметических операций. Удивительно, что в эту оценку не входит sn — число ненулевых элементов матрицы P. В частности, при $s \approx \sqrt{n}$ получается сложность, пропорциональная n, а не $n^{3/2}$, как можно было ожидать (к сожалению, за это есть и плата в виде множителя ε^{-2}). Тем не менее, еще раз повторим, что несмотря на все возможные ускорения вычислений, использование таких методов в наших целях, к сожалению, не представляется возможным.

Отмеченные в этом пункте методы (МПИ и Поляка–Трембы) не исчерпывают множество методов, гарантирующих малость $\|\tilde{p}_T - \nu\|$, в оценку скорости сходимости которых входит α. К таким методам можно отнести, например метод Д. Спилмана [62], являющийся вариацией описанного выше метода Поляка–Трембы.

Однако мы сосредоточимся далее на так называемых методах Монте-Карло [18]. Наряду с тем, что эти методы являются численными методами поиска вектора PageRank, они также позволяют по-новому проинтерпретировать вектор PageRank.

4. Markov Chain Monte Carlo и Google problem

В предыдущем пункте для вычисления вектора PageRank был описан МПИ, в основе которого лежит равенство (1). С одной стороны, в предположении достаточной величины спектральной щели этот метод обеспечивает очень быструю сходимость по норме приближенного решения к вектору PageRank. С другой стороны, одна итерация этого метода требует порядка sn арифметических операций, что может привести к невозможности его применения, если не выполнено неравенство $s \ll n$. К тому же, каждому поисковому запросу обычно соответствует свой web-граф релевантных страниц. Поэтому, как правило, для каждого запроса нужно вычислять свой вектор PageRank. Следовательно, для эффективной работы реальной поисковой системы скорость вычисления вектора PageRank оказывается важнее точности его вычисления. МПИ не позволяет в полной мере пожертвовать точностью в угоду скорости вычисления. В самом деле, согласно оценке (3), переход, скажем, от точности $\varepsilon \approx 10^{-6}$ к точности $\varepsilon \approx 10^{-3}$ уменьшит время счета всего лишь вдвое. Предложим другой алгоритм, который будет более чувствителен к точности по сравнению с МПИ, но при этом будет иметь меньшую стоимость итерации.

В основе этого алгоритма лежит метод *Markov Chain Monte Carlo* (MCMC) [44]. Основная идея заключается в практическом использовании эргодической теоремы из пункта 2. То есть, грубо говоря, нужно запустить человека и достаточно долго подождать, собирая статистику посещенных им вершин графа (т. е. web-страниц). При оценке эффективности работы такого метода важно ответить на два вопроса: насколько эффективным можно сделать шаг и сколько шагов надо сделать человеку.

Наиболее вычислительно затратная операция на каждом шаге метода — это случайный выбор ребра для перехода из текущей вершины. Случай, в котором из каждой вершины (web-страницы) выходит не более $s \ll n$ ребер, является простым. На каждом шаге за порядка s арифметических операций разыгрывается случайная величина в соответствии с вероятностями переходов. При этом память после каждой операции освобождается, т. е. не требуется хранить в памяти ничего, кроме матрицы P и текущего вектора частот. Именно в такой ситуации ($s \ll n$) мы и будем находиться, изучая во второй части этой работы конкретные web-графы, порожденные моделью Бакли–Остгуса. Отметим, тем не менее, как можно организовать вычисления в сложном случае, когда s является достаточно большим. Можно до старта итераций алгоритма подготовить специальным образом память за порядка sn арифметических операций, например, следующим образом. Каждой вершине ставится в соответствие свое разбиение отрезка $[0, 1]$ так, чтобы число подотрезков равнялось числу выходящих ребер, а длины подотрезков были пропорциональны вероятностям перехода по этим ребрам. Тогда на каждом шаге достаточно один раз генерировать равномерно распределенную на отрезке $[0, 1]$ случайную величину (стандартные генераторы, как правило, именно это, в первую очередь, и предлагают) и нанести ее на соответствующий текущей вершине подотрезок. В зависимости от того, в какой из подотрезков она попала, выбрать исходящее ребро и сдвинуть человека по нему. Недостаток этого подхода — использование дополнительно памяти для хранения порядка sn чисел и высокая стоимость подготовки этой памяти. К счастью, такая подготовка, требующая порядка sn арифметических операций, нужна только один раз, в отличие от МПИ, в котором таких затрат требует каждая итерация. При дополнительных предположениях о компактном описании формул расчета $\{p_{ij}\}_{j=1}^{n}$ существуют быстрые способы случайного выбора ребра для перехода из текущей вершины, которые приводят к тому, что шаг выполняется за $\sim \log_2 s \le \log_2 n$ арифметических операций. В частности, если имеет место равенство между собой отличных от нуля внедиагональных элементов матрицы P в каждой строке, то сложность шага $\sim \log_2 s$ достигается, например, методом Кнута–Яо [18].

Перейдем теперь к обсуждению оценки числа шагов, которые нужно сделать человеку для достижения заданной точности приближения вектора PageRank. Постараемся грубо сформулировать основной результат [17].

Пусть $\nu(T)$ — вектор частот (случайный вектор) пребывания в различных вершинах блуждающего человечка после T шагов и $T \gg T_0 \overset{\mathrm{def}}{=} C\alpha^{-1}\ln(n/\varepsilon)$. Тогда с вероятностью не меньше $1-\sigma$ (здесь и везде в дальнейшем $\sigma \in (0,1)$ — произвольный доверительный уровень) имеет место неравенство

$$\|\nu(T)-\nu\|_2 \le \mathrm{C}\sqrt{\frac{\ln n + \ln(\sigma^{-1})}{\alpha T}}, \tag{8}$$

где $\|x\|_2 = \sqrt{\sum_{k=1}^n x_k^2}$.

Приведем сравнительные характеристики упомянутых к настоящему моменту методов. Под сложностью понимается количество арифметических операций типа умножения двух чисел, которые достаточно осуществить, чтобы (в случае MCMC с вероятностью не меньше $1-\sigma$) достичь точности ε по целевой функции[3], которая в таблице, приведенной ниже, обозначается целью. Напомним, что вектор $\tilde{p}_T$ — это результат работы метода, а ν — вектор PageRank (2). Стоит также пояснить обозначение $\|x\|_\infty = \max_{k=1,\ldots,n}|x_k|$.

Таблица. Сравнение методов решения задачи поиска вектора PageRank

Метод	Сложность	Цель
МПИ [52]	$\frac{sn}{\alpha}\ln\left(\frac{2}{\varepsilon}\right)$	$\left\|\tilde{p}_T-\nu\right\|_1$
Поляка–Трембы [30]	$\frac{2sn}{\varepsilon}$	$\left\|P^\top\tilde{p}_T-\tilde{p}_T\right\|_1$
Д. Спилмана [62]	$C\left(n+\frac{s^2}{\alpha\varepsilon}\ln\left(\frac{1}{\varepsilon}\right)\right)$	$\left\|\tilde{p}_T-\nu\right\|_\infty$
MCMC [17]	$C\left(n+\frac{\log_2(n)\ln(n/\sigma)}{\alpha\varepsilon^2}\right)$	$\left\|\tilde{p}_T-\nu\right\|_2$
Вариация метода условного градиента [2]	$C\left(n+\frac{s^2\ln n}{\varepsilon^2}\right)$	$\left\|P^\top\tilde{p}_T-\tilde{p}_T\right\|_2$

Сравнение затрудняется тем, что у всех методов разные целевые функции, а значит, и разные критерии точности. Кроме того, разные методы могут по-разному быть ускорены с использованием параллельных вычислений. Для задач огромных размеров, к которым, безусловно, относится и Google problem, этот вопрос выходит на передний план. Если умножение матрицы на вектор в МПИ хорошо параллелится, то например, возможность применения параллельных вычислений к MCMC в описанном здесь варианте не ясна.

5. Параллельный метод MCMC для Google problem.

Идею распараллеливания, собственно также как и идеи остальных описанных методов (МПИ, MCMC), подсказывает "жизнь". А именно, в реальном Интернете "блуждает" не один человек (пользователь), а много пользователей. Обозначим число независимо блуждающих пользователей через N. Рассмотрим сначала для простоты ситуацию, когда в графе всего две вершины, т. е. $n=2$. Поскольку пользователи блуждают независимо, то для каждого из них можно ожидать, что на шаге $T_0 = C\alpha^{-1}\ln(2/\varepsilon)$ вероятность

[3] Заметим, что для метода MCMC имеет смысл рассматривать только $\varepsilon \ll n^{-1/2}$, поскольку $\left\|(n^{-1},\ldots,n^{-1})\right\|_2 = n^{-1/2}$.

обнаружить пользователя в вершине 1 с хорошей точностью ($\sim \varepsilon$) равна ν_1. Соответственно, вероятность обнаружить его в вершине 2 равна $\nu_2 = 1 - \nu_1$. Таким образом, посмотрев на то, в какой вершине находился каждый человек на шаге T_0, мы с хорошим приближением получим так называемую *простую выборку* X (независимые одинаково распределенные случайные величины) из *распределения Бернулли* с вероятностью успеха (выпадения "орла"), равной ν_1. Последнее означает, что каждый человек (подобно монетке) на шаге T_0 независимо от всех остальных с вероятностью ν_1 будет обнаружен в состоянии 1 (выпал "орлом"), а с вероятностью $1 - \nu_1$ будет обнаружен в состоянии 2 (выпал "решкой"). Чтобы оценить вектор PageRank в данном простом случае достаточно оценить по этой выборке ν_1 — частоту выпадения "орла". Интуиция подсказывает, что в качестве оценки неизвестного параметра ν_1 следует использовать r/N — долю пользователей, которые оказались в состоянии 1. Интуиция подсказывает правильно! Далее будет показано, что такой способ, действительно, оптимальный.

Вероятность, что число людей r, которые наблюдались в состоянии 1, равно k, может быть посчитана по следующей формуле (*биномиальное распределение*):

$$P(r = k) = C_N^k \nu_1^k (1 - \nu_1)^{N-k}.$$

Используя грубый вариант *формулы Стирлинга* $k! \simeq (k/e)^k$, получаем *оценку типа Санова* [34]:

$$C_N^k \nu_1^k (1 - \nu_1)^{N-k} \simeq \exp(-N \cdot KL(k/N, \nu_1)),$$

где

$$KL(p, q) = -p \ln\left(\frac{p}{q}\right) - (1 - p) \ln\left(\frac{1 - p}{1 - q}\right).$$

Отсюда по *неравенству Пинскера* [42] $KL(p, q) \geq 2(p - q)^2$ следует, что *с вероятностью не меньше* $1 - \sigma$ *имеет место неравенство*

$$\left|\frac{r}{N} - \nu_1\right| \leq C\sqrt{\frac{\ln(\sigma^{-1})}{N}}, \tag{9}$$

которое иллюстрирует рис. 4.

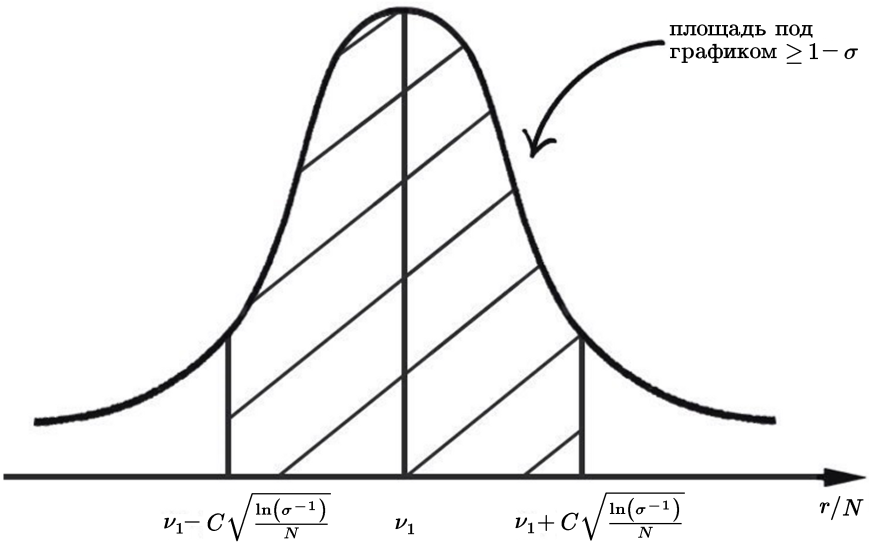

Рис. 4. График зависимости $P(r) = C_N^r \nu_1^r (1 - \nu_1)^{N-r}$ при большом значении N

По сути, этот рисунок отражает тот факт, что биномиальное распределение (биномиальная мера) с ростом числа исходов (людей) N *концентрируется* вокруг ν_1.

Вектор $\nu = (\nu_1, \nu_2)^\top$, в малой окрестности которого с большой вероятностью на больших временах находится реальный вектор распределения людей по web-страницам (в данном случае двум), естественно называть *равновесием макросистемы*, описываемой блуждающими по web-графу людьми. *Таким образом, мы пришли к еще одному пониманию вектора PageRank ν как равновесию описанной выше макросистемы.*

Достаточно большое количество примеров макросистем с анализом их равновесий собрано в [9, гл. 6] (см. также [4, 7, 8, 10, 12, 13, 15, 16, 19, 23, 24, 49, 61]). Одним из таких примеров является “модель Эренфестов” [23, 24], поясняющая различные парадоксы в термодинамике, другим примером макросистемы, которая может быть известна читателям, является “Кинетика социального неравенства” [7].

На соотношение (9) можно посмотреть и с другой стороны, т. е. с точки зрения математической статистики [27]. А именно, вспомним, что ν_1 нам не известно. В то время как реализацию случайной величины (статистики) r/N мы можем измерить. Соотношение (9) от такого взгляда на себя не перестанет быть верным.

В общем случае для n вершин можно провести аналогичные рассуждения с заменой биномиального распределения *мультиномиальным* [4, 9, 13, 17]. Чтобы получить аналог оценки (9), для *концентрации мультиномиальной меры*, можно использовать векторное неравенство Хефдинга [4, 17, 42, 64]. Оказывается, *с вероятностью не меньше* $1 - \sigma$ *имеет место следующее неравенство* [17]:

$$\left\| \frac{r}{N} - \nu \right\|_2 \le C \sqrt{\frac{\ln\left(\sigma^{-1}\right)}{N}}, \tag{10}$$

где вектор $r = (r_1, \ldots, r_n)^\top$ *описывает распределение людей по web-страницам в момент наблюдения* T_0. Заметим, что константу C здесь можно оценить сверху числом 4. Неравенства (9), (10) являются представителями класса *неравенств концентрации меры*, играющего важную роль в современной теории вероятностей и ее приложениях [20, 39, 42, 53, 64].

Если запустить $N \sim \varepsilon^{-2} \ln\left(\sigma^{-1}\right)$ человек, дав каждому сделать $T_0 \sim \alpha^{-1} \ln\left(n/\varepsilon\right)$ шагов (стоимость шага $\sim \log_2 n$), то (вообще говоря, случайный) вектор r/N, характеризующий распределение людей по web-страницам на шаге T_0, с вероятностью не меньше $1 - \sigma$ обладает таким свойством: $\left\| r/N - \nu \right\|_2 \le \varepsilon$. Это следует из оценки (10). Таким образом, мы приходим к оценке сложности (параллельного варианта) метода MCMC:

$$C\left(n + \frac{\log_2(n) \ln\left(n/\varepsilon\right) \ln\left(\sigma^{-1}\right)}{\alpha \varepsilon^2}\right),$$

которая с точностью до логарифмического множителя совпадает с оценкой, приведенной в пункте 4 (см. таблицу). Однако отличие данного подхода в том, что можно пустить блуждать людей параллельно. То есть можно распараллелить работу программы на N процессорах. Разумеется, можно распараллелить и на любом меньшем числе процессоров, давая процессору следить сразу за несколькими людьми. Можно еще немного “выиграть”, если сначала запустить меньшее число людей, но допускать, что со временем люди случайно производят клонов, которые с момента рождения начинают жить независимой жизнью, так чтобы к моменту T_0 наблюдалось, как и раньше, N человек.

Пока мы не пояснили, в каком смысле и почему выбранный способ рассуждения оптимален. Для большей наглядности снова вернемся к случаю $n = 2$, и заметим, что оценка r/N неизвестного параметра ν_1 может быть получена следующим образом:

$$\frac{r}{N} = \arg\max_{\nu\in[0,1]} \nu^r \ (1-\nu)^{N-r}. \tag{11}$$

Действительно, максимум у функций

$$f_1(\nu) = \nu^r \ (1-\nu)^{N-r_1}$$

и

$$f_2(\nu) = \ln\left(\nu^r \ (1-\nu)^{N-r}\right) = r\ln\nu + (N-r)\ln(1-\nu)$$

достигается в одной и той же точке, поэтому рассмотрим задачу $f_2(\nu) \to \max_{\nu\in[0,1]}$. Из ***принципа Ферма*** [36] (максимум лежит либо на границе, либо среди точек, в которых производная обращается в ноль) получаем условие $df_2(\nu)/d\nu = 0$, которое в данном случае примет вид

$$\frac{r}{\nu} - \frac{N-r}{1-\nu} = 0, \qquad \text{т. е.} \qquad \frac{r/N}{\nu} = \frac{1-r/N}{1-\nu}.$$

Отсюда и следует формула (11).

Функция правдоподобия, стоящая в правой части (11), отражает вероятность конкретного[4] распределения людей по вершинам (web-страницам), для которого число людей в первой вершине равно ν. То есть оценка r/N может быть проинтерпретирована как *оценка максимального правдоподобия*. Указанный выше способ построения оценок является общим способом получения хороших оценок неизвестных параметров. А именно, пусть есть схема эксперимента (параметрическая модель), согласно которой можно посчитать вероятность того, что мы будем наблюдать то, что наблюдаем. Эта вероятность (плотность вероятности) зависит от неизвестного вектора параметров θ. Будем максимизировать эту вероятность по параметрам этого вектора при заданных значениях наблюдений, т. е. при заданной выборке. Тогда получим зависимость неизвестных параметров от выборки X. Именно эту зависимость $\hat{\theta}^N(X)$ в общем случае и называют *оценкой максимального правдоподобия* вектора неизвестных параметров. Оказывается (***теорема Фишера***), *что в случае выполнения довольно общих условий регулярности используемой параметрической модели такая оценка является асимптотически оптимальной* (***теория ле Кама*** [21]). Последнее можно понимать так, что для каждой отдельной компоненты k вектора θ можно написать неравенство, аналогичное (9), справедливое с вероятностью не меньшей $1-\sigma$:

$$\left|\hat{\theta}_k^N(X) - \theta_k\right| \le C_{k,N}(\theta)\sqrt{\frac{\ln(\sigma^{-1})}{N}}$$

с $C_{k,N}(\theta) \to C_k(\theta)$ при $N\to\infty$, где N — объем выборки, т. е. число наблюдений. При этом $C_k(\theta) > 0$ являются равномерно по θ и k наименее возможными. Имеется в виду, что если как-то по-другому оценивать θ (обозначим другой способ оценивания $\tilde{\theta}^N(X)$), то для любого θ и k с вероятностью не меньшей $1-\sigma$:

$$\left|\tilde{\theta}_k^N(X) - \theta_k\right| > \tilde{C}_{k,N}(\theta)\sqrt{\frac{\ln(\sigma^{-1})}{N}},$$

где $\underline{\lim}_{N\to\infty} \tilde{C}_{k,N}(\theta) \ge C_k(\theta)$.

Строго говоря, именно такая зависимость правой части от σ имеет место не всегда. В общем случае при зафиксированном N с уменьшением σ правая часть может вести себя хуже, однако при не очень маленьких значениях[5] σ написанное верно всегда.

[4] Поскольку стоит "конкретного", то не нужно писать биномиальный коэффициент.

[5] Пороговое значение σ_0 удовлетворяет следующей оценке: $\ln N \ll \ln\sigma_0^{-1} \ll N$.

К сожалению, приведенная выше весьма вольная формулировка не отражает в полной мере всю мощь теоремы Фишера. Ведь в таком виде открытым остается вопрос об оценках совместного распределения отклонений сразу нескольких компонент оценки максимального правдоподобия от истинных значений. На самом деле можно сформулировать результат (делается это чуть более громоздко) об асимптотической оптимальности оценки максимального правдоподобия и в таких общих категориях. Мы не будем здесь этого делать. Детали — в классической монографии [21].

Основной недостаток теоремы Фишера в приведенном нами варианте ле Кама заключается в том, что оценка оптимальна только в пределе $N \to \infty$. В реальности же нам дана только конечная выборка, пусть и большого объема. Сформулированный выше результат ничего не говорит о том, насколько хорошей будет такая оценка при конечном N. Также теорема не говорит, как получать точные оценки на $C_{k,N}(\theta)$. Теорема лишь предлагает эффективный способ расчета $C_k(\theta)$[6]. Скажем, приведенная нами ранее оценка (9) хотя по форме и выглядит так, как нужно, но все же она далеко не оптимальна. В частности, в оптимальный вариант оценки (9) нужно прописывать под корнем в числителе еще множитель $\nu_1\,(1-\nu_1)$, оцененный нами сверху в (9) константой $1/4$. Мы сделали такую замену в (9), чтобы в правую часть неравенства не входил неизвестный параметр ν_1. При значениях ν_1, близких к нулю или единице, такая "замена" оказывается слишком грубой. Мы привели простой пример, в котором смогли проконтролировать грубость своих рассуждений. В общем случае, к сожалению, это не так просто сделать (если говорить о потерях в мультипликативных константах типа C). Поэтому, если мы хотим использовать оценки типа (9), (10), то борьба за "оптимальные константы" сводится не только к выбору оптимального способа оценивания неизвестных параметров, но и к способу оценивания концентрации вероятностного распределения выбранной оценки вокруг истинного значения. За последние несколько лет теория ле Кама была существенно пересмотрена как раз в контексте отмеченных выше проблем. Современный (неасимптотический) вариант параметрической математической статистики, по-прежнему базирующийся на теореме Фишера, недавно был построен В.Г. Спокойным [63]. Многие отмеченные проблемы удалось в большой степени решить.

Другая проблема — это зависимость C_k от неизвестного параметра θ. Грубо, но практически эффективно, проблема решается подстановкой $C_k\big(\hat{\theta}^N(X)\big)$. Более точно надо писать неравенство концентрации для $C_k\big(\hat{\theta}^N(X)\big)$, исходя из неравенства на $\hat{\theta}^N(X)$. Казалось бы, что возникает "порочный круг", и это приводит к иерархии "зацепляющихся неравенств". Однако, если выполнить описанное выше огрубление[7] для того, чтобы обрезать эту цепочку, не сразу, а на каком-то далеком вложенном неравенстве, то чем оно дальше, тем к меньшей грубости это в итоге приведет. Детали мы также вынуждены здесь опустить.

В заключение, для закрепления материала, мы предлагаем читателям оценить, сколько надо опросить человек на выходе с избирательных участков большого города, чтобы с точностью 5 % (0.05) и с вероятностью 0.99 оценить победителя во втором туре выборов (два кандидата, графы против всех нет). Для решения этой задачи рекомендуется воспользоваться неравенством (9) с явно выписанными константами

$$\left|\frac{r}{N}-\nu_1\right| \le \frac{1}{2}\sqrt{\frac{\ln(2/\sigma)}{N}}.$$

[6] $C_k(\theta)$ не универсальны и зависят от использующейся параметрической модели.

[7] Речь идет об оценке $\nu_1\,(1-\nu_1) \le 1/4$.

Благодарности. Авторы выражают благодарность рецензенту за предложения по улучшению текста статьи.

Литература

1. **Агаев Р.П., Чеботарев П.Ю.** Сходимость и устойчивость в задачах согласования характеристик (обзор базовых результатов) // Управление большими системами. — 2010. — Т. 30, № 1. — С. 470–505.
2. **Аникин А.С., Гасников А.В., Горнов А.Ю., Камзолов Д.И., Максимов Ю.В., Нестеров Ю.Е.** Эффективные численные методы решения задачи PageRank для дважды разреженных матриц // Труды МФТИ. — 2015. — Т. 7, № 4. — С. 74–94.
3. **Арнольд В.И.** Цепные дроби. — М.: Изд-во МЦНМО, 2001. — (Библиотека “Математическое просвещение”; вып. 14).
4. **Баймурзина Д.Р., Гасников А.В., Гасникова Е.В.** Теория макросистем с точки зрения стохастической химической кинетики // Труды МФТИ. — 2015. — Т. 7, № 4. — С. 95–103.
5. **Батищева Я.Г., Веденяпин В.В.** II-й закон термодинамики для химической кинетики // Мат. моделирование. — 2005. — Т. 17, № 8. — С. 106–110.
6. **Бланк М.Л.** Устойчивость и локализация в хаотической динамике. — М.: Изд-во МЦНМО, 2001.
7. **Богданов К.Ю.** Кинетика социального неравенства // Квант. — 2004. — № 5. — С. 7–12.
8. **Богданов К.Ю.** Хищник и жертва: уравнение сосуществования // Квант. — 2014. — № 4-5. — С. 13–17.
9. **Бузун Н.О., Гасников А.В., Гончаров Ф.О., Горбачев О.Г., Гуз С.А., Крымова Е.А., Натан А.А., Черноусова Е.О.** Стохастический анализ в задачах. Часть 1 / А.В. Гасников. — М.: Изд-во МФТИ, 2016.
10. **Вайдлих В.** Социодинамика: системный подход к математическому моделированию в социальных науках. Пер. с англ. 3-е изд. — М.: УРСС, 2010.
11. **Гальперин Г.А., Земляков А.Н.** Математические бильярды. — М.: Наука, 1990. — (Серия “Библиотечка Квант”; вып. 77).
12. **Гардинер К.В.** Стохастические методы в естественных науках. — М.: Мир, 1986.
13. Введение в математическое моделирование транспортных потоков. 2-е изд. / А.В. Гасников. — М.: Изд-во МЦНМО, 2013.
14. **Гасников А.В.** Эффективные численные методы поиска равновесий в больших транспортных сетях: Дис. ... докт. физ.-мат. наук: 05.13.18. — М.: Изд-во МФТИ, 2016.
15. **Гасников А.В., Гасникова Е.В.** Об энтропийно-подобных функционалах, возникающих в стохастической химической кинетике при концентрации инвариантной меры и в качестве функций Ляпунова динамики квазисредних // Мат. заметки. — 2013. — Т. 94, № 6. — С. 816–824.
16. **Гасников А.В., Гасникова Е.В., Мендель М.А., Чепурченко К.В.** Эволюционные выводы энтропийной модели расчета матрицы корреспонденций // Мат. моделирование. — 2016. — Т. 28, № 4. — С. 111–124.
17. **Гасников А.В., Дмитриев Д.Ю.** Об эффективных рандомизированных алгоритмах поиска вектора PageRank // Журн. вычисл. матем. и мат. физики. — 2015. — Т. 55, № 3. — С. 355–371.
18. **Ермаков С.М.** Метод Монте-Карло в вычислительной математике. Вводный курс. — М.: Бином. Лаборатория знаний, 2009.

19. **Занг В.-Б.** Синергетическая экономика: время и перемены в нелинейной экономической теории. — М.: Мир, 1999.
20. **Зорич В.А.** Математический анализ задач естествознания. — М.: Изд-во МЦНМО, 2008.
21. **Ибрагимов И.А., Хасьминский Р.З.** Асимптотическая теория оценивания. — М.: Наука, 1979.
22. **Калинкин А.В.** Марковские ветвящиеся процессы с взаимодействием // Успехи мат. наук. — 2002. — Т. 57, № 2(344). — С. 23–84.
23. **Кац М.** Вероятность и смежные вопросы в физике. — М.: Мир, 1965.
24. **Кельберт М.Я., Сухов Ю.М.** Марковские цепи как отправная точка теории случайных процессов и их приложений. Вероятность и статистика в примерах и задачах. Т. 2. — М.: Изд-во МЦНМО, 2010.
25. **Красносельский М.А., Крейн С.Г.** Замечание о распределении ошибок при решении системы линейных уравнений при помощи итерационного процесса // Успехи мат. наук. — 1952. — Т. 7, № 4(50). — С. 157–161.
26. **Красносельский М.А., Лифшиц Е.А., Соболев А.В.** Позитивные линейные системы. Метод положительных операторов. — М.: Наука, 1985.
27. **Лагутин М.Б.** Наглядная математическая статистика: учебное пособие. 3-е изд. — М.: Бином. Лаборатория знаний, 2013.
28. **Малышев В.А., Пирогов С.А.** Обратимость и необратимость в стохастической химической кинетике // Успехи мат. наук. — 2008. — Т. 63, № 1(379). — С. 4–36.
29. **Никайдо Х.** Выпуклые структуры и математическая экономика. — М.: Мир, 1972.
30. **Поляк Б.Т., Тремба А.А.** Решение задачи PageRank для больших матриц с помощью регуляризации // Автоматика и телемеханика. — 2012. — Вып. 11. — С. 144–166.
31. **Пурмаль А.П., Слободецкая Е.М., Травин С.О.** Как превращаются вещества. — М.: Наука, 1984. — (Серия “Библиотечка Квант”; вып. 36).
32. **Разжевайкин В.Н.** Анализ моделей динамики популяций. Учебное пособие. — М.: Изд-во МФТИ, 2010.
33. **Райгородский А.М.** Модели интернета. — Долгопрудный: Изд. дом “Интеллект”, 2013.
34. **Санов И.Н.** О вероятности больших отклонений случайных величин // Мат. сборник. — 1957. — Т. 42(84), № 1. — С. 11–44.
35. **Синай Я.Г.** Как математики изучают хаос // Мат. просвещение. — 2001. — Т. 5. — С. 32–46.
36. **Тихомиров В.М.** Рассказы о максимумах и минимумах. — М.: Наука, 1986. — (Серия “Библиотечка Квант”; вып. 56).
37. **Avrachenkov K., Lebedev D.** PageRank of scale-free growing networks // Internet Math. — 2006. — Vol. 3, № 2. — P. 207–232.
38. **Aldous D., Fill J.** Reversible Markov Chains and Random Walks on Graphs. — Berkeley: University of California, 2002.
39. **Ball K.** An elementary introduction to modern convex geometry // Flavors of Geometry. MSRI Publ. — 1997. — Vol. 31. — P. 1–58.
40. **Blum A., Hopcroft J., and Kannan R.** Foundations of Data Science. — http://www.cs.cornell.edu/jeh/book.pdf.
41. **Bogolubsky L., Dvurechensky P., Gasnikov A., Gusev G., Nesterov Yu., Raigorodskii A., Tikhonov A., and Zhukovskii M.** Learning Supervised PageRank with Gradient-Based and Gradient-Free Optimization Methods // The Thirtieth Annual Conference on Neural Information Processing Systems (NIPS 2016). — Barcelona, 2016.

42. **Boucheron S., Lugosi G., and Massart P.** Concentration Inequalities: A Nonasymptotic Theory of Independence. — Oxford University Press, 2013.
43. **Brin S., Page L.** The anatomy of a large-scale hypertextual web search engine // Comput. Network ISDN Syst. — 1998. — Vol. 30, № 1-7. — P. 107–117.
44. **Diaconis P.** The Markov chain Monte Carlo revolution // Bulletin (New Series) of the AMS. — 2009. — Vol. 49, № 2. — P. 179–205.
45. **Ethier N.S., Kurtz T.G.** Markov Processes. — New York: John Wiley & Sons Inc., 1986.
46. **Chung F.** Laplacians and the Cheeger inequality for directed graphs // Annals of Combinatorics. — 2005. — Vol. 9, № 1. — P. 1–19.
47. **Franceschet M.** PageRank: Standing on the shoulders of giant // Communication of ACM. — 2011. — Vol. 54, № 6. — P. 92–101.
48. **Grechnikov E.A.** The degree distribution and the number of edges between nodes of given degrees in the Buckley–Osthus model of a random web graph // Internet Math. — 2012. — Vol. 8, № 3. — P. 257–287.
49. **Jaynes E.T.** Probability Theory. The Logic of Science. — Cambridge University Press, 2003.
50. **Joulin A., Ollivier Y.** Curvature, concentration and error estimates for Markov chain Monte Carlo // The Annals of Probability. — 2010. — Vol. 38, № 6. — P. 2418–2442.
51. **Kapur J.N.** Maximum — Entropy Models in Science and Engineering. — New York: John Wiley & Sons Inc., 1989.
52. **Langville A.N., Meyer C.D.** Google's PageRank and Beyond: The Science of Search Engine Rankings. — Princeton University Press, 2006.
53. **Ledoux M.** The Concentration of Measure Phenomenon // Math. Surveys Monogr. — Providence, RI: AMS, 2001. — Vol. 89.
54. **Levin D.A., Peres Y., and Wilmer E.L.** Markov Chain and Mixing Times. — AMS, 2009.
55. **Lezaud P.** Chernoff-type bound for finite Markov chains // The Annals of Applied Probability. — 1998. — Vol. 8, № 3. — P. 849–867.
56. **Meyn S.P., Tweedie R.L.** Markov Chains and Stochastic Stability. — London: Springer-Verlag, 2005.
57. **Montenegro R., Tetali P.** Mathematical aspects of mixing times in Markov chains // Foundations and Trends in Theoretical Computer Science. — 2016. — Vol. 1, № 3. — P. 237–354.
58. **Nesterov Yu., Nemirovski A.** Finding the stationary states of Markov chains by iterative methods // Appl. Math. and Comput. — 2015. — Vol. 255. — P. 58–65.
59. **Pandurangan G., Raghavan P., and Upfal E.** Using PageRank to characterize web structure // Internet Math. — 2006. — Vol. 3, № 1. — P. 1–20.
60. **Paulin D.** Concentration inequalities for Markov chains by Marton couplings // Electron J. Probab. — 2015. — Vol. 20, № 79. — P. 1–32.
61. **Sandholm W.** Population Games and Evolutionary Dynamics. Economic Learning and Social Evolution. — Cambridge: MIT Press, 2010.
62. **Spielman D.** Lecture № 7. — 2009. — http://www.cse.cuhk.edu.hk/ chi/csc5160/notes/L07.pdf.
63. **Spokoiny V.** Parametric estimation. Finite sample theory // The Annals of Statistics. — 2012. — Vol. 10, № 6. — P. 2877–2909.
64. **Van Handel R.** Probability in High Dimension. — Princeton University, 2014. — (Lecture Notes; ORF 570).
65. Chaos IX: Chaotic or not? Chaos research today. — http://www.chaos-math.org/fr/chaos-ix-chaotique-ou-pas.

Поступила в редакцию 7 марта 2017 г.,
в окончательном варианте 15 мая 2017 г.

Литература в транслитерации

1. **Agaev R.P., Chebotarev P.Yu.** Skhodimost' i ustoychivost' v zadachah soglasovaniya harakteristik (obzor bazovyh rezul'tatov) // Upravlenie bol'shimi sistemami. — 2010. — T. 30, № 1. — S. 470–505.
2. **Anikin A.S., Gasnikov A.V., Gornov A.Yu., Kamzolov D.I., Maksimov Yu.V., Nesterov Yu.E.** Effektivnye chislennye metody resheniya zadachi PageRank dlya dvazhdy razrezhennyh matrits // Trudy MFTI. — 2015. — T. 7, № 4. — S. 74–94.
3. **Arnol'd V.I.** Tsepnye drobi. — M.: Izd-vo MTSNMO, 2001. — (Biblioteka "Matematicheskoe prosveshchenie"; vyp. 14).
4. **Baymurzina D.R., Gasnikov A.V., Gasnikova E.V.** Teoriya makrosistem s tochki zreniya stohasticheskoy himicheskoy kinetiki // Trudy MFTI. — 2015. — T. 7, № 4. — S. 95–103.
5. **Batishcheva Ya.G., Vedenyapin V.V.** II-y zakon termodinamiki dlya himicheskoy kinetiki // Mat. modelirovanie. — 2005. — T. 17, № 8. — S. 106–110.
6. **Blank M.L.** Ustoychivost' i lokalizatsiya v haoticheskoy dinamike. — M.: Izd-vo MTSNMO, 2001.
7. **Bogdanov K.Yu.** Kinetika sotsial'nogo neravenstva // Kvant. — 2004. — № 5. — S. 7–12.
8. **Bogdanov K.Yu.** Hishchnik i zhertva: uravnenie sosushchestvovaniya // Kvant. — 2014. — № 4-5. — S. 13–17.
9. **Buzun N.O., Gasnikov A.V., Goncharov F.O., Gorbachev O.G., Guz S.A., Krymova E.A., Natan A.A., Chernousova E.O.** Stohasticheskiy analiz v zadachah. Chast' 1 / A.V. Gasnikov. — M.: Izd-vo MFTI, 2016.
10. **Vaydlih V.** Sotsiodinamika: sistemnyy podhod k matematicheskomu modelirovaniyu v sotsial'nyh naukah. Per. s angl. 3-e izd. — M.: URSS, 2010.
11. **Gal'perin G.A., Zemlyakov A.N.** Matematicheskie bil'yardy. — M.: Nauka, 1990. — (Seriya "Bibliotechka Kvant"; vyp. 77).
12. **Gardiner K.V.** Stohasticheskie metody v estestvennyh naukah. — M.: Mir, 1986.
13. Vvedenie v matematicheskoe modelirovanie transportnyh potokov. 2-e izd. / A.V. Gasnikov. — M.: Izd-vo MTSNMO, 2013.
14. **Gasnikov A.V.** Effektivnye chislennye metody poiska ravnovesiy v bol'shih transportnyh setyah: Dis. ... dokt. fiz.-mat. nauk: 05.13.18. — M.: Izd-vo MFTI, 2016.
15. **Gasnikov A.V., Gasnikova E.V.** Ob entropiyno-podobnyh funktsionalah, voznikayushchih v stohasticheskoy himicheskoy kinetike pri kontsentratsii invariantnoy mery i v kachestve funktsiy Lyapunova dinamiki kvazisrednih // Mat. zametki. — 2013. — T. 94, № 6. — S. 816–824.
16. **Gasnikov A.V., Gasnikova E.V., Mendel' M.A., Chepurchenko K.V.** Evolyutsionnye vyvody entropiynoy modeli rascheta matritsy korrespondentsiy // Mat. modelirovanie. — 2016. — T. 28, № 4. — S. 111–124.
17. **Gasnikov A.V., Dmitriev D.Yu.** Ob effektivnyh randomizirovannyh algoritmah poiska vektora PageRank // ZHurn. vychisl. matem. i mat. fiziki. — 2015. — T. 55, № 3. — S. 355–371.
18. **Ermakov S.M.** Metod Monte-Karlo v vychislitel'noy matematike. Vvodnyy kurs. — M.: Binom. Laboratoriya znaniy, 2009.
19. **Zang V.-B.** Sinergeticheskaya ekonomika: vremya i peremeny v nelineynoy ekonomicheskoy teorii. — M.: Mir, 1999.
20. **Zorich V.A.** Matematicheskiy analiz zadach estestvoznaniya. — M.: Izd-vo MTSNMO, 2008.
21. **Ibragimov I.A., Has'minskiy R.Z.** Asimptoticheskaya teoriya otsenivaniya. — M.: Nauka, 1979.
22. **Kalinkin A.V.** Markovskie vetvyashchiesya protsessy s vzaimodeystviem // Uspekhi mat. nauk. — 2002. — T. 57, № 2(344). — S. 23–84.

23. **Kats M.** Veroyatnost' i smezhnye voprosy v fizike. — M.: Mir, 1965.

24. **Kel'bert M.Ya., Suhov Yu.M.** Markovskie tsepi kak otpravnaya tochka teorii sluchaynyh protsessov i ih prilozheniy. Veroyatnost' i statistika v primerah i zadachah. T. 2. — M.: Izd-vo MTSNMO, 2010.

25. **Krasnosel'skiy M.A., Kreyn S.G.** Zamechanie o raspredelenii oshibok pri reshenii sistemy lineynyh uravneniy pri pomoshchi iteratsionnogo protsessa // Uspekhi mat. nauk. — 1952. — T. 7, № 4(50). — S. 157–161.

26. **Krasnosel'skiy M.A., Lifshits E.A., Sobolev A.V.** Pozitivnye lineynye sistemy. Metod polozhitel'nyh operatorov. — M.: Nauka, 1985.

27. **Lagutin M.B.** Naglyadnaya matematicheskaya statistika: uchebnoe posobie. 3-e izd. — M.: Binom. Laboratoriya znaniy, 2013.

28. **Malyshev V.A., Pirogov S.A.** Obratimost' i neobratimost' v stohasticheskoy himicheskoy kinetike // Uspekhi mat. nauk. — 2008. — T. 63, № 1(379). — S. 4–36.

29. **Nikaydo H.** Vypuklye struktury i matematicheskaya ekonomika. — M.: Mir, 1972.

30. **Polyak B.T., Tremba A.A.** Reshenie zadachi PageRank dlya bol'shih matrits s pomoshch'yu regulyarizatsii // Avtomatika i telemekhanika. — 2012. — Vyp. 11. — S. 144–166.

31. **Purmal' A.P., Slobodetskaya E.M., Travin S.O.** Kak prevrashchayutsya veshchestva. — M.: Nauka, 1984. — (Seriya “Bibliotechka Kvant”; vyp. 36).

32. **Razzhevaykin V.N.** Analiz modeley dinamiki populyatsiy. Uchebnoe posobie. — M.: Izd-vo MFTI, 2010.

33. **Raygorodskiy A.M.** Modeli interneta. — Dolgoprudnyy: Izd. dom “Intellekt”, 2013.

34. **Sanov I.N.** O veroyatnosti bol'shih otkloneniy sluchaynyh velichin // Mat. sbornik. — 1957. — T. 42(84), № 1. — S. 11–44.

35. **Sinay Ya.G.** Kak matematiki izuchayut haos // Mat. prosveshchenie. — 2001. — T. 5. — S. 32–46.

36. **Tihomirov V.M.** Rasskazy o maksimumah i minimumah. — M.: Nauka, 1986. — (Seriya “Bibliotechka Kvant”; vyp. 56).

37. **Avrachenkov K., Lebedev D.** PageRank of scale-free growing networks // Internet Math. — 2006. — Vol. 3, № 2. — P. 207–232.

38. **Aldous D., Fill J.** Reversible Markov Chains and Random Walks on Graphs. — Berkeley: University of California, 2002.

39. **Ball K.** An elementary introduction to modern convex geometry // Flavors of Geometry. MSRI Publ. — 1997. — Vol. 31. — P. 1–58.

40. **Blum A., Hopcroft J., and Kannan R.** Foundations of Data Science. — http://www.cs.cornell.edu/jeh/book.pdf.

41. **Bogolubsky L., Dvurechensky P., Gasnikov A., Gusev G., Nesterov Yu., Raigorodskii A., Tikhonov A., and Zhukovskii M.** Learning Supervised PageRank with Gradient-Based and Gradient-Free Optimization Methods // The Thirtieth Annual Conference on Neural Information Processing Systems (NIPS 2016). — Barcelona, 2016.

42. **Boucheron S., Lugosi G., and Massart P.** Concentration Inequalities: A Nonasymptotic Theory of Independence. — Oxford University Press, 2013.

43. **Brin S., Page L.** The anatomy of a large-scale hypertextual web search engine // Comput. Network ISDN Syst. — 1998. — Vol. 30, № 1-7. — P. 107–117.

44. **Diaconis P.** The Markov chain Monte Carlo revolution // Bulletin (New Series) of the AMS. — 2009. — Vol. 49, № 2. — P. 179–205.

45. **Ethier N.S., Kurtz T.G.** Markov Processes. — New York: John Wiley & Sons Inc., 1986.

46. **Chung F.** Laplacians and the Cheeger inequality for directed graphs // Annals of Combinatorics. — 2005. — Vol. 9, № 1. — P. 1–19.

47. **Franceschet M.** PageRank: Standing on the shoulders of giant // Communication of ACM. — 2011. — Vol. 54, № 6. — P. 92–101.

48. **Grechnikov E.A.** The degree distribution and the number of edges between nodes of given degrees in the Buckley–Osthus model of a random web graph // Internet Math. — 2012. — Vol. 8, № 3. — P. 257–287.

49. **Jaynes E.T.** Probability Theory. The Logic of Science. — Cambridge University Press, 2003.

50. **Joulin A., Ollivier Y.** Curvature, concentration and error estimates for Markov chain Monte Carlo // The Annals of Probability. — 2010. — Vol. 38, № 6. — P. 2418–2442.

51. **Kapur J.N.** Maximum — Entropy Models in Science and Engineering. — New York: John Wiley & Sons Inc., 1989.

52. **Langville A.N., Meyer C.D.** Google's PageRank and Beyond: The Science of Search Engine Rankings. — Princeton University Press, 2006.

53. **Ledoux M.** The Concentration of Measure Phenomenon // Math. Surveys Monogr. — Providence, RI: AMS, 2001. — Vol. 89.

54. **Levin D.A., Peres Y., and Wilmer E.L.** Markov Chain and Mixing Times. — AMS, 2009.

55. **Lezaud P.** Chernoff-type bound for finite Markov chains // The Annals of Applied Probability. — 1998. — Vol. 8, № 3. — P. 849–867.

56. **Meyn S.P., Tweedie R.L.** Markov Chains and Stochastic Stability. — London: Springer-Verlag, 2005.

57. **Montenegro R., Tetali P.** Mathematical aspects of mixing times in Markov chains // Foundations and Trends in Theoretical Computer Science. — 2016. — Vol. 1, № 3. — P. 237–354.

58. **Nesterov Yu., Nemirovski A.** Finding the stationary states of Markov chains by iterative methods // Appl. Math. and Comput. — 2015. — Vol. 255. — P. 58–65.

59. **Pandurangan G., Raghavan P., and Upfal E.** Using PageRank to characterize web structure // Internet Math. — 2006. — Vol. 3, № 1. — P. 1–20.

60. **Paulin D.** Concentration inequalities for Markov chains by Marton couplings // Electron J. Probab. — 2015. — Vol. 20, № 79. — P. 1–32.

61. **Sandholm W.** Population Games and Evolutionary Dynamics. Economic Learning and Social Evolution. — Cambridge: MIT Press, 2010.

62. **Spielman D.** Lecture № 7. — 2009. — http://www.cse.cuhk.edu.hk/ chi/csc5160/notes/L07.pdf.

63. **Spokoiny V.** Parametric estimation. Finite sample theory // The Annals of Statistics. — 2012. — Vol. 10, № 6. — P. 2877–2909.

64. **Van Handel R.** Probability in High Dimension. — Princeton University, 2014. — (Lecture Notes; ORF 570).

65. Chaos IX: Chaotic or not? Chaos research today. — http://www.chaos-math.org/fr/chaos-ix-chaotique-ou-pas.