

Московский физико-технический институт
(государственный университет)

Институт теоретической физики имени Л. Д. Ландау РАН

А. Б. Теслюк

**О СЛУЧАЙНОМ БЛУЖДЕНИИ
ПО ГРАФУ ВЕБ-ДОКУМЕНТОВ**

Дипломная работа по кафедре
проблем теоретической физики.

Научный руководитель:
доктор физ.-мат. наук, проф.
Щур Л. Н.

Черноголовка - 2004

Содержание

Введение	3
Определения и понятия	4
Модели случайных графов	7
Модель случайного графа Эрдоса	7
Модель Ваттса-Строгатса	8
Модель Барабаши-Альберта	8
Вычисления и результаты	9
Выводы	11
Список литературы	20

Введение

В современных информационных системах, таких как Internet, чрезвычайно важную роль играют механизмы поиска нужной информации. Ключевая задача поисковой системы состоит в том, чтобы выделить из огромного множества информации именно те документы, которые наиболее точно подходят к запросу пользователя, и расположить их так, чтобы пользователю не пришлось их искать среди множества ненужной информации. Например, в ответ на достаточно специфический запрос "random walks" поисковая система Google находит около 540 тысяч различных документов, содержащих эти слова. Однако Internet был бы совершенно бесполезным инструментом, если для поиска информации по случайным блужданиям пользователю каждый раз приходилось бы просматривать сотни тысяч документов. Алгоритмы ранжирования информации большинства поисковых систем¹ основаны на моделях случайного блуждания по графам документов. В основе этих алгоритмов лежат модели, предложенные более полувека назад в работах Каца [12], Хаббелла [11]. Схожие модели применяются и в библиометрике [7, 20]. Хороший обзор исследований в этой области представлен в работе Кляйнберга [14].

Однако несмотря на столь обширные области применения, математический аппарат случайных блужданий по графам развит недостаточно хорошо. Для задач ранжирования информации необходимо уметь находить стационарные решения случайного блуждания по графу, а также их свойства, основываясь на свойствах заданного графа. Но к сожалению, стационарные решения случайных блужданий по графу удается получить в аналитическом виде, пригодном для практического применения, только в редких случаях, когда граф неориентированный и связный. Метода нахождения стационарного решения случайных блужданий по ориентированным графам в общем случае нам не известно.

Теория конечных марковских цепей дает лишь критерии существования стационарных решений, а также связь стационарных решений случайного блуждания с собственными векторами стохастической матрицы марковской цепи². Однако на практике вычисление собственных векторов матриц большого размера зачастую невозможно. Например для того, чтобы построить марковский процесс случайного блуждания по графу веб-документов ($\sim 10^{10}$ документов) потребуется сохранить в памяти матрицу размером

¹например, системы Google [9]

²определение стохастической матрицы марковской цепи и критерий существования стационарного решения см. в следующем разделе

$\sim 10^{20}$ байт, что во много раз превосходит возможности современных вычислительных систем.

В настоящей работе мы предлагаем способ нахождения функции распределения элементов вектора стационарного решения случайного блуждания по графу. Мы исследовали случайные блуждания на нескольких известных моделях случайных графов, а именно на классической модели случайного графа Эрдоша [6], модели Барабаши-Альберта [2] и для модели Ваттса-Строгатса [23]. Нами установлено, что функция распределения элементов вектора стационарного решения случайного блуждания по графу (для этих моделей) полностью определяется функцией распределения координационных чисел вершин графа. Кроме того, нами получен вид функции распределения элементов вектора стационарного решения для случайного блуждания с переходом на произвольную вершину. Именно такое случайное блуждание используется в алгоритмах поисковой системы Google.

Статья организована следующим образом. В разделе "Определения и понятия" мы приводим основные определения и понятия, используемые в работе. В разделе "Модели" Мы даем описания моделей случайных графов, которые используются в работе. В разделе "Вычисления и результаты" приводятся детали вычислений и основные результаты. В заключительной части статьи мы обсуждаем взаимосвязь наших результатов со степенными законами распределения популярности веб-документов и веб-сайтов [22].

Определения и понятия

Рассмотрим граф (G, E) , где G - множество вершин, а E - множество ребер. *Матрицей связности* графа называется матрица, у которой элемент $a_{ij} = 1$, если существует ребро исходящее из i -й вершины и входящее в j -ю. $a_{ij} = 0$, если такого ребра нет. В случае, когда граф неориентированный, матрица связности симметрична, т.е $a_{ij} = a_{ji}$. Количество исходящих ребер называется координационным числом вершины, и обозначается d_i .

Зададим случайное блуждание по графу следующим образом: выбрав одну вершину в качестве начальной точки, мы произвольным образом перемещаемся в одну из ее соседних вершин. Затем повторяем случайное перемещение из вновь выбранной вершины и т.д. Получившаяся последовательность вершин называется *реализацией случайного блуждания по графу (G, E)* . Нетрудно видеть, что вероятность перехода из i -й вершины в j -ю равна

$$p_{ij} = \frac{a_{ij}}{d_i} = \mathbf{D} \cdot \mathbf{A}, \quad (0.1)$$

где \mathbf{D} - диагональная матрица $(\mathbf{D})_{ij} = \delta_{ij}/d_i$.

В терминах теории конечных марковских цепей случайное блуждание по графу есть конечная марковская цепь. Множество вершин графа \mathbf{G} называется множеством состояний. Матрица вероятностей перехода между состояниями:

$$\mathbf{P} = (p_{ij}) = \mathbf{D} \cdot \mathbf{A} \quad (0.2)$$

называется стохастической матрицей или матрицей переходных вероятностей.

Найдем матрицу вероятностей перехода между состояниями за 2 шага. Элемент $p_{ij}^{(2)}$ есть вероятность за 2 шага попасть из i -го состояния в j -ое. Нетрудно видеть, что

$$\mathbf{P}^{(2)} = (p_{ij}^{(2)}) = \left(\sum_k p_{ik} p_{kj} \right) = \mathbf{P}^2 \quad (0.3)$$

По индукции легко доказать, что матрица вероятностей перехода за n -шагов будет равна

$$\mathbf{P}^{(n)} = \mathbf{P}^n. \quad (0.4)$$

Цепь Маркова называется неприводимой, если любое ее состояние достижимо из любого другого состояния. Состояние G_j называется периодическим, если существует такое t , что для любого n , не кратного t , $p_{jj}^{(n)} = 0$. Непериодическое состояние называется эргодическим, если среднее время возвращения (среднее число шагов) в это состояние конечно. Неприводимая цепь Маркова называется эргодической, если все ее состояния эргодические.

Один из самых замечательных фактов, известных в теории конечных марковских цепей состоит в том, что в эргодической марковской цепи существуют не зависящие от начального состояния пределы:

$$\pi_k = \lim_{n \rightarrow \infty} p_{ik}^{(n)}, \quad \pi_k > 0, \quad (0.5)$$

$$\sum_k \pi_k = 1,$$

$$\pi_k = \sum_i \pi_i p_{ik} = \mathbf{P}^T \cdot \boldsymbol{\pi}. \quad (0.6)$$

Пределы π_i называются инвариантными, или стационарными распределениями для данной цепи Маркова. Часто вектор π называют стационарным решением случайного блуждания. Вектор π есть собственный вектор транспонированной стохастической матрицы \mathbf{P}^T , соответствующий собственному значению 1.

В терминах случайного блуждания по графу элемент вектора π_i имеет смысл вероятности того, что в произвольный момент времени траектория случайного блуждания пройдет через i -ю вершину.

Из формулы (0.6) видно, что задача нахождения стационарного распределения марковского процесса сводится к задаче о нахождении собственного вектора стохастической матрицы. В случаях, когда марковский процесс имеет огромное количество состояний, решение этой задачи может быть затруднительно даже численно. Однако, к счастью, в некоторых случаях можно найти стационарное распределение в простом аналитическом виде. Нетрудно проверить, что в случае, когда граф ненаправленный и связный, то стационарное распределение случайного блуждания существует и выражается формулой [18]:

$$\pi_k = \frac{d_k}{2m}, \quad (0.7)$$

где $m = \frac{1}{2} \sum_k d_k$ - количество ребер в графе. Действительно, в ненаправленном графе матрица связности симметрична $a_{ij} = a_{ji}$. Тогда:

$$\begin{aligned} \sum_j \pi_j p_{ji} &= \sum_j \frac{d_j}{2m} \frac{a_{ji}}{d_j} = \\ &= \frac{1}{2m} \sum_j a_{ji} = \frac{1}{2m} \sum_j a_{ij} = \frac{d_i}{2m} = \pi_i. \end{aligned} \quad (0.8)$$

Обратим внимание на такой простой факт: если граф (G, E) есть случайный связные неориентированный граф, и известна функция распределения координационных чисел вершин $p_d(x)$, то функция распределения элементов вектора $p_\pi(x)$ будет равна

$$p_\pi(x) = p_d\left(\frac{x}{2m}\right). \quad (0.9)$$

Соотношение (0.9) является простым следствием формулы (0.7), однако оно будет играть ключевую роль в дальнейшем изложении. Во избежании путаницы в обозначениях договоримся, что буква p с двумя индексами обозначает элемент стохастической матрицы, а функция $p(x)$ - функцию распределения случайной величины x .

Помимо случайного блуждания по графу, которое мы определили выше, в литературе встречается его модификация, которую назовем *случайным блужданием по графу с переходом на произвольную вершину* [9]. Допустим у нас имеется граф (G, E) . На каждом шаге случайного блуждания мы с некоторой вероятностью γ переходим на произвольную вершину из всего множества G , или с вероятностью $1 - \gamma$ выбираем в качестве следующей вершины одну из соседей. Такая модель, применительно к случайному блужданию по графу веб-документов, имеет наглядную интерпретацию. Пользователь, передвигаясь в веб-пространстве либо открывает новую страницу (или как вариант попадает на нее из поисковой системы), либо переходит по ссылке со страницы, на который он находится. Эти два типа перемещений отражены в модели случайного блуждания с переходом на произвольную страницу [9].

Стохастическая матрица марковской цепи, задаваемой таким случайным блужданием, будет определяться формулой:

$$\mathbf{P} = \frac{\gamma}{N}\mathbf{I} + (1 - \gamma)\mathbf{D} \cdot \mathbf{A}, \quad (0.10)$$

где N - количество состояний в марковской цепи (количество вершин в графе), \mathbf{I} - матрица, все элементы которой равны 1, \mathbf{D} - диагональная матрица $(\mathbf{D})_{ij} = \delta_{ij}d_i^{-1}$, \mathbf{A} - матрица связности графа.

Нетрудно доказать, что построенная таким образом марковская цепь будет неприводимой и эргодической. Действительно, на каждом шаге любое состояние достижимо с вероятностью заведомо не меньшей, чем $\frac{\gamma}{N}$. Следовательно все состояния достижимы за конечное время $t \sim \frac{N}{\gamma}$. Следовательно, независимо от свойств графа, будут существовать стационарные пределы (0.5). К сожалению, простого аналитического способа, нахождения стационарного распределения, аналогичного (0.7), нам неизвестно.

Модели случайных графов

В этом разделе представлены краткие сведения о моделях случайных графов, которые мы анализировали в нашей работе.

Модель случайного графа Эрдоша.

Самая первая модель случайного графа была предложена в 1959 году Полом Эрдошем [6]. Случайный граф строится следующим образом: имеется N вершин. Последовательно перебираются все пары вершин и, с некоторой вероятностью p , каждая пара соединяется ребром. Изначально модель предлагалась для неориентированных

графов, однако несложно естественным образом обобщить ее на случай ориентированных графов. Для этого, при построении графа, необходимо перебирать все упорядоченные пары вершин и соединять их с некоторой вероятностью направленным ребром. Среднее число ребер в таком графе равно Np . Нетрудно видеть, что функция распределения координационного числа вершины подчиняется биномиальному закону $p(d) = C_{N-1}^d p^d (1-p)^{N-d-1}$. При больших N удобно ввести параметр $\lambda = Np$, а биномиальное распределение заменить распределением Пуассона с параметром λ , $p(d) = e^{-\lambda} \frac{\lambda^d}{d!}$.

Модель Ваттса-Строгатса.

В 1998 году Дунканом Ваттсом и Стивенем Строгатсом была предложена модель случайного графа, которая активно применяется при описании так называемых *small-world* сетей [23]. Кстати этими же авторами было предложено само понятие “small-world” сети, сети с небольшой длиной пути между двумя произвольными вершинами. Рассмотрим регулярный граф, состоящий из N вершин. Каждая вершина связана с $2k$ соседями. Мы последовательно перебираем все ребра графа и с некоторой вероятностью p отсоединяем ребро от одной вершины и подсоединяем его к произвольно выбранной, но так, что между двумя вершинами может быть не более одного ребра. Полученный в результате граф назовем случайным графом Ваттса-Строгатса. Варьируя параметр p можно получить как регулярный граф (при $p = 0$), так и случайный граф Эрдоша, когда $p = 1$. Граф Ваттса-Строгатса лежит где-то в промежутке между упорядоченным регулярным графом и совершенно случайным графом Эрдоша. Тем не менее, как мы увидим ниже, с точки зрения случайных блужданий, модель Ваттса-Строгатса мало чем отличается от модели Эрдоша. Все дело в том, что функции распределения координационного числа вершины (ниже будет показана ее связь со стационарным распределением случайного блуждания) в этих двух моделях очень похожи. Действительно, координационное число вершины в модели Ваттса-Строгатса есть сумма $2k$ и разности двух случайных величин - количества подсоединенных ребер (обозначим η) и количества отсоединенных ребер (обозначим ξ). Величины ξ и η распределены по биномиальным законам $p(\xi) = C_k^\xi p^\xi (1-p)^{k-\xi}$, а $p(\eta) = C_{(N-1-2k)}^\eta z^\eta (1-z)^{N-1-2k-\eta}$, где $z = \frac{p}{N-1}$. При $N \gg k \gg 1$ функция распределения координационных чисел описывается нормальным распределением $N(k, pk)$.

Модель Барабаши-Альберта.

Две модели, описанные выше, обладают одной общей особенностью: координационные числа вершин в обеих моделях распределены по нормальному закону (при $N \rightarrow \infty$).

Во многих реальных системах, таких как генные сети или WWW наблюдается степенное распределение координационных чисел вершин. В 1999 году Барабаши и Альберт предложили модель случайного графа [2], со степенным распределением координационных чисел. Граф строится следующим образом: на начальном этапе имеется m_0 вершин. Затем на каждом шаге в граф добавляется новая вершина и соединяется $m < m_0$ ребрами с другими вершинами, при этом вероятность подсоединения ребра к вершине пропорциональна ее координационному числу: $P(k_i) = k_i / \sum_i k_i$. На t -ом шаге граф состоит из $m_0 + t$ вершин и mt ребер. Авторы модели показали, что в таком растущем графе, функция распределения координационных чисел вершин не зависит от времени (и соответственно от размера графа) и описывается степенным законом с показателем $\gamma = 2.9 \pm 0.1$.

Вычисления и Результаты.

В этом разделе мы приводим результаты численного моделирования случайных блужданий на трех моделях случайных графов, представленных в предыдущем разделе. Мы сначала приводим детали вычислений стационарных распределений. Затем мы демонстрируем связь стационарных распределений и функции распределения координационных чисел вершин.

Как уже было упомянуто в разделе (), задача о нахождении стационарных распределений сводится к задаче о нахождении собственного вектора транспонированной матрицы связности графа. Однако прямое вычисление собственного вектора для графов с большим количеством вершин неэффективно, потому что требует огромного количества операций с матрицами большого размера. Более оптимальный метод - метод итераций, предложенный авторами в [17]. Обозначим стохастическую матрицу марковской цепи буквой \mathbf{P} и рассмотрим следующую последовательность итераций:

$$\begin{aligned} u_i^{(0)} &= \frac{1}{N} \\ u_i^{(n+1)} &= \mathbf{P}^T u_i^{(n)} \end{aligned} \tag{0.11}$$

Из уравнений уравнения (0.5) и (0.4), нетрудно увидеть, что при $n \rightarrow \infty$, $u_i^{(n)}$ стремится к стационарному распределению марковской цепи, задаваемой стохастической матрицей \mathbf{P} . Причем такой метод итераций экспоненциально быстро сходится, а показатель экспоненты определяется вторым собственным значением матрицы связности

графа. Подробную информацию об оптимизации итеративного метода можно найти в работе [10].

В наших вычислениях мы брали случайные графы, состоящие из $N = 10240$ вершин. Для всех моделей было достаточно 50 итераций для достижения точности порядка 1%. Для более эффективного использования памяти, мы хранили структуру графа не в форме матрицы, а в форме списка связей. Т.к. матрица связности графа содержит немало нулей, это позволило достичь значительной экономии памяти и заметно увеличить производительность вычислений.

Перейдем теперь к результатам. Сначала мы проверили справедливость формулы (0.9) для случайного блуждания по графу Эрдоса. Для этого мы построили случайный граф Эрдоса, нашли с помощью метода итераций стационарное распределение случайного блуждания и вычислили эмпирические функции распределения как элементов стационарного решения, так и координационных чисел вершин. Обе функции распределения представлены на графике (1). Видно, что две кривые полностью совпадают. Этот результат подтверждает формулу (0.9).

Напомним, что формула (0.9) была выведена в предположении, что случайное блуждание происходит по неориентированному графу. Встает вопрос: а как будут соотноситься функции распределения элементов стационарного решения и координационных чисел вершин для случайного блуждания по ориентированному графу? Для ответа на этот вопрос мы построили ориентированный граф Эрдоса и вычислили искомые функции распределения. Результат представлен на графике (2). Видно, что две кривые с поразительной точностью совпадают и опять выполняется соотношение (0.9). Таким образом мы видим, что для случайного блуждания по ориентированному графу Эрдоса функция распределения стационарного решения полностью определяется соотношением (0.9), несмотря на то, что это соотношение было выведено при более слабых предположениях относительно структуры графа. Можно предположить, что на функцию распределения элементов стационарного решения не влияет какой мы имеем граф, ориентированный или нет.

Проверим теперь как выполняется соотношение (0.9) для других моделей случайных графов. На рис. (3), (4) представлены кривые для функций распределения элементов стационарного решения и координационных чисел случайного блуждания по графам Ваттса-Строгатса и Барабаши-Альберта. Видно, что для обоих типов случайных графов уравнение (0.9) выполняется.

Таким образом мы видим, что соотношение между функциями распределения эле-

ментов стационарного решения и координационных чисел графа выполняется для случайных блужданий по всем трем моделям случайных графов, причем даже в случае случайного блуждания по ориентированному графу.

Рассмотрим теперь случайные блуждания с переходом на произвольную вершину. Глядя на формулу (0.10) можно предположить, что для стационарного распределения случайного блуждания с переходом могло бы быть справедливо соотношение:

$$\pi_i = \frac{\gamma}{N} + (1 - \gamma) \pi_i^0 = \frac{\gamma}{N} + (1 - \gamma) \frac{d_i}{2m}, \quad (0.12)$$

где π_i^0 есть стационарное распределение для случайного блуждания по тому же графу без перехода. В общем случае это предположение неверно. чтобы в этом убедиться достаточно проверить, что вектор π_i не является собственным вектором стохастической матрицы. Однако наш анализ показывает, что на основе формулы (0.12) можно построить соотношение для функций распределения элементов стационарного решения $p_\pi(x)$ и координационных чисел вершин $p_d(x)$, которое будет выполняться:

$$p_\pi(x) = p_d\left(\frac{\gamma}{N} + (1 - \gamma)\frac{x}{2m}\right). \quad (0.13)$$

Продемонстрируем справедливость выражения (0.13) на примере трех моделей случайных графов. Для этого мы с помощью метода итераций нашли стационарные решения случайного блуждания с переходом для всех трех моделей графов. Значение параметра γ мы выбрали равным 0.15. Это было сделано не случайно. Случайное блуждание с переходом на произвольную вершину часто используется в моделях перемещения пользователя по графу веб-документов. Исследования показывают, что в среднем пользователь, попав на страницу, примерно 6 раз переходит по ссылкам внутри сайта прежде чем уйти с этой страницы на другую. И именно такое значение γ используется в модели *Pagerank* в системе Google. На рисунках (5), (6), (7) представлены функции распределения координационных чисел вершин и элементов стационарного решения для трех моделей случайных графов. Видно, что эти функции очень хорошо совпадают и соотношение (0.13) выполняется.

Выводы.

Мы исследовали функцию распределения элементов стационарного распределения случайного блуждания по трем наиболее популярным моделям случайных графов. Эти модели различаются между собой, имеют разное распределение координационных чисел

вершин, различаются по областям приложений. Кроме того, мы рассматривали случайное блуждание как по неориентированному графу, так и по ориентированному. Во всех случаях функции распределения элементов стационарного распределения случайного блуждания зависели только от функции распределения координационных чисел вершин и выполнялось соотношение (0.9). Можно предположить, что это соотношение является общей зависимостью для всех типов графов, однако четкого теоретического обоснования этого факта нам пока неизвестно.

Мы также рассмотрели случайное блуждание с переходом на произвольную вершину. В этом случае мы тоже обнаружили соотношение между функциями распределения стационарного решения и координационных чисел вершин - соотношение (0.13).

Соотношения (0.9) и (0.13) позволяют установить взаимосвязь между свойствами стационарных решений случайных блужданий и свойствами топологии графа. Так можно связать степенные законы популярности веб-документов [8, 3, 13, 5, 22] со степенным распределением координационных чисел вершин [15, 16, 1, 4, 2, 19, 21]. Многочисленные исследования показывают, что координационные числа d_i в графе веб-документов распределены по степенному закону $1/d^\alpha$, с показателем $\alpha \simeq 1.1 \pm 0.3$ (обычно авторы приводят показатель функции плотности вероятности, который на 1 больше). А исследования популярности веб-документов показывают, что ранговое распределение веб-документов может описываться степенным законом с показателем $s \simeq 1$. Ранговое распределение документов по популярности есть ни что иное, как функция распределения популярности веб-документов в "перевернутых" координатах (подробный анализ свойств ранговых распределений веб-документов представлен в работе [22]). Если в качестве модели поведения пользователя взять случайное блуждание по графу веб-документов, то мерой популярности документа будет как раз элемент стационарного решения случайного блуждания - вероятность нахождения пользователя на выбранном документе. В такой модели наши соотношения (0.9) и (0.13) как раз связывают популярность документа и топологию графа веб-документов и лежат в согласии с экспериментом.

В пользу нашего объяснения степенных законов популярности веб-документов может служить следующий факт. Очевидно, что поисковые системы оказывают огромное влияние на популярность ресурсов. Популярен будет тот ресурс, который поисковая система выдает первым на ключевой запрос. Алгоритмы ранжирования большинства поисковых систем основаны на случайном блуждании по графу веб-документов, и это может служить "положительной обратной связью". Действительно, те веб-документы,

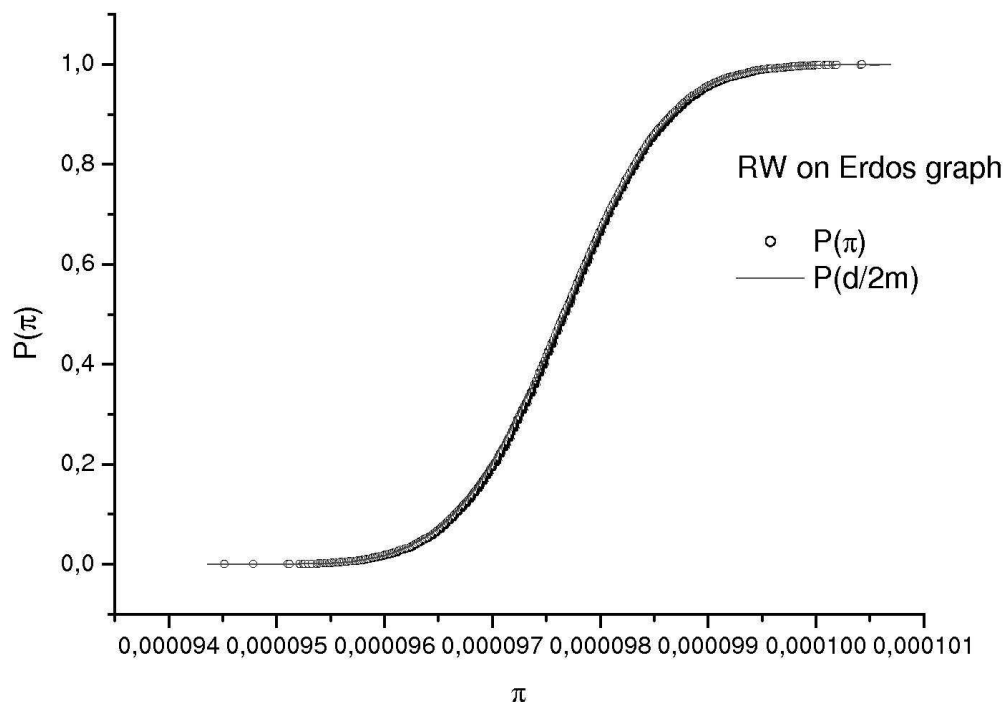


Рис. 1. Рис. 1 Случайное блуждание по графу Эрдоша. Количество вершин $N = 10240$, вероятность соединения двух вершин ребром $\lambda = 0.6$. Распределение элементов стационарного решения показано кружками, а распределение координационных чисел вершин - сплошной линией.

которые имеют больший вес по мнению поисковой системы (большее значение соответствующего элемента стационарного решения) приобретают большую популярность среди пользователей, вследствие работы самой поисковой системы. Таким образом поисковые системы "подстраивают" реальное распределение популярности веб-документов под алгоритм случайного блуждания. Можно сказать, что топология графа веб-документов определяет их популярность.

Заметим, что ясной модели, объясняющей степенные распределения популярности веб-документов до сих пор предложено не было.

Автор благодарит Льва Николаевича Шура и Сергея Александровича Крашакова за неоценимую помощь в работе над статьей. Их ценные советы, дополнения и замечания оказали огромное влияние на подготовку этой работы.

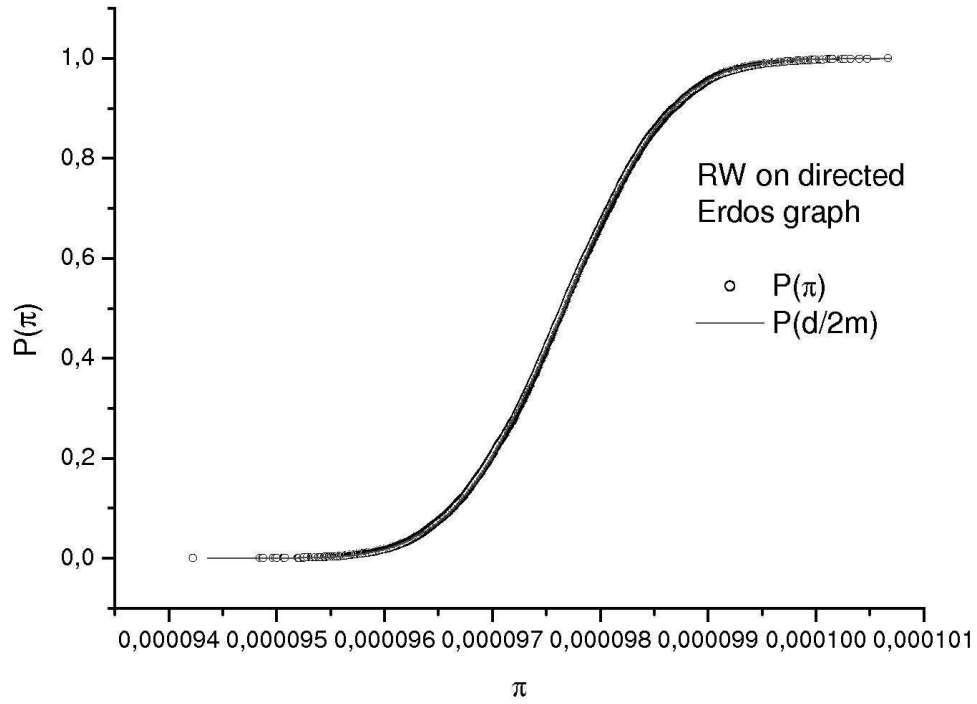


Рис. 2. Рис. 2. Случайное блуждание по ориентированному графу Эрдоша. Количество вершин $N = 10240$, вероятность соединения двух вершин ребром $\lambda = 0.6$. Распределение элементов стационарного решения показано кружками, а распределение координационных чисел вершин - сплошной линией.

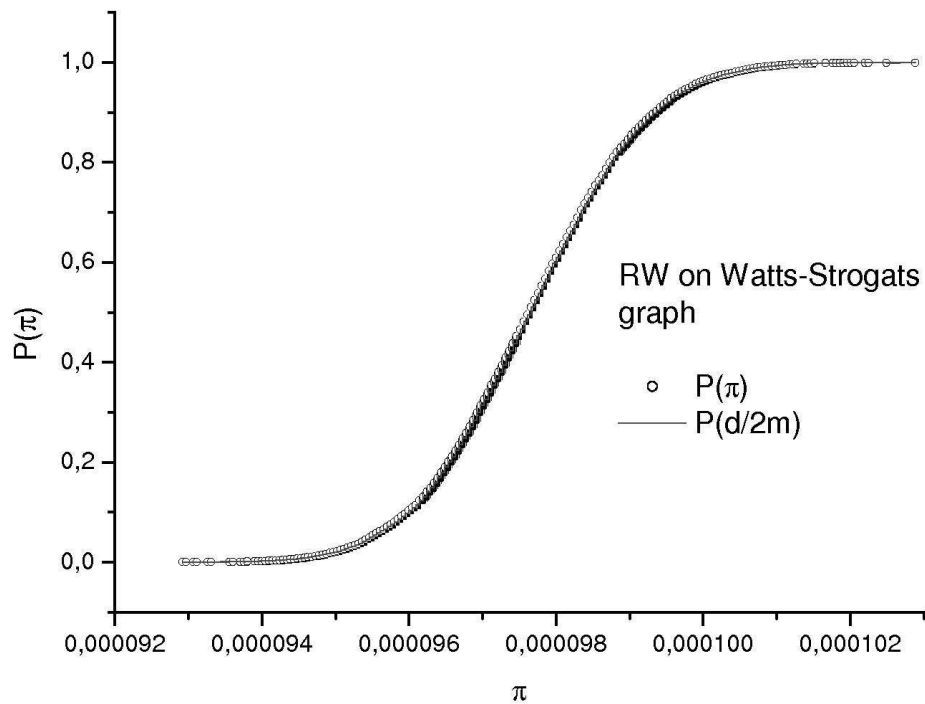


Рис. 3. Рис. 3. Случайное блуждание по графу Ваттса-Строгатса. Количество вершин $N = 10240$, начальное координационное число вершины $K = 1000$, вероятность отсоединения ребра $p = 0.5$. Распределение элементов стационарного решения показано кружками, а распределение координационных чисел вершин - сплошной линией.

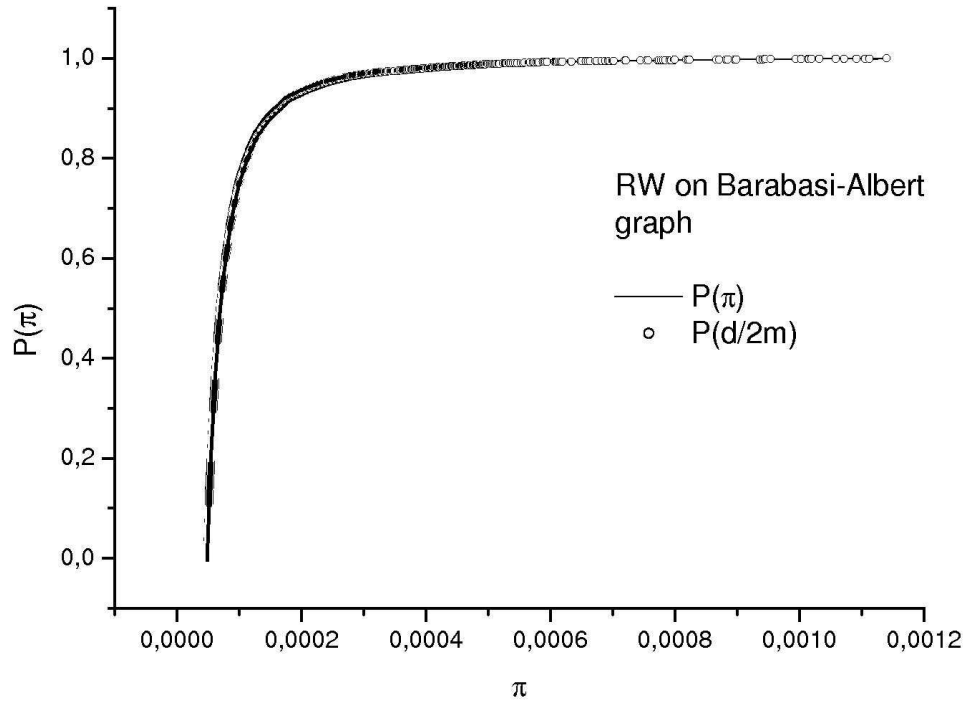


Рис. 4. Рис. 4. Случайное блуждание по графу Барабаша-Альберта. Количество вершин $N = 10240$. Координационное число новой вершины $M = M_0 = 50$. Распределение элементов стационарного решения показано кружками, а распределение координационных чисел вершин - сплошной линией.

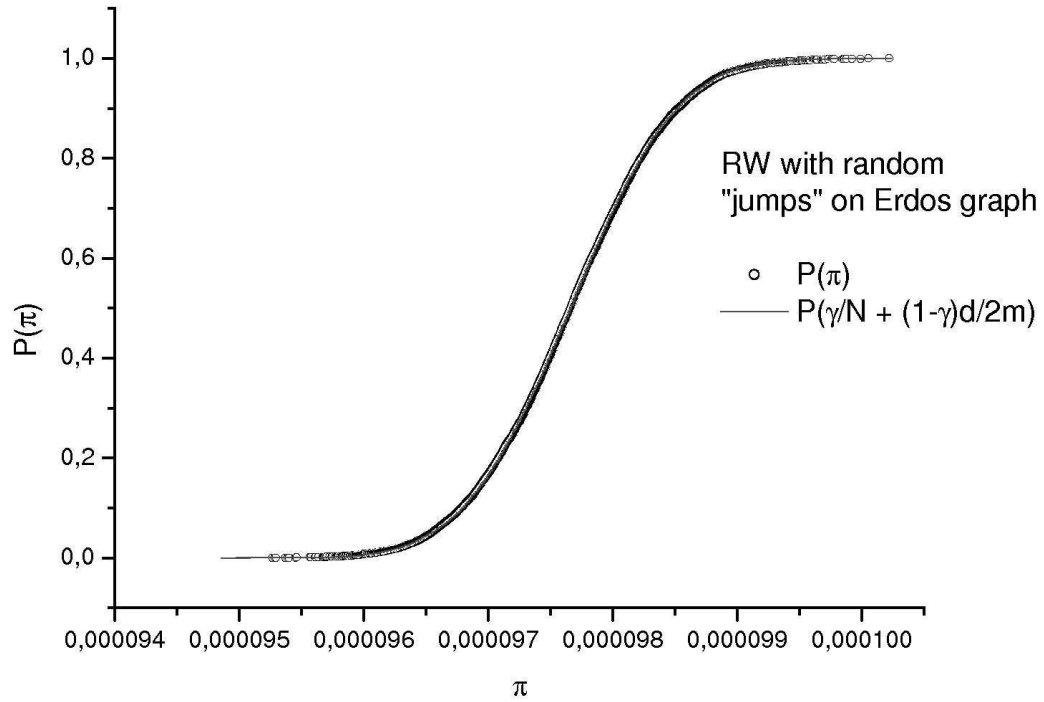


Рис. 5. Рис. 5. Случайное блуждание со случайным переходом по графу Эрдоша. Количество вершин $N = 10240$, вероятность соединения двух вершин ребром $\lambda = 0.6$. $N=10240$. Распределение элементов стационарного решения показано кружками, а распределение координационных чисел вершин - сплошной линией.

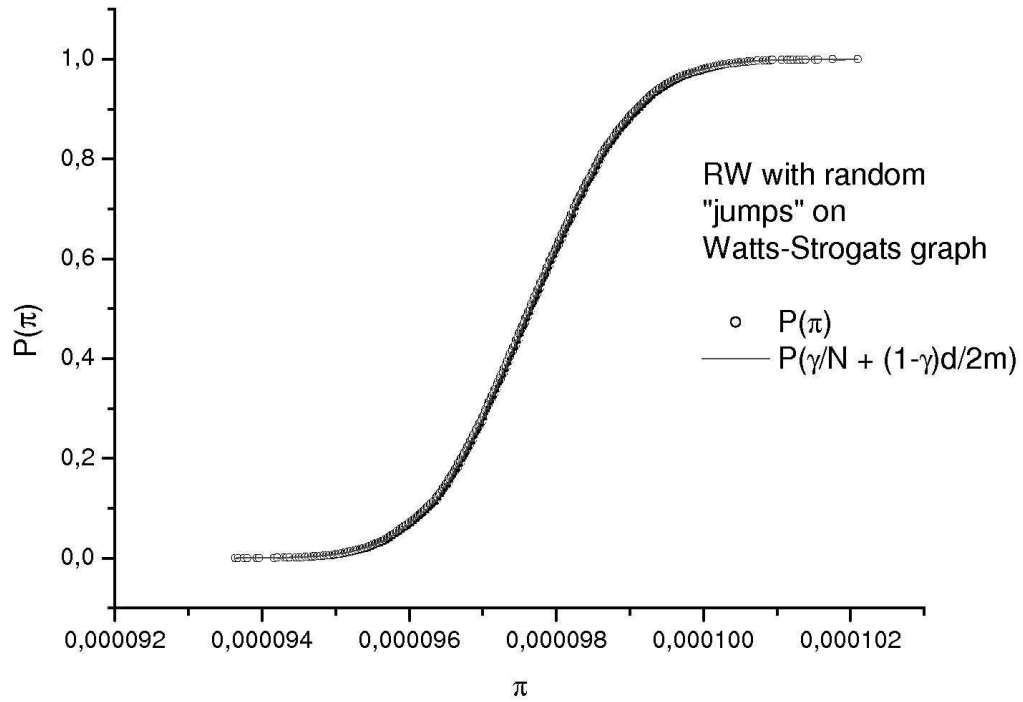


Рис. 6. Рис. 6. Случайное блуждание со случайным переходом по графу Ватса-Строгатса. Количество вершин $N = 10240$, начальное координационное число вершины $K = 1000$, вероятность отсоединения ребра $p = 0.5$. Распределение элементов стационарного решения показано кружками, а распределение координационных чисел вершин - сплошной линией.

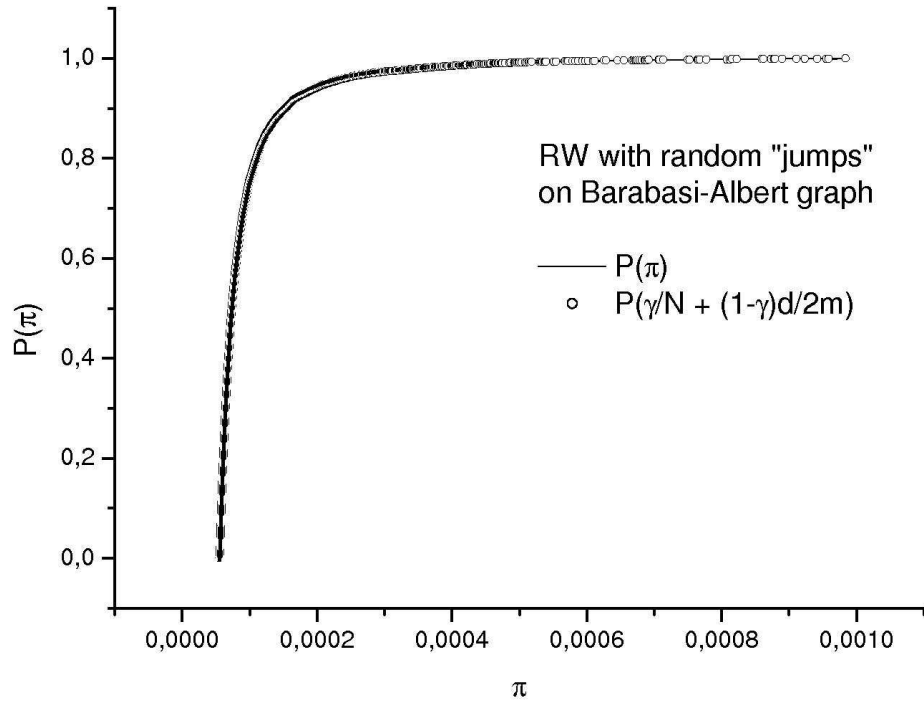


Рис. 7. Рис. 7. Случайное блуждание по графу Барабаша-Альберта. Количество вершин $N = 10240$. Координационное число новой вершины $M = M_0 = 50$. Распределение элементов стационарного решения показано кружками, а распределение координационных чисел вершин - сплошной линией.

Список литературы

- [1] L. Adamin, B. Huberman, *Power-law distribution of the World Wide Web*, Science 287, 2115, 2000.
- [2] A. Barabasi, R. Albert, *Emergence of Scaling in Random Networks*, Science 286, 509-512, 1999.
- [3] L. Breslau, P. Chao, L. Fan, G. Phillips, S. Shenker, *Web Caching and Zipf-like Distributions: Evidence and Possible Implications*, Proc. IEEE INFOCOM '99: 18th Annual Joint Conference of the IEEE Computer and Communications Societies, Volume: 1, p. 126-134 1999.
- [4] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rahagopalan, R. Stata, A. Tomkins, J. Wiener, *Graph Structure in the Web*, Computer Networks, 33, 309-320, 2000.
- [5] R. P. Doyle, J. S. Chase, S. Gadde, A. M. Vahdat. *The trickle-down effect: Web caching and server request distribution*. Computer Communications, **25**, 345-356 (2002).
- [6] P. Erdos, A. Renyj, *On Random Graphs*, Publicationes Mathematicae, **6**, 290-297 (1959).
- [7] , Garfield E., *Citation analysis as a tool in journal evaluation*, Science 178, 471-79, 1972
- [8] S. Glassman, *A caching Relay for the World Wide Web*. Proc. 1st International Conference on the World-Wide Web, CERN, Geneva (Switzerland), May 1994. Elsevier Science, p.69-76, also in Computer Networks and ISDN Systems, 27(2), 165-173 (1994)
- [9] S. Brin, L. Page, *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, Computer Networks and ISDN Systems, 30(1-7), 107-117 (1998).
- [10] T. H. Haveliwala, *Efficient computation of PageRank*, Technical report, Computer Scientist Department, Stanford University, 1999.
- [11] Hubbel C. H., *An input-output approach to clique identification*, Sociometry 28, 377-399, 1965
- [12] Katz, L., *A new status index derived from sociometric analysis*, Psychometrika 18, 39 - 43, 1953

- [13] T. Kelly, J. Mogul, *Aliasing on the world wide web: prevalence and performance implications*. Proc. 11th Int. WWW Conf., Honolulu, May 2002, ACM Press, pp.281 - 292.
- [14] J. Kleinberg. *Authoritative sources in a hyperlinked environment*, Proc. 9th ACM-SIAM Symposium on Discrete Algorithms, 1998. Extended version in Journal of the ACM 46(5), pp. 668-677 (1999), Also appears as IBM Research Report RJ 10076, May 1997
- [15] R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins, *Trawling the Web for Cyber Communities*, Proc. 8th Int. WWW Conf., Toronto, Canada, May 11-14, 1999; Computer Networks, 31(11-16)1481-1493 (1999).
- [16] R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins, *Extracting Large Scale Knowledge from the Web*, Proc. VLDB'99, Sep 1999.
- [17] A. Langville, C. Meyer, *Fiddling with PageRank*, CRSC Tech Report crsc-tr03-34.
- [18] L. Lovasz, *Random Walks on Graphs: A Survey*, in: Combinatorics, Paul Erdos is Eighty, Vol. 2 (ed. D. Miklos, V. T. Sos, T. Szonyi), Janos Bolyai Mathematical Society, Budapest, 1996, 353–398.
- [19] D. Pennock, G. Flake, S. Lawrence, E. Glover, C. Giles, *Winners don't take all: Characterizing the competition for link on the web*, Proc Natl. Acad. Sci. USA, 99, 5207-5211, 2002.
- [20] Pinski, G., and Narin F., *Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics*, Inf. Proc. Manage. 12, 297-312, 1976
- [21] B. Tadic, *Adaptive random walks on the class of Web graphs*, Eur. Phys J. B 23, 221-228, 2001.
- [22] A. B. Teslyuk, S. A. Krashakov, L.N. Shchur, *On the Universality of rank-distributions of Web-Site Populattities*, arXiv: cs.NI/0404010
- [23] D. J. Watts, S. H. Strogatz, *Collective dynamics of 'small-world' networks*, Nature, June 1998, vol. 393, pp. 440-442.