

# Библиография как источник исследования процессов научной коммуникации

С. А. ИВАНОВ

Технический институт культуры,  
г. Москва, Россия

*Рассматриваются различные виды библиографий, отражающие процессы передачи научной информации в научном сообществе. Информетрический анализ библиографий позволил выявить основные закономерности формирования потока научной информации: экспоненциальный рост публикаций, рассеяние статей по широкому кругу изданий. Определяется динамика "незримых коллективов" ученых и редакционных коллегий, участвующих в коммуникационных процессах в одной тематической области.*

## ВВЕДЕНИЕ

Появившаяся статья Ю. Дина [1] о процессах коммуникации в науке и количественных методах исследования в информатике, ориентированных на формальные связи между документами (цитирование), требует более детального изложения методов проведения информетрических исследований. Главным становится вопрос об определении основных закономерностей создания и использования научной информации на основе изучения динамики научных исследований в различных областях науки. В рамках этого процесса наибольший интерес для научного сообщества представляет установление объективных закономерностей функционирования мировой системы научной коммуникации. Тогда становятся понятными многие частные исследования и эмпирические оценки коммуникационных процессов (указателей цитирования статей различных журналов и авторов, факторов влияния журналов на развитие науки и т. д.), проводимые, в основном, в Институте научной информации Ю. Гарфилда в США.

Эти исследования не касаются главного вопроса: какими устойчивыми закономерностями обладает мировая научная коммуникационная система? В основе научных исследований лежит тематический принцип: науку можно представить как большое число сообществ ученых, работающих в разных тематических областях. И совсем не обязателен охват цитированием всех публикаций, входящих в ту или иную тематическую область.

Поэтому гораздо удобнее рассматривать множество публикаций, объединенных одной тематикой. Такое множество документов более адекватно отражает процессы научной коммуникации. Многие значимые научные документы не цитируются по самым разным причинам, но нет оснований считать, что их содержание не известно ученым, занимающимся смежной или аналогичной тематикой. В этом смысле исследование кластеров цитируемых документов в одной тематической области не позволяет исследовать всю совокупность коммуникационных процессов в научном сообществе, занимающемся данной проблемой. На этом пути до сегодняшнего времени не выявлено сколько-нибудь ощутимых результатов по нахождению объективных закономерностей функционирования мировой системы научной коммуникации.

Исключение составляют работы [2, 3], где исследуются гиперболические распределения размеров групп кластеров статей, различающихся по числу процитированных статей других авторов. Эти исследования проводились в базах данных Science Citation Index (SCI) и Social Science Citation Index

(SSCI) за 1984 г. общим объемом 660 тыс. статей. Построенные распределения численности групп кластеров позволили автору высказать предположение, что такая система кластеров представляет собой фрактальную структуру с дробной размерностью.

Объектом исследований в информатике должно быть множество документов, объединенных не по формальным признакам (цитируемости), а на основании семантических связей между документами. Такими информационными объектами являются тематические библиографии, объединяющие научные документы по их содержанию. Они отражают процессы развития одной тематической области и являются объективным документальным подтверждением коммуникационных процессов между учеными.

Важная роль библиографий состоит в том, что для новых поколений ученых представляет интерес не только интегрированное научное знание, аккумулированное в монографиях, но также многие идеи, рассеянные в большом числе оригинальных работ.

Составление таких библиографий, как правило, происходит в момент наибольшего интереса научного сообщества к данной тематике, когда в тематическую область приходят новые поколения ученых.

Такие исследования актуальны в настоящее время в связи с большими объемами документальной информации, накопленными человечеством. Самая большая Библиотека конгресса США содержит 21 млн единиц хранения, из них переведено в электронную форму примерно 5 млн документов и доступно через Интернет [4].

Появление в конце XX в. мировой информационной сети Интернет и электронных журналов и библиотек в этой сети не изменило информационных потребностей ученых, изменилась только информационная среда. Сегодня практически вся научная информация стала доступной большинству ученых [5]. Электронные средства обработки информации значительно облегчили доступ к накопленным научным сообществом информационным массивам документов по различным отраслям знания.

Однако проблемы семантического поиска информации в Интернете остались те же, поэтому осталась и актуальной проблема рассеяния научной информации. Можно высказать предположение, что она во многом связана с закономерностями объединения интеллектуальных усилий научного сообщества над решением конкретных научных проблем.

## ИССЛЕДОВАНИЕ ИНФОРМАЦИОННЫХ ПОТОКОВ

Исторически сложилось так, что в информатике наибольшее распространение получили исследования информационных потоков.

Многие ученые анализировали динамику публикации научных документов по различным научным направлениям и на основе такого анализа делали прогнозы развития научных исследований. Эти методы были основаны на применении научных публикаций в качестве показателей развития отдельных научных областей. В качестве примера таких исследований можно привести работу [6], где исследовался информационный поток научных документов по полупроводникам, являющимся основным элементом информационной индустрии. В работе исследовался рост информационного потока в базе данных INSPEC, предназначенной для информационного обслуживания физиков и инженеров и содержащей более 6 млн записей из 60 стран мира.

Информационный поток за 1978–1997 гг. насчитывает 275254 документа, среди которых 184233 статьи написаны 125990 учеными (с учетом соавторов) и опубликованы в 1877 журналах.

Статьи в журналах вместе с докладами на конференциях (88807) составили более 99% всего тематического информационного потока.

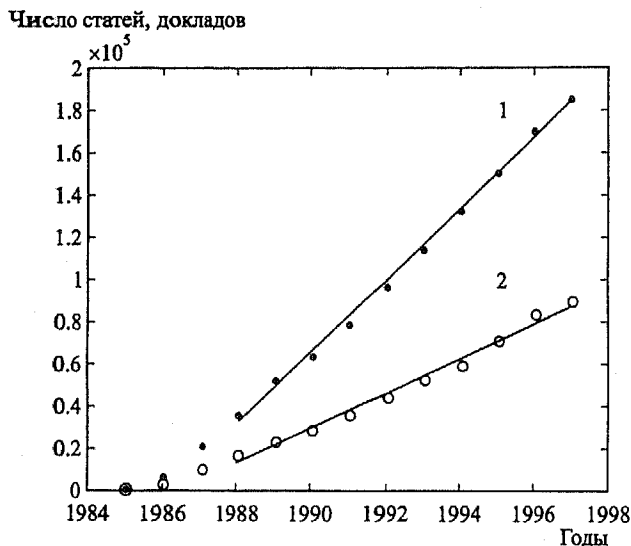


Рис. 1. Увеличение числа статей (1)  $\nu_{ст}=16700/\text{год}$  и докладов на конференциях (2)  $\nu_{д}=8150/\text{год}$  по полупроводникам в 1986–1997 гг.

Для этого информационного массива, включающего все публикации по полупроводникам, наблюдаются линейные зависимости (рис. 1) роста во времени:

числа статей

$$N_{ст}(t) = N_{ст}(t_0 = 1988) + \nu_{ст}(t - t_0), \quad \nu_{ст} = 16700/\text{год},$$

числа докладов

$$N_{д}(t) = N_{д}(t_0 = 1988) + \nu_{д}(t - t_0), \quad \nu_{д} = 8150/\text{год}.$$

На десятилетнем интервале наблюдаются стационарные информационные потоки, имеющие небольшие флуктуации, составляющие единицы процентов.

Авторы также отмечают, что даже такие большие информационные массивы подчиняются гиперболическим законам рассеяния статей по журналам С. Бредфорда и продуктивности ученых А. Лотки.

Изученные информационные потоки практически не изменялись в десятилетнем периоде, однако многие информационные потоки в науке изменяются во времени, что будет показано на различных примерах.

## МИРОВАЯ СИСТЕМА НАУЧНОЙ КОММУНИКАЦИИ

Научные документы на бумажных носителях доступны научному сообществу в библиотеках или в электронном виде через Интернет.

Можно выделить две основные системы подготовки научных документов:

— мировая система периодических и продолжающихся изданий,

— мировая система научных книжных издательств.

Каждая научная группа выделяет из потока этих двух систем документы, относящиеся к тематике своих исследований, создает собственные библиографии, отражающие их взгляд на проблему. Эти библиографии представляют интерес для всего научного сообщества.

Каждая новая статья расширяет область исследований и одновременно интегрирует уже ранее полученные знания, тем самым увеличивая размер предметной области. Эти закономерности можно проследить, анализируя тематические массивы документов

## ИСТОРИЯ ИНФОРМЕТРИЧЕСКИХ ИССЛЕДОВАНИЙ

В информатике хорошо известны первые работы по статистическим исследованиям тематических информационных массивов на коротких временных интервалах в 2–5 лет: статья Лотки о распределении ученых по продуктивности [7], работа Бредфорда о рассеянии статей по периодическим изданиям [8], монография Ципфа, обобщающая основные методы статистического анализа различных социальных систем [9] и предлагающая ожившую статистическую трактовку закона Бредфорда. История этих исследований отражена в наиболее полной тематической библиографии [10].

Дальнейшее развитие методика статистических исследований информационных массивов получила в статье [11], где для документов по библиометрическим исследованиям из компьютерной базы данных было построено пространственно-временное распределение публикаций по периодическим изданиям.

Исторический анализ показал, что данная тема не привлекла внимание широкого круга исследователей. С момента появления статьи А. Лотки в 1926 г. до 1969 г. было напечатано всего 65 статей, а затем только в одном 1969 г. было опубликовано 17 статей на эту тему.

Этот год можно считать отправной точкой перехода библиометрических исследований в информатике из патентного периода к широкому фронту исследований, сопровождающихся экспоненциальным увеличением числа статей. Информационный массив, исследуемый в статье [11], содержал 191 публикацию из 19 наиболее продуктивных изданий за период 1941–1982 гг.

Динамика числа статей и журналов, где эти статьи напечатаны (1965–1982 гг.) аппроксимируется экспоненциальными зависимостями:

$$N_{ст}(t) = N_{ст}(t_0 = 1965) \exp\{\lambda_{ст}(t - t_0)\}, \quad \lambda_{ст} = 0,13/\text{год}.$$

Среднее время удвоения числа статей составляет — 5 лет.

$$N_{ж}(t) = N_{ж}(t_0 = 1965) \exp\{\lambda_{ж}(t - t_0)\}, \lambda_{ж} = 0,06/\text{год}.$$

Среднее время удвоения числа журналов — 11 лет.

Ранговое распределение для первых 19 наиболее продуктивных изданий выражается формулой:

$$N_{ст}(r) = \frac{N_{ст}(r=1)}{r^\gamma}, \gamma = 0,86,$$

диапазон рангов для аппроксимации гиперболической функцией составляет ( $1 \leq r \leq 16$ ).

На рис. 2 показано пространственно-временное распределение статей из основных профильных изданий за 40 лет на трех осях: время, ранг периодического издания, число публикаций в издании. Представленная в трехмерном пространстве поверхность наглядно показывает, что заполнение статьями по этой тематике периодических и продолжающихся изданий происходило в соответствии с гиперболическим распределением журналов.

Число статей

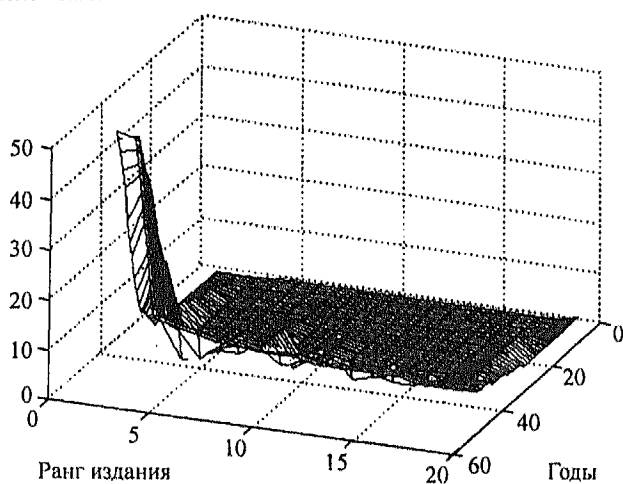


Рис. 2. Пространственно-временное распределение журналов по числу статей

Дальнейшим шагом в развитии информетрических исследований стало определение основного объекта исследований — полной библиографии, включающей, по мнению составителей, все опубликованные документы по тематике исследований [12]. Статистический анализ этих библиографий показал, что для них действует закон рассеяния статей Бредфорда. Такие библиографии являются уникальными информационными объектами, для их подготовки требуются усилия большого числа ученых. Их создание стало возможным с появлением Интернета, и они представляют наибольший интерес для информетрических исследований.

## ОБЪЕКТ ИССЛЕДОВАНИЙ

Анализ информационных массивов в различных областях естественных наук позволил выдвинуть гипотезу о едином механизме формирования различных информационных массивов, определяемом процессами коммуникации. Ранее исследования проводились на информационных массивах, тематика которых не была точно определена. Как правило, это были годовые авторские и журнальные указатели в реферативных журналах.

В настоящей статье рассматриваются тематические коллекции документов, обладающие разной степенью интеграции научных знаний в тематической области. Ими являются:

- списки трудов ученых, проработавших всю активную научную жизнь в одной тематике;
- тематические библиографии при научных обзорах;
- библиографии при первых монографиях по новым научным направлениям;
- полные библиографии, созданные известными специалистами.

Исследование динамики коммуникационных процессов проводится на массивах документов, формирующих тематические области.

## СПИСКИ ТРУДОВ УЧЕНЫХ

Один из руководителей международной ассоциации History & Computing, ведущий специалист в области исторической информатики проф. МГУ Л. И. Бородкин любезно предоставил автору для библиометрического анализа список своих научных трудов, включающий 215 наименований за 1975–2000 гг. (4 монографии, 3 учебных пособия, 51 статья из периодических изданий, 97 статей из неперидических сборников, 43 опубликованных тезиса выступлений на конференциях и симпозиумах и т. д.).

За период 1976–2000 гг. им была опубликована 51 статья в 23 журналах. Увеличение во времени числа статей и журналов может быть аппроксимировано линейной функцией:

$$N_{ст}(t) = N_{ст}(t_0 = 1985) + \nu_{ст}(t - t_0).$$

Средняя скорость увеличения числа статей в год:  $\nu_{ст} = 2,0/\text{год}$ , ( $N_{ст}(t_0 = 1985) = 16$ ), т. е. ежегодно в среднем публикует по две статьи.

Увеличение во времени числа журналов может быть аппроксимировано также линейной функцией:

$$N_{ж}(t) = N_{ж}(t_0 = 1985) + \nu_{ж}(t - t_0).$$

Средняя скорость увеличения числа изданий:  $\nu_{ж} = 1,31/\text{год}$ , ( $N_{ж}(t_0 = 1985) = 4$ ), т. е. ежегодно Л. И. Бородкин публикует одну статью в новом для него издании.

Ранговое распределение журналов по числу опубликованных статей представляется гиперболической функцией:

$$N_{ст}(r) = \frac{N_{ст}(r=1)}{r^\gamma}, (1 \leq r \leq 20), \gamma = 0,92.$$

Значение  $\gamma \approx 1,0$  свидетельствует о соответствии распределения классическому закону Ципфа. В журнале первого ранга с максимальным числом статей ("История СССР") опубликовано 11 статей, в журнале второго ранга ("Новая и новейшая история") — 8 статей, третьего ранга ("Экономическая история. Обзорение") — 4 статьи.

В гуманитарных науках широкое распространение получили тематические сборники (неперидические издания), которые отражают научные интересы различных институтов и научных школ. В 1976–2000 гг. Л. И. Бородкин опубликовал в них 97 статей. Увеличение числа статей во времени в сборниках происходит по экспоненциальной зависимости:

$$N_{ст}(t) = N_{ст}(t_0 = 1980) \exp\{\lambda_{ст}(t - 1980)\}, \lambda_{ст} = 0,13/\text{год}.$$

Среднее время удвоения числа статей составляет 5 лет при начальном значении  $N_{ст}(t_0 = 1980) = 8$ .

Экспоненциальное увеличение числа статей в различных сборниках свидетельствует о проявлении интереса различных научных школ к тематике работы ассоциации. Статьи в сборниках свидетельствуют об устойчивых контактах ассоциации с различными историческими институтами и о растущем авторитете ученого в научном сообществе.

Личные контакты ученого в 1976–2000 гг. нашли отражение в 43 выступлениях на различных конференциях и симпозиумах. Увеличение числа опубликованных тезисов выступлений соответствует линейной функции:

$$N_{\text{тез}}(t) = N_{\text{тез}}(t_0 = 1980) + \nu_{\text{тез}}(t - 1980), \nu_{\text{тез}} = 1,9/\text{год}.$$

В среднем составляет два выступления ежегодно при начальном значении  $N_{\text{тез}}(t_0 = 1980) = 6$ .

Библиометрический анализ списка трудов Л. И. Бородинки показал закономерности отражения научной активности ученого в мировой системе научной коммуникации, использование им различных каналов для представления результатов своих исследований научному сообществу.

## БИБЛИОГРАФИЯ

### ПРИ ТЕМАТИЧЕСКИХ ОБЗОРАХ

В 1989 г. был опубликован обзор [13], посвященный открытому Бредфордом закону рассеяния научных публикаций в мировой системе периодических изданий. Этот закон стал предметом активного исследовательского фронта ученых в библиометрии с середины 60-х гг. XX в. В библиографии обзора собраны наиболее значимые статьи за 1934–1987 гг., относящиеся к различным аспектам закона Бредфорда.

Закон рассеяния научной информации в периодических изданиях Бредфорда — один из наиболее известных и изученных законов информатики. Сформулированный более 50 лет назад, он привлекает внимание ученых и сегодня. Закон Бредфорда сыграл значительную роль в развитии теории научно-технической информации и стал одной из теоретических основ информатики. Он дал понимание функционирования больших социальных систем, создающих и представляющих человечеству научные знания. Закон рассеяния научной информации находит и сегодня применение для оценки документальных потоков научной информации по отдельным тематическим направлениям.

Этот обзор интересен тем, что для него отобраны статьи, касающиеся только интегрального распределения Бредфорда, хотя аналогичные исследования проводятся в рамках эмпирических законов Ципфа и Лотки. Это объясняется позицией автора, отмечающего статистический характер других законов, хотя распределение Бредфорда является интегральной формой рангового распределения Ципфа.

Автор выделяет три основных аспекта, отраженных в обзоре:

- формулировка этого эмпирического закона,
- числовые оценки параметров распределения этого закона,
- соотношение других наукометрических распределений с законом Бредфорда.

В библиографию входят 73 публикации из 20 периодических и продолжающихся изданий за 1934–1987 гг., а также другие виды научных документов. Латентный период составил 30 лет (1934–1963 гг.).

Зависимость увеличения числа публикаций за 20 лет (1963–1982 гг.):

$$N_{\text{ст}}(t) = N_{\text{ст}}(t_0 = 1963)\exp\{\lambda_{\text{ст}}(t - t_0)\}, \lambda_{\text{ст}} = 0,16/\text{год}.$$

Время удвоения числа статей равно 5 годам, при начальном значении  $N_{\text{ст}}(t_0 = 1963) = 4$ .

А зависимость увеличения числа изданий во времени:

$$N_{\text{ж}}(t) = N_{\text{ж}}(t_0 = 1963)\exp\{\lambda_{\text{ж}}(t - t_0)\}, \lambda_{\text{ж}} = 0,11/\text{год}.$$

Время удвоения числа журналов равно 6 годам, при начальном значении  $N_{\text{ж}}(t_0 = 1963) = 3$ .

Ранговое распределение периодических изданий выражается гиперболической функцией:

$$N_{\text{ст}}(r) = A/r^\gamma, (0 \leq r \leq 10), \gamma = 1,26.$$

Значение  $\gamma > 1,0$  свидетельствует о концентрации статей в ряде профильных изданий по информатике, образованных в 60–70-е гг.: “Journal of Documentation”, “Journal of the American Society for Information Science” и др.

Аллометрическая зависимость между числом опубликованных статей и числом журналов имеет вид:

$$N_{\text{ст}} = AN_{\text{ж}}^\rho, \rho = 1,62, A — константа.$$

Значение показателя степени  $\rho = 1,62$  больше, чем классическое значение  $\rho = 1,2$ , соответствующее  $\gamma = 1,0$ , что также подтверждает концентрацию статей в ряде профильных изданий.

Можно отметить, что, несмотря на небольшой объем информационного массива (73 статьи за 54 года), ранговое распределение журналов подчиняется закону Ципфа.

## БИБЛИОГРАФИИ

### ПРИ ПЕРВЫХ МОНОГРАФИЯХ

Монография Т. Гианика и В. И. Пасечника “Бислойные липидные пленки: структура и механические свойства” [14] относится к одному из современных и перспективных направлений в биофизике — исследованию механических свойств модельных липидных пленок, применяемых при создании биосенсоров. Авторы, являющиеся пионерами в изучении электрических и механических свойств бислойных липидных пленок в нашей стране и Европе, обобщают результаты своих теоретических и экспериментальных исследований более чем за двадцатилетний период.

Библиография включает описания 337 статей, опубликованных в 94 периодических и продолжающихся изданиях в 1925–1995 гг., а также описания диссертаций, монографий, статей в сборниках и т. д. Латентный период составил 35 лет (1925–1959 гг.). Рост числа статей и журналов в 1959–1995 гг. происходит по экспоненте, среднее время удвоения числа статей  $\approx 4$  года, а среднее время удвоения числа журналов, где публикуются статьи, составляет  $\approx 6$  лет.

Ранговое распределение периодических и продолжающихся изданий выражается гиперболической функцией:

$$N_{\text{ст}}(r) = A/r^\gamma, (2 \leq r \leq 60), \gamma = 1,13.$$

Значение  $\gamma = 1,13$  близко к  $\gamma = 1,0$ , что соответствует классическому распределению Ципфа.

Увеличение числа цитированных книг происходит по экспоненте на интервале 1955–1975 гг.:

$$N_m(t) = N_m(t_0 = 1955) \exp\{\lambda_m(t - t_0)\}, \lambda_m = 0,14/\text{год}.$$

Среднее время удвоения числа цитированных книг  $\approx 5$  лет.

Для этой библиографии по липидным пленкам выполняются основные наукометрические закономерности.

Монография П. Халмоща “Теория меры” [15] — одна из немногих книг по чистой математике, оказавшая существенное влияние на развитие науки в XX в. Созданный на ее основе функциональный анализ явился теоретическим основанием для решения большого числа некорректно поставленных практических задач по обработке экспериментальных данных — цифровой обработке изображений, компьютерной томографии. Многие основные утверждения меры Лебега, теории измеримых функций и интеграла Лебега являются основаниями современной теории вероятностей.

Библиография включает 50 статей из 22 периодических изданий, 51 статью из различных сборников и описания 20 монографий за 1909–1949 гг.

Увеличение числа статей и числа журналов во времени показано на рис. 3:

$$N_{ст}(t) = N_{ст}(t_0 = 1920) \exp\{\lambda_{ст}(t - t_0)\}, \lambda_{ст} = 0,14/\text{год},$$

среднее время удвоения числа публикаций  $\approx 5$  лет.

$$N_{ж}(t) = N_{ж}(t_0 = 1920) \exp\{\lambda_{ж}(t - t_0)\}, \lambda_{ж} = 0,11/\text{год},$$

среднее время удвоения числа журналов  $\approx 6$  лет.

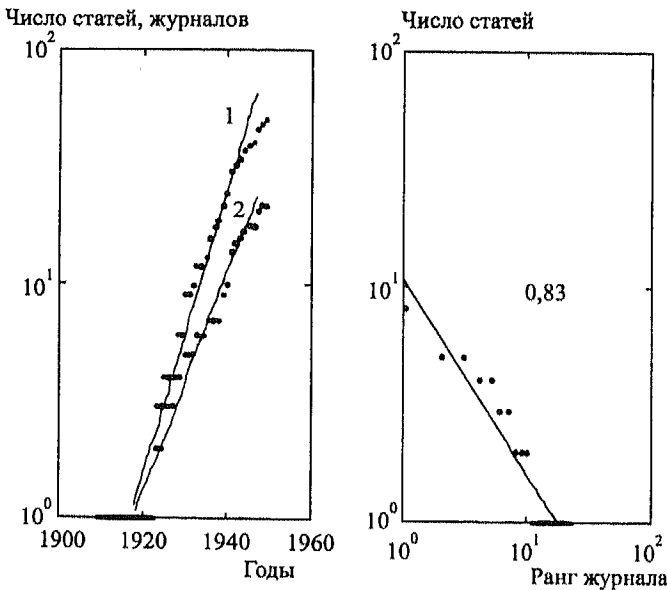


Рис. 3. Увеличение во времени числа статей (1)  $\lambda_{ст} = 0,14/\text{год}$  и числа журналов (2)  $\lambda_{ж} = 0,11/\text{год}$  в книге П. Халмоща за 1909–1949 гг. (50 статей, 22 журнала). Ранговое распределение журналов с  $\gamma = 0,83$

Ранговое распределение журналов по числу опубликованных статей представляется гиперболической функцией:

$$N_{ст}(r) = A/r^\gamma, (2 \leq r \leq 20), \gamma = 0,83.$$

Здесь основные информетрические закономерности выполняются на небольшом информационном массиве, что позволяет предположить, что механизм формирования информационного тематического массива не зависит от его объема.

Появление значимых публикаций распределено по всей системе периодических изданий.

## ПОЛНЫЕ ТЕМАТИЧЕСКИЕ БИБЛИОГРАФИИ

В Интернете на сайте Florida State University в математической электронной библиотеке помещена библиография по закону взаимности в теории чисел [16, 17]. Эта библиография относится к узкой тематической области в чистой математике.

Она содержит ссылки на 1063 научных документа за 1640–2001 гг.: 877 статей 377 авторов опубликовали в 276 изданиях в 1732–2001 гг., 123 книг вышли в 1640–2001 гг., 32 диссертации, 9 опубликованных писем ученых. Письмо П. Ферма по этой тематике датируется 1640 годом (P. de Fermat, Lettre à Mersenne, 25. Dec. 1640; oeuvres II, 212–217).

Эта библиография позволяет рассмотреть процессы научной коммуникации на протяжении более 350 лет. Предложенная методика может стать основанием для проведения статистических исследований полных библиографий, включающих тысячи документов [18].

Полные тематические библиографии далее анализируются по следующим показателям.

*Динамика публикации книг.* В число книг включены монографии, справочники, различные книжные сборники, увеличение числа которых во времени подчиняется экспоненциальному закону (рис. 4) и выражается формулой:

$$N_k(t) = N_k(t_0 = 1840) \exp\{\lambda_k(t - t_0)\}, \lambda_k = 0,01/\text{год}.$$

Среднее время удвоения размеров массива составляет  $\approx 70$  лет.

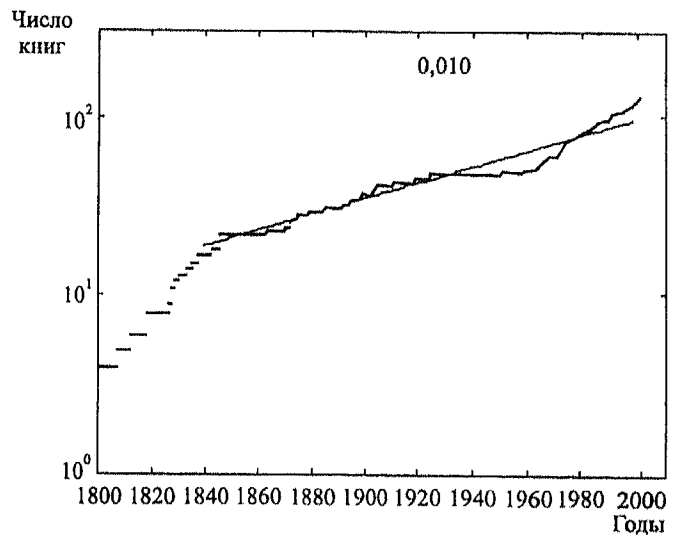


Рис. 4. Увеличение числа книг во времени (123 книги,  $\lambda_k = 0,01/\text{год}$ )

*Распределение статей по периодическим изданиям.* В библиографии содержится 877 статей, опубликованных в 276 журналах в 1732–2001 гг. Латентный период продолжался до 1820 г., за 1732–1820 гг. было опубликовано 11 статей.

Развитие научного направления по закону взаимности в теории чисел с 1820 г. отразилось в экспоненциальном увеличении числа статей (рис. 5):

$$N_{ст}(t) = N_{ст}(t_0 = 1820) \exp\{\lambda_{ст}(t - t_0)\}, \lambda_{ст} = 0,02/\text{год}.$$

Среднее время удвоения статей и журналов  $\approx 35$  лет.

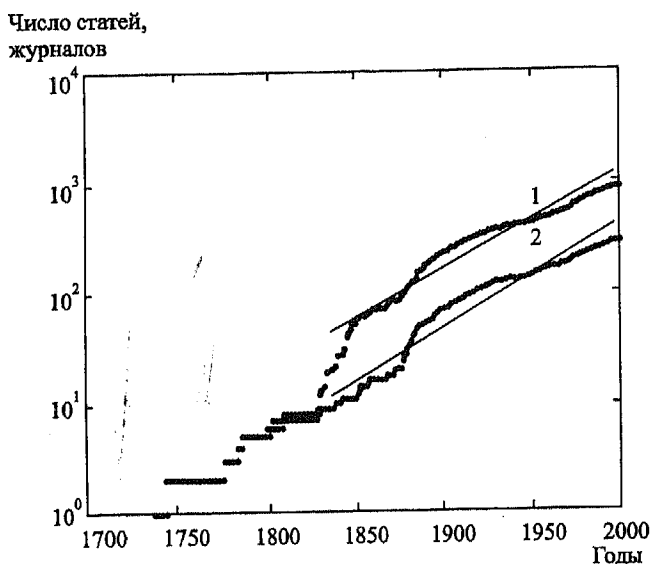


Рис. 5. Увеличение во времени числа статей и числа журналов из библиографии по закону взаимности (1736-1999 гг., 877 статей, 276 журналов,  $\lambda_{ст} = 0,02/\text{год}$ ,  $\lambda_{ж} = 0,02/\text{год}$ )

Ранговое распределение журналов по числу опубликованных статей в двойном логарифмическом масштабе показано на рис. 6 и имеет аналитическое выражение:

$$N_{ст}(r) = \frac{N_{ст}(r=1)}{r^\gamma},$$

где  $\gamma = 0,86$ , значение  $\gamma \approx 1,0$ , что соответствует классическому распределению Ципфа.

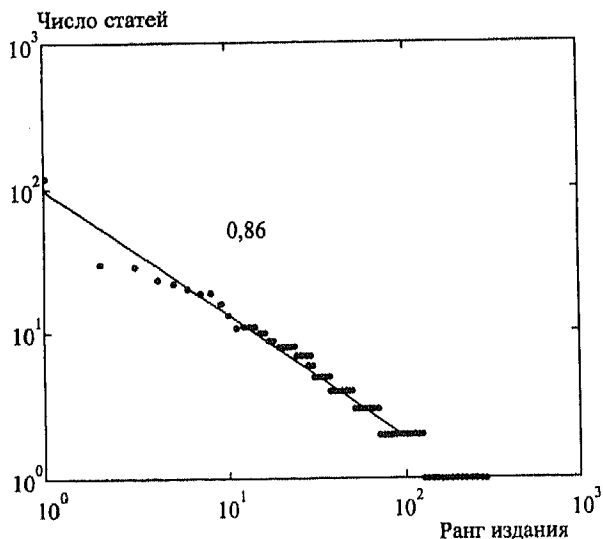


Рис. 6. Ранговое распределение журналов по числу статей с  $\gamma = 0,86$

*Распределение авторов по продуктивности.* Общее число авторов статей также увеличивается по экспоненте, и среднее время удвоения  $\approx 40$  лет.

Частотное распределение авторов по числу опубликованных статей представлено на рис. 7 и аппроксимируется формулой:

$$N_{авт}(i) = \frac{N_{авт}(i=1)}{i^{\alpha+1}}, \text{ где } \alpha = 0,88.$$

Это гиперболическое распределение с  $\alpha \approx 1,0$  соответствует эмпирическому закону Лотки.

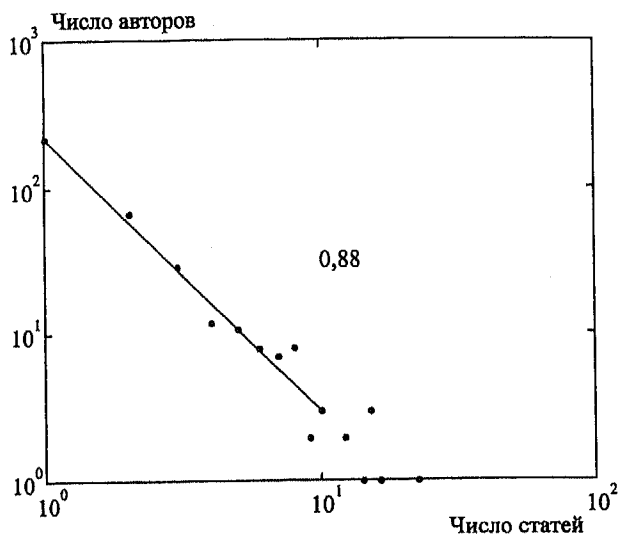


Рис. 7. Частотное распределение авторов по числу опубликованных статей с  $\alpha = 0,88$

Экспоненциальный рост числа статей продолжается до момента создания библиографии (2001 г.). Эта библиография подготовлена в период бурного развития данного научного направления и повышенного интереса научного сообщества к этой тематике новых поколений ученых.

*Динамика "незримого коллектива" ученых, работающих в одной тематической области.* Наибольший интерес для наукометрических исследований представляет не увеличение во времени общего числа ученых, а поведение "незримого коллектива" ученых, занимающихся научной проблемой и публикующих статьи в одно время.

На рис. 8 показано изменение "незримого коллектива" авторов, работающих в этом направлении в разные годы. Из рассмотрения были исключены авторы, опубликовавшие только по одной статье. Максимальный размер группы не превысил 30 человек за все время существования тематики.

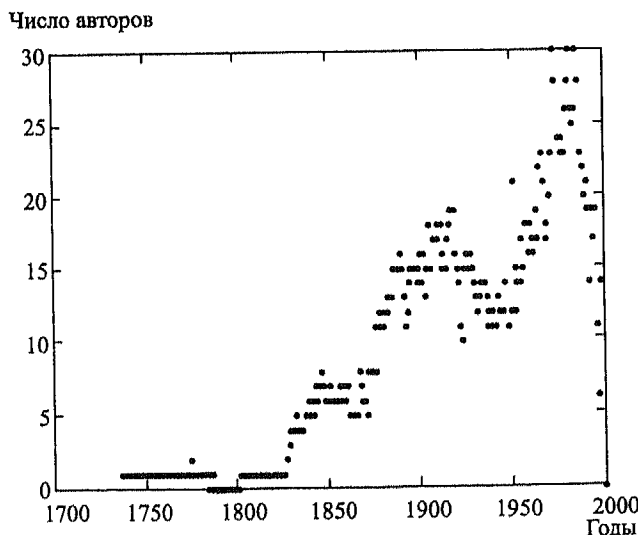


Рис. 8. Численность "незримого коллектива" авторов, печатающих статьи в различные моменты времени

Распределение длительности работы ученых в этой области составило для 377 авторов: до года работали 236 ученых, до двух лет — 17, трех — 10, четырех — 13, пяти — 7 и т. д.

Можно предложить показательный закон распределения длительности работы:

$N_a(t) = A \exp(-\mu_a t)$ ,  $\mu_a = 0,10/\text{год}$ , что соответствует среднему времени  $T = 1/\mu \approx 10$  годам.

Показательный закон распределения длительности работы ученых в тематическом направлении позволяет высказать предположение о том, что вероятность выхода ученого из тематической области (прекращение публикации статей) практически не зависит от того, сколько времени он проработал. Наиболее важный результат состоит в том, что число авторов, занимающихся тематикой, относительно невелико (до 30 человек) на протяжении двух столетий.

*Динамика "незримого коллектива" редакционных коллегий журналов, публикующих статьи по тематике.* Аналогичные зависимости получены для множества периодических изданий.

Распределение сроков публикации статей разными изданиями (интервал между первой и последней тематическим статьями) составляет для 276 изданий: один год — 136 журналов, два года — 8, три года — 5, четыре года — 4, пять лет — 4 и т. д.

Численность "незримых коллективов" редакционных коллегий журналов, публикующих статьи по тематике в различные моменты времени, была относительно невелика и не превышала 60 изданий (рис. 9).

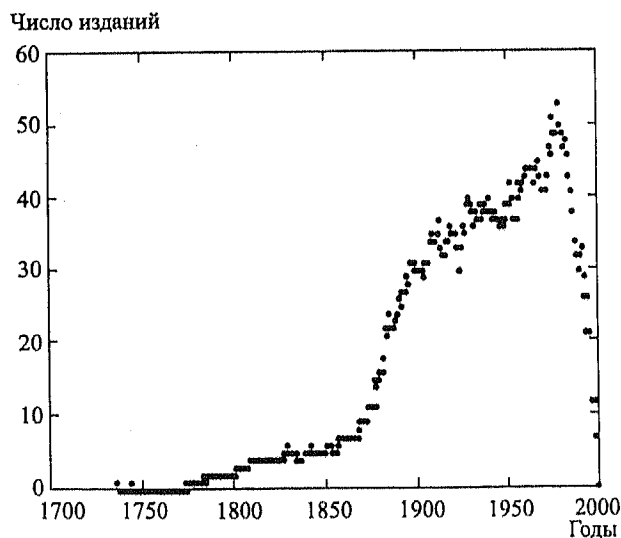


Рис. 9. Число изданий, печатающих статьи в различные моменты времени

Но средняя длительность участия в коммуникационных процессах в тематической области существенно различается: для авторов — 10 лет, для журналов — 30 лет.

*Флуктуации параметров гиперболических распределений.* На рис. 10 показаны изменения (флуктуации) значений показателей степени гиперболических  $\alpha$ -частотного и  $\gamma$ -рангового распределений для журналов за 1850–2000 гг. Размеры информационного массива возросли в десятки раз, а гиперболические распределения оставались практически без изменения.

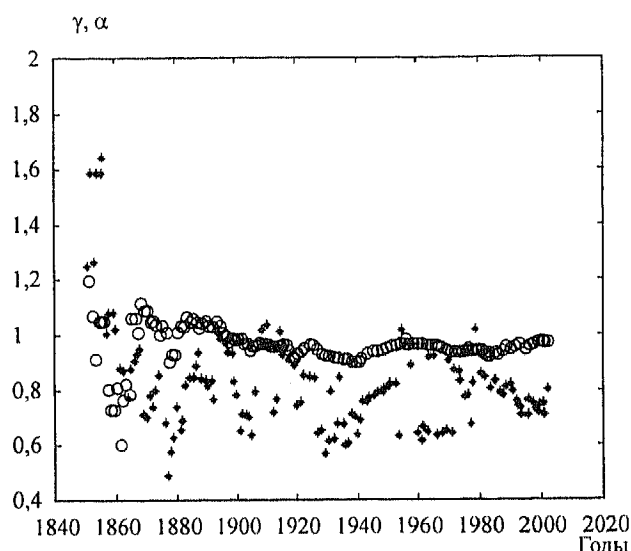


Рис. 10. Флуктуации показателей степени  $\gamma$ -рангового (o) и  $\alpha$ -частотного (\*) распределений для множества журналов

Флуктуации параметра  $\alpha$  несколько больше, чем параметра  $\gamma$ , этому есть вполне очевидное объяснение: "хвосты" распределения подвержены большим изменениям, чем изменения в "ядерной" группе изданий.

На рис. 11 показаны изменения (флуктуации) значений показателей степени гиперболических  $\alpha$ -частотного и  $\gamma$ -рангового распределений для множества авторов статей за 1850–2000 гг. Для множества авторов значение показателя  $\alpha$ -частотного распределения монотонно увеличивается, что объясняется расширением круга авторов, пришедших в тематику и опубликовавших пока небольшое число статей.

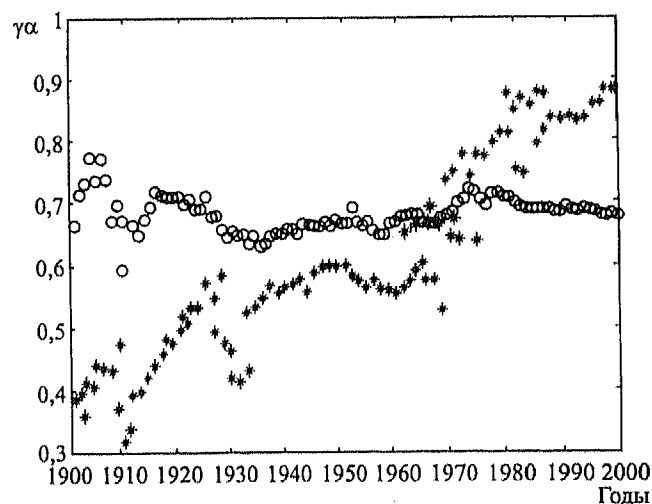


Рис. 11. Флуктуации показателей степени  $\gamma$ -рангового (o) и  $\alpha$ -частотного (\*) распределений для множества авторов

Анализ этой библиографии позволяет высказать предположение о выполнении закона рассеяния научной информации даже для узких тематических областей на протяжении длительных периодов.

Частной научной проблемой занимаются даже ученые, работающие в разных научных школах и странах, что находит отражение в рассеянии статей. Изученная нами библиография еще раз подтверждает мысль о едином процессе развития науки в обществе и происходящих коммуникационных процессах, охватывающих всё научное сообщество.

## ЗАКЛЮЧЕНИЕ

Рассмотренные закономерности формирования библиографий в мировой системе научной коммуникации позволяют сделать следующие выводы:

— к моменту создания библиографий документов наблюдается экспоненциальный рост статей во времени;

— рассеяние статей в системе периодических изданий является объективным процессом, связанным с объединением усилий научного сообщества для решения научных задач, и закон Бредфорда выполняется на протяжении всего времени развития тематического научного направления (рис. 10 и 11);

— всегда существуют небольшие группы ученых, объединенных одной научной проблемой, поддерживающих связи по формальным каналам коммуникации;

— сроки работы различных ученых различаются, причем больше половины из них публикует по одной статье и работает менее года;

— аналогичная ситуация имеет место и для редакционных коллегий научных журналов, более половины журналов публикует по одной статье и, соответственно, участвует в тематике менее одного года.

Это приводит к флуктуациям параметров гиперболических распределений для авторов и журналов и, в свою очередь, является причиной, по которой трудно выделить информационные потоки на начальной стадии образования новых тематических направлений. Такой анализ можно провести только на основе изучения полных библиографий, включающих все документы за большие промежутки времени.

Основные усилия в информетрии должны быть направлены на изучение поведения “незримых коллективов” ученых, публикующих регулярно статьи по новой тематике. Эти коллективы, развиваясь, объединяют и направляют усилия большого числа ученых и являются связующими звеньями в научном сообществе.

Анализ нескольких видов библиографий позволяет выдвинуть гипотезу о едином характере формирования информационных массивов документов.

Научные точки зрения ученых, собранные в библиографиях, позволяют новым ученым приступить к работе в этой области науки и находить новые нерешенные задачи.

Семантический подход к объединению научных документов в библиографиях позволил сформулировать основные положения для описания коммуникационных процессов в науке.

## ЛИТЕРАТУРА

1. Дин Ю. Научная коммуникация и библиометрия. Ч. 2. Процесс научной коммуникации // Международный форум по информации и документации.— 1999.— Т. 24, № 1.— С. 3–17.

2. Van Raan A. F. J. Fractal dimension of co-citation // *Nature*.— 1990.— Vol. 347.— P. 626.

3. Van Raan A. F. J. Fractal geometry of information space as represented by cocitation clustering // *Scientometrics*.— 1991.— Vol. 20, № 3.— P. 439–449.

4. Dowling M. Libraries, librarians and Library Association in the United States in 2001: Making a difference in the knowledge age // *IFLA Journal*.— 2001.— Vol. 27, № 3.— P. 133–142.

5. Шумский С. А. Самоорганизующиеся нейронные сети / Научная сессия МИФИ— 2001. III Всероссийская научно-техническая конференция “Нейроинформатика-2001”.— М.: МИФИ, 2001.— С. 182–211.

6. Tsay Ming-Yueh, Jou Shiow-Jen, Ma Sheau-Shin. A bibliometric study of semiconductor literature, 1978–1997 // *Scientometrics*.— 2000.— Vol. 49, № 3.— P. 491–509.

7. Lotka A. J. The frequency distribution of scientific productivity // *Journal of the Washington Academy of Science*.— 1926.— Vol. 16.— P. 317–323.

8. Bradford S. C. The documentary chaos / Bradford S. C. Documentation.— London: Crosby Lockwood, 1948.— P. 108–121.

9. Zipf G. K. Human behaviour and the principle of least effort.— Cambridge: Addison — Wesley, 1949.

10. Vlachy J. Frequency distributions of scientific performance. A bibliography of Lotka's law and related phenomena // *Scientometrics*.— 1978.— Vol. 1.— P. 109–130.

11. Chen Ye-Sho. Analysis of Lotka's law: The Simon — Yule approach // *Information Processing & Management*.— 1989.— Vol. 25.— P. 527–544.

12. Goffman W., Warren K. S. Dispersion of papers among journals based on a mathematical analysis of two diverse medical literatures // *Nature*.— 1969.— Vol. 221.— P. 1205–1207.

13. Lockett M. W. The Bradford distribution: A review of the literature, 1934–1987 // *Library and Information Science Research*.— 1989.— Vol. 11, № 1.— P. 21–36.

14. Hianik T., Passechnik V. I. Bilayer lipid membranes: Structure and mechanical properties.— Bratislava: Ister Science Ltd.— 1995.— 436 p.

15. Halmos P. R. Measure theory.— London: D. Van Nostrand Company, Inc., 1950.

16. Math WWW VL: Bibliographics [FSU Math].— “<http://euclid.math.fsu.edu/Science/Biblo.html>”

17. Bibliography on reciprocity laws.— “<http://www.rzuser.uni-heidelberg.de/~hb3/recbib.html>”

18. Beckmann P. E. Bibliography of chaos.— Johannes Gutenberg — Universitat Mainz, 2001.

<http://www.uni-mainz/FB/Physik/Chaos/Chaosbib.htm>