Высший химический колледж Российской Академии Наук

Избранные главы курса дискретной математики

Перечисление деревьев

С. В. Дужин

— Счел ли ты деревья?
— Как счесть деревья? — смеясь, сказал Степан Аркадьич,...— Сочесть пески, лучи планет хотя и мог бы ум высокий...

Л. Толстой. Анна Каренина

1 Формулировка результатов

Деревом называется связный граф, не имеющий циклов. На рисунке 1 изображены все деревья с числом вершин, не превосходящим 5. (Точный смысл этого утверждения заключается в том, что любое дерево, имеющее не более 5 вершин, *изоморфно* одному из деревьев, нарисованных на картинке, а никакие два из этих деревьев не изоморфны друг другу.)

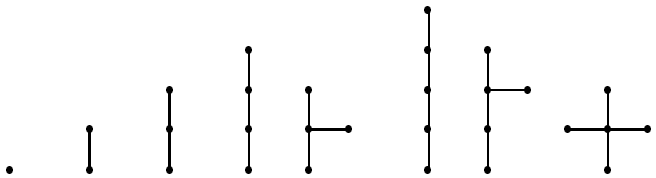

Рис. 1. Деревья с числом вершин ≤ 5

Задача 1. Докажите, что связный граф является деревом тогда и только тогда, когда число его ребер на единицу меньше числа вершин.

Задача перечисления деревьев, т.е. нахождения формул для числа t_n различных деревьев с n вершинами — одна из основных, и притом самых красивых, задач перечислительной комбинаторики. Впервые эту задачу поставил знаменитый английский математик Артур Кэли (см. [1]) в связи с задачей определения числа изомеров предельных углеводородов C_nH_{2n+2}. Каждый такой изомер описывается способом соединения атомов углерода, т.е. деревом с n вершинами валентности, не превосходящей 4. Надо сказать, что А. Кэли, найдя изящное рекуррентное соотношение для числа всех деревьев, не нашел сколько-нибудь удовлетворительной формулы для числа предельных алканов — такая формула была впервые получена венгерским математиком Д. Пойа в работе [2].

Задача 2. Попробуйте вывести эту формулу, дочитав этот текст до конца и используя изложенные здесь методы.

Все известные науке способы перечисления деревьев основываются на предварительном перечислении так называемых *корневых* деревьев.

Корневым деревом называется дерево, в котором отмечена одна вершина, называемая корнем.

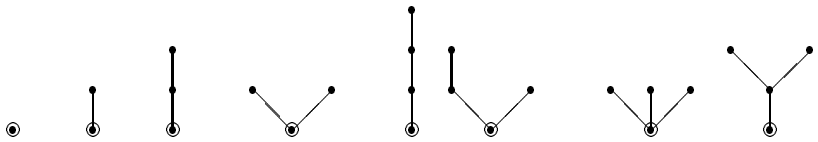

Рис. 2. Корневые деревья с числом вершин ≤ 4

Задача 3. Деревья и корневые деревья естественно возникают в различных, иногда весьма отдаленных друг от друга областях математики. Вот пример, недавно использованный Е. В. Бабаевым (МГУ) в его топологической теории структуры химических молекул. Проверьте, что различные, с точностью до гомеоморфизма, расположения n непересекающихся окружностей на сфере перечисляются деревьями с $n+1$ вершиной, а различные расположения n непересекающихся окружностей на плоскости перечисляются корневыми деревьями с $n+1$ вершиной. (Указание: области, на которые окружности делят поверхность, соответствуют вершинам дерева; при этом неограниченная часть плоскости отвечает корню.)

Корневые деревья считаются одинаковыми (изоморфными), если между ними существует изоморфизм, при котором корень переходит в корень. Сколько есть различных способов отметить корень в данном дереве? Столько, сколько у данного дерева неподобных (существенно различных) вершин, т.е. вершин, не переводимых друг в друга посредством автоморфизмов дерева как графа. Точное определение числа неподобных вершин — это число орбит в множестве вершин $V(G)$ данного дерева Γ при действии группы автоморфизмов $\mathrm{Aut}(\Gamma)$.

Пусть t_n — число неизоморфных деревьев с n вершинами, а T_n — число неизоморфных корневых деревьев с n вершинами. Значения t_n и T_n для малых n приведены в нижеследующей таблице:

n	1	2	3	4	5	6	7
t_n	1	1	1	2	3	6	11
T_n	1	1	2	4	9	20	48

Глядя на эту таблицу, трудно угадать какую-либо закономерность, управляющую поведением чисел t_n и T_n. Чтобы найти такую закономерность, оказывается чрезвычайно полезным изучать числа t_n и T_n не по отдельности, а объединенными в формальные степенные ряды ("производящие функции")

$$t(x) = \sum_{n=1}^{\infty} t_n x^n = x + x^2 + x^3 + 2x^4 + \dots$$

$$T(x) = \sum_{n=1}^{\infty} T_n x^n = x + x^2 + 2x^3 + 4x^4 + \dots$$

В терминах производящих функций рекуррентные соотношения между коэффициентами T_n и t_n допускают следующую компактную запись.

Теорема 1

$$T(x) = x \exp \sum_{k=1}^{\infty} \frac{T(x^k)}{k}$$

Теорема 2

$$t(x) = T(x) - \frac{1}{2}(T(x)^2 - T(x^2))$$

Задача 4. Считая известным лишь начальный отрезок разложения $T(x) = x + x^2 + 2x^2 + \ldots$, найдите по теореме 1 значение T_4, а затем по теореме 2 значения t_i для $i \leq 4$.

Для доказательства теоремы 1 мы воспользуемся теорией перечисления Пойа. Предварительно напомним ее основные положения.

2 Теория перечисления Пойа

Теория перечисления Пойа имеет целью нахождение числа сложных комбинаторных конфигураций в условиях действия группы подстановок. Она развивается в следующих предпосылках.

Имеется некоторое конечное множество M, элементы которого называются "позициями". На множестве позиций задано действие конечной группы G.

Позиции можно заполнять "фигурами", т.е. элементами некоторого, возможно бесконечного, множества Φ. Каждой фигуре из множества Φ приписан некоторый "вес", являющийся неотрицательным целым числом, и для каждого $n \geq 0$ известно число F_n фигур веса n. Производящая функция $F(x) = \sum_{n=0}^{\infty} F_n x^n$ называется "ряд, перечисляющий фигуры по их весам".

В некоторых ситуациях вместо позиций удобнее говорить об "объектах", а вместо фигур — о свойствах этих объектов.

Конфигурацией называется расстановка фигур по позициям, т.е. отображение $f : M \to \Phi$. *Вес* конфигурации — это сумма весов всех расставленных фигур:

$$w(f) = \sum_{\mu \in M} w(f(\mu)),$$

где через w обозначена функция веса.

Действие группы G на множестве M естественным образом порождает ее действие на множестве конфигураций (фигуры, расставленные по позициям, переставляются вместе с позициями). Конфигурации, эквивалентные относительно действия группы G, имеют равные веса. Пусть K_n — число конфигураций веса n, различных с точки зрения действия группы G, т.е. число G-орбит в множестве конфигураций, имеющих вес n.

Теория перечисления Пойа объясняет, как найти числа K_n, зная числа F_n.

Оказывается, что эту связь удобнее всего сформулировать в виде соотношения между производящими функциями $F(x)$ и $K(x)$, используя специальный многочлен Z_G от нескольких переменных, который определяется группой G и ее действием на множестве M и называется *цикловой индекс* группы G.

По определению,

$$Z_G(y_1, y_2, \ldots, y_m) = \frac{1}{|G|} \sum_{g \in G} y_1^{b_1(g)} y_2^{b_2(g)} \ldots y_m^{b_m(g)},$$

где $b_i(g)$ — число циклов длины i, входящих в разложение подстановки множества M, которое отвечает элементу группы g. (Таким образом, каждый моном $y_1^{b_1(g)} y_2^{b_2(g)} \ldots y_m^{b_m(g)}$ войдет в сумму с коэффициентом, равным числу элементов группы G, действие которых на множестве M имеет одно и то же разложение на циклы.) Число переменных, от которых зависит многочлен Z_G, определяется наибольшей длиной цикла среди встречающихся при рассматриваемом действии группы; во всяком случае, оно не превосходит числа элементов множества M.

Ну вот, мы дали все определения, необходимые для того, чтобы выписать, наконец, основную формулу теории перечисления Пойа. После таких титанических усилий формула выглядит очень коротко и просто (трудно в ученье — легко в бою, говорил Суворов).

Теорема 3 *(Формула перечисления Пойа).*

$$K(x) = Z_G(F(x), F(x^2), \ldots, F(x^m)).$$

Или словами: перечисляющий ряд для конфигураций получается при подстановке перечисляющего ряда для фигур в цикловой индекс группы.

Обратите внимание на способ подстановки функции $F(x)$ в многочлен Z_G: каждая переменная y_k заменяется на выражение $F(x^k)$ (не путать с $F(x)^k$, как мы однажды перепутали на лекции!).

Стандартным примером конфигурации является ожерелье из n бусинок двух цветов, скажем белых и красных, рассматриваемое с точностью до самосовмещений в пространстве. В этом случае множество фигур состоит из двух элементов (красная и белая бусинка), которым можно приписать веса 0 и 1 соответственно. Многочлен, перечисляющий фигуры, имеет вид $1 + x$. Множество позиций представляет собой набор вершин правильного n-угольника, на котором действует группа всех его движений D_n.

Задача 5. Чтобы набить руку, напишите многочлен, перечисляющий двуцветные ожерелья из 4 и 5 бусинок. Ответ проверьте при помощи здравого смысла.

Другой, особенно важный для химиков пример возникает в следующей ситуация. Множество позиций есть некоторое множество точек в пространстве образующее пространственную форму молекулы (например, вершины некоторого многогранника M). G — группа вращений этого многогранника. Фигуры

— это атомы или группы атомов, которые имеют одинаковую валентность и могут располагаться в вершинах многогранника M. В этом случае конфигурации — это изомеры замещения данного состава.

Единственная трудность, которая возникает в этой химической ситуации, заключается в том, что для полного перечисления молекул по количеству входящих в них атомов разного вида необходимо, вообще говоря, вводить векторные веса и производящие функции многих переменных. Например, если изучаются молекулы, состоящие из атомов трех типов, то перечисляющий многочлен для фигур будет $x+y+z$, а коэффициент при мономе $x^p y^q z^r$ в перечисляющем ряде для конфигураций есть число конфигураций, состоящих из p атомов первого типа, q атомов второго типа и r атомов третьего типа.

Задача 6. Напишите многочлен, который перечисляет изомеры, имеющие форму октаэдра и состоящие из шести трехвалентных атомов типа A, B и C.

3 Доказательства

Для доказательства теоремы 1 заметим, что при отбрасывании корня корневое дерево с $n+1$ вершиной распадается на несколько корневых деревьев с суммарным числом вершин n. Эти части однозначно, с точностью до перестановки, определены данным деревом. Обратно, из нескольких корневых деревьев всегда можно составить новое большое корневое дерево. При этом дерево с 1 вершиной получается, если составить вместе 0 деревьев.

Постараемся переформулировать это простое наблюдение как некоторое свойство производящей функции $T(x)$, используя теорию перечисления Пойа.

В качестве множества фигур у нас будет выступать множество всех корневых деревьев. Весом дерева мы будем считать число вершин в нем. Тогда перечисляющий ряд для фигур есть в точности производящая функция для числа корневых деревьев $T(x)$.

Рассмотрим составление нового корневого дерева со степенью корня r как образование конфигурации из r фигур. Множество позиций есть множество ребер, выходящих из корня большого дерева. Это множество состоит из r элементов, и на нем действует полная группа подстановок S_r. Вес дерева, составленного из компонент, рассматриваемого как конфигурация, есть по определению сумма весов составляющих его фигур, т.е. на 1 меньше числа вершин в нем. Таким образом, перечисляющий ряд для конфигураций по весам есть $\sum_{n=0}^{\infty} T_{n+1}^{(r)} x^n$, где $T_{n+1}^{(r)}$ есть число корневых деревьев с данной спененью корня r. По формуле Пойа

$$\sum_{n=0}^{\infty} T_{n+1}^{(r)} x^n = \frac{1}{r!} Z_{S_r}(T(x), T(x^2), \ldots).$$

Чтобы получить соотношение для числа всех корневых деревьев, надо просуммировать полученное равенство по всем r от 0 до бесконечности. (Значению 0 отвечает, как мы уже говорили, дерево с 1 вершиной; при этом группа S_0 есть единичная группа, действующая на пустом множестве, а ее цикловой ин-

декс есть константа 1 — в отличие от группы S_1, которая тоже единичная, но действует на одноэлементном множестве и имеет цикловой индекс, равный y_1.)

Суммирование дает:

$$\sum_{n=0}^{\infty} T_{n+1}x^n = \sum_{r=0}^{\infty} \frac{1}{r!} Z_{S_r}(T(x), T(x^2), ...).$$

Для суммы цикловых индексов всех симметрических групп существует формула, которая доказывается сравнением коэффициентов при одинаковых мономах в левой и правой части равенства.

Задача 7. Докажите формулу

$$\sum_{r=0}^{\infty} \frac{1}{r!} Z_{S_r}(y_1, y_2, ...) = \exp \sum_{k=1}^{\infty} \frac{y_k}{k}.$$

Теорема 1 непосредственно вытекает из утверждения задачи 7 и предыдущего соотношения.

Следствие 1 *Имеет место следующая рекуррентная формула для числа корневых деревьев:*

$$T_n = \frac{1}{n-1} \sum_{m=1}^{n} (\sum_{p|m} pT_p) T_{n-m}.$$

(Внутренняя сумма распространяется на все делители p числа m).

Доказательство следствия. Ряд $T(x)$ можно продифференцировать почленно:

$$T'(x) = \sum_{n=1}^{\infty} nT_n x^{n-1} = \sum_{n=0}^{\infty} (n+1)T_{n+1}x^n$$

Теперь возьмем производную от обеих частей уже доказанного равенства

$$T(x) = x \exp \sum_{k=1}^{\infty} \frac{T(x^k)}{k},$$

заменяя там, где надо, выражение $\exp \sum_{k=1}^{\infty} \frac{T(x^k)}{k}$ на $T(x)/x$. Получим:

$$xT'(x) - T(x) = T(x) \sum_{k=1}^{\infty} x^k T'(x^k).$$

Подставив сюда найденное выше выражение для T', находим:

$$\sum_{n=1}^{\infty} (n-1)T_n x^n = \sum_{q=1}^{\infty} T_q x^q \cdot \sum_{k=1}^{\infty} x^k (\sum_{p=1}^{\infty} pT_p x^{(p-1)k}).$$

Меняя здесь порядок суммирования, можно переписать это соотношение как

$$\sum_{n=1}^{\infty} (n-1)T_n x^n = \sum_{k,p,q} pT_p T_q x^{q+kp} = \sum_{n=1}^{\infty} \sum_{q+kp=n} pT_p T_q x^n.$$

(Последний знак суммы означает, для данного n, суммирование по всем тройкам целых чисел k, p, q, удовлетворяющим соотношению $q + kp = n$.) Наконец, приравняем коэффициенты при x^n в левой и правой части последнего равенства:

$$(n-1)T_n = \sum_{q+kp=n} pT_pT_q = \sum_{m=1}^{n} (\sum_{p|m} pT_p)T_{n-m},$$

где новый индекс суммирования m соответствует старому kp. Следствие доказано.

Задача 8. Запрограммируйте полученную формулу на компьютере и найдите числа T_n для $n \leq 15$.

Задача 9. Попробуйте найти прямое комбинаторное доказательство этой формулы, т.е. смысл разбиения числа $(n-1)T_n$ на слагаемые, указанные в правой части формулы.

Задача 10. При помощи тождественных преобразований со степенными рядами покажите, что теорема 1 эквивалентна следующей формуле А. Кэли:

$$T(x) = x \prod_{n=1}^{\infty} (1 - x^n)^{-T_n}.$$

Перейдем к доказательству теоремы 2, которая устанавливает связь между числом деревьев и корневых деревьев. Доказательство, которое будет дано, использует понятие *симметричного ребра* графа и понятие *реберно-корневого дерева*, и мы вначале их объясним.

Ребро e некоторого графа Г называется симметричным, если существует автоморфизм графа, меняющий местами концевые вершины e.

Задача 11. Докажите, что дерево не может иметь болеее одного симметричного ребра.

Таким образом, все деревья делятся на два типа: имеющие одно симметричное ребро и не имеющие ни одного симметричного ребра. (Укажите примеры тех и других).

Оказывается, такое разделение играет важную роль в подсчете числа существенно различных (неподобных) вершин и ребер дерева. (В п. 1 мы говорили, что связь между числом деревьев и корневых деревьев осуществляется путем подсчета числа неподобных вершин.)

А именно, имеет место следующая лемма.

Лемма (теорема о характеристике неподобия для деревьев). *Пусть p — число неподобных вершин, а q — число неподобных ребер некоторого дерева. Эти числа связаны между собой так:*

$p = q$, если дерево обладает симметричным ребром,

$p = q + 1$, если дерево не обладает симметричным ребром.

Доказательство леммы. По формуле Бернсайда для числа орбит при действии группы автоморфизмов графа $A = \mathrm{Aut}(\Gamma)$ на множестве вершин:

$$p = \frac{1}{|A|} \sum_{\alpha \in A} v(\alpha),$$

где $v(\alpha)$ - число вершин, неподвижных под действием автоморфизма α.

Аналогично, для действия A на множестве ребер:

$$q = \frac{1}{|A|} \sum_{\alpha \in A} e(\alpha),$$

где $e(\alpha)$ - число ребер, неподвижных под действием α.

Утверждение леммы вытекает из следующих наблюдений. (i) Если дерево не обладает симметричным ребром, то ребро может остаться на месте под действием автоморфизма, только если оба его конца остаются на месте. Следовательно, число ребер, которые сдвигаются, равно числу вершин, которые сдвигаются. А поскольку общее число вершин на 1 больше, мы получаем $v(\alpha) = e(\alpha) + 1$ для любого α. (ii) Если дерево имеет симметричное ребро, то группа автоморфизмов распадается на две равные части: автоморфизмы, сохраняющие оба конца симметричного ребра, и автоморфизмы, переставляющие эти концы. Для автоморфизмов первого типа будет $v(\alpha) = e(\alpha) + 1$, а для автоморфизмов второго типа, наоборот, $v(\alpha) = e(\alpha) - 1$ (на самом деле, как нетрудно проверить, в этом случае $v(\alpha) = 0$, $e(\alpha) = 1$). Следовательно, суммы в обеих формулах Бернсайда совпадут. Лемма доказана.

Упражнение. Проверьте утверждение леммы для деревьев, изображенных на рисунке 1.

Чтобы вывести теорему 2 из доказанной леммы, перепишем ее утверждение в виде тождества

$$1 = p - (q - s),$$

где, как и выше, p и q обозначают, соответственно, число неподобных вершин и ребер некоторого дерева, а s есть число симметричных ребер в этом дереве.

Теперь просуммируем обе части этого тождества по всем деревьям с данным числом вершин n. Что мы при этом получим? В левой части равенства получится, очевидно, просто число всех деревьев с n вершинами t_n. Первое слагаемое правой части даст в точности число всех корневых деревьев T_n.

А каков гносеологический смысл суммы чисел $q-s$, т.е. чисел *несимметричных* ребер, по всем деревьям с данным числом вершин? Чтобы его установить, нам как раз и потребуется заявленное раньше понятие реберно-корневого дерева. По аналогии с обычными корневыми (или вершинно-корневыми) деревьями, *реберно-корневое* дерево — это дерево, в котором отмечено одно ребро, называемое, естественно, корневым ребром. (Проверьте для разминки, что число реберно-корневых деревьев с 4 вершинами равно 3, что коренным образом отличается от числа вершинно-корневых деревьев с 4 вершинами.)

Теперь мы готовы к тому, чтобы объяснить смысл суммы $\sum(q-s)$, которая нас так заинтриговала. Это число L_n реберно-корневых деревьев с n вершинами, у которых запрещается брать в качестве корня симметричное ребро. Давайте их перечислим. Для этого подумаем: к чему приводит удаление корневого ребра? Правильно, к распадению дерева на две (ровно на две!) части с таким же общим числом вершин. Эти части, как нетрудно понять, естественным образом снабжены структурой корневого (именно вершинно-корневого!) дерева. Если при этом не трогать симметричных ребер (не брать их в качестве корней — согласно уговору, который дороже денег), то мы получим такое выражение числа L_n через числа T_k:

$$L_n = \sum_{\substack{k+l=n \\ k<l}} T_k T_l,$$

(суммирование только по парам неравных чисел!). Тот же факт можно выразить на языке производящих функций, для чего достаточно рассмотреть произведение ряда $T(x)$ сам на себя в количестве (только!) двух экземпляров. Немного поразмыслив, вы поймете, что формула

$$L(x) = \frac{1}{2}(T(x)^2 - T(x^2))$$

как раз и выражает необходимое соотношение (вычитание $T(x^2)$ — это запрет на одинаковые индексы k и l, т.е. запрет на симметричные корни; двойка в знаменателе — дань тому факту, что перестановка частей при составлении реберно-корневого дерева не меняет результат этого действия).

Задача 12. Используя доказанную теорему и результат задачи 8, найдите значения t_n для $n \leq 15$.

Литература

[1] A. Cayley. Mathematical Papers. Cambridge, 1891–1897: v.3, 242-246 (On the theory of the analytical forms called trees), v.9, 202-204 (On the math. theory of isomers), v.11, 365-367 (On the analytical forms called trees), v.13, 26-28 (A theorem on trees).

[2] G. Pólya. Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen, *Acta Math.*, **68** (1937), 145–254.

[3] Ф. Харари, Э. Палмер. Перечисление графов. М., Мир, 1977.

[4] В. В. Белов, Е. М. Воробьев, В. Е. Шаталов. Теория графов. М., Высшая школа, 1976.