Лекция 3

Графы

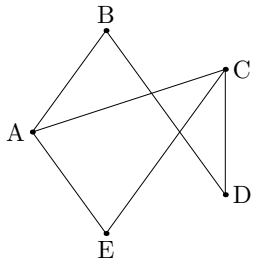
Чтобы решить какую-то задачу, часто бывает полезно нарисовать картинку, иллюстрирующую её условие. В этой главе мы рассмотрим один вид таких картинок: «графы». Граф — это набор точек («вершин»), соединённых линиями («рёбрами»). При этом важно, какие точки соединены, а как именно это ребро нарисовано, не имеет значения.

Прежде чем давать точные определения соответствующих понятий, мы разберём несколько задач, в которых подобные картинки помогают.

3.1 Примеры

3.1.1 Граф авиарейсов

Задача. Представим себе страну, в которой есть пять городов A, B, C, D, E, между которыми летают самолёты. Есть шесть рейсов: A–B, A–C, A–E, B–D, C–D, C–E (каждый рейс в обе стороны). Можно ли долететь из города A в город D прямым рейсом? с одной пересадкой? с двумя пересадками? Сколькими способами?



Это совсем простая задача: чтобы её решить, достаточно нарисовать картинку. Сразу видно, что прямого рейса нет, с одной пересадкой есть два способа A–B–D и A–C–D, а с двумя пересадками есть единственный вариант A–E–C–D.

Ту же картинку можно использовать, чтобы ответить на более сложный вопрос.

Задача. Докажите, что после отмены любого из шести рейсов оставшиеся пять позволяют добраться из любого города в любой («не теряется связность», как говорят).

Решение. Можно по очереди вычеркнуть каждый из рейсов и убедиться в требуемом. Если хотеть сказать короче, можно заметить, что можно составить два кольца («цикла») A–B–D–C–A и A–C–E. Всякий рейс входит в одно из колец; если он отменится, остальные рейсы кольца могут его заменить (пусть и с пересадками).

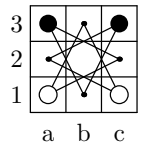
3.1.2 Перестановка коней

Следующая задача часто разбирается на математических кружках для школьников, так что, возможно, вы её уже видели.

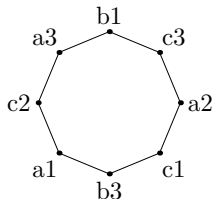
Задача. На доске 3×3 в углах a1, c1, a3, c3 поставлены белые и чёрные кони (белые снизу, чёрные сверху). Можно ли их поменять местами, чтобы чёрные оказались внизу, а белые наверху?

(По шахматным правилам конь ходит буквой «Г», на две клетки вперёд и на одну вбок, и не может стать на уже занятую клетку.)

Решение. Давайте отметим на доске, из каких клеток можно попасть в какие ходом коня. (Из центральной клетки никуда пойти нельзя, оставаясь на доске).



Картина станет яснее, если мы «распутаем» эти линии, составив схему возможных движений с точки зрения коня. Например, из a1 конь может пойти в b3 и c2, из b3 можно пойти обратно в a1 и дальше в c1, и так далее.



Теперь видно, что «топологически» (если не обращать внимание на географическое положение, а только на возможности ходов) кони могут двигаться по кругу, как в

хороводе. Чтобы поменять чёрных и белых местами, достаточно в этом круге сделать полоборота против часовой стрелки. Сначала каждый конь делает по одному шагу:

$$\mathrm{a1} \to \mathrm{b3},\ \mathrm{c1} \to \mathrm{a2},\ \mathrm{c3} \to \mathrm{b1},\ \mathrm{a3} \to \mathrm{c2}.$$

Контрольный вопрос. Нарисуйте (на шахматной доске) положения коней в этот момент.

Затем кони делают ещё шаг в том же направлении по кругу:

$$\mathrm{b3} \to \mathrm{c1},\ \mathrm{a2} \to \mathrm{c3},\ \mathrm{b1} \to \mathrm{a3},\ \mathrm{c2} \to \mathrm{a3}.$$

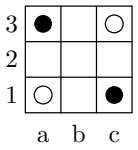
Контрольный вопрос. Убедитесь, что кони снова в углах шахматной доски, но цвета ещё не те.

Чтобы прийти к искомому положению, достаточно повторить такой сдвиг по кругу ещё два раза. Задача решена.

Задача. Сколько всего ходов сделали кони в нашем решении? Можно ли обойтись меньшим числом ходов?

Решение. На первый вопрос ответить легко: каждый конь сделал четыре хода, поэтому всего понадобилось 16 ходов. Второй вопрос немного сложнее. Заметим, что порядок коней по кругу не меняется при движении (как для машин на круговом шоссе, где нет возможности обгона). Значит, белый конь a1 должен попасть на c3, а белый конь c1 должен попасть на a3, а не наоборот, и каждый из них должен сделать по четыре хода (ему надо попасть в противоположную точку кольца). Аналогично для чёрных коней, и потому меньше 16 ходов не получится.

Задача. Можно ли поменять местами одного чёрного и одного белого коня, получив показанную на рисунке позицию? (Начальная позиция та же, что и в предыдущих задачах.)



Решение. Нет — в терминах кольца это означает, что один конь должен обогнать другого, а это невозможно (чёрные кони были рядом, и останутся рядом, аналогично и для белых).

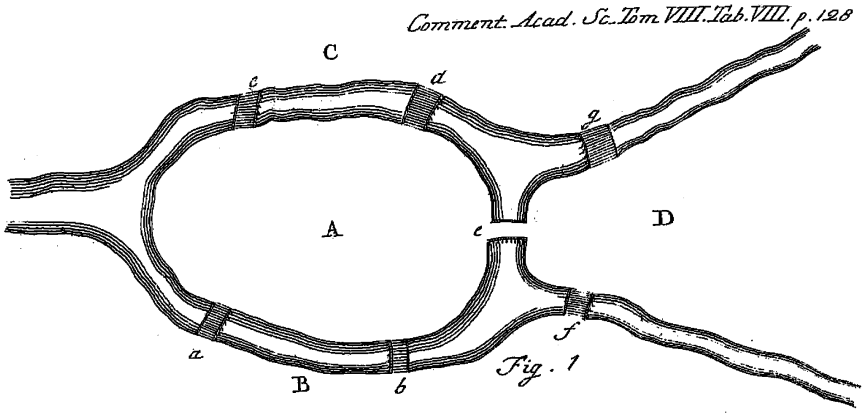
Задача для самостоятельного решения: про коней на доске 4×3.

3.1.3 Эйлер и мосты в Кёнигсберге

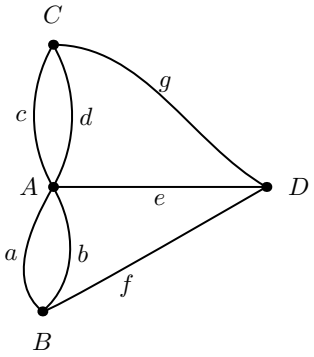
Следующая задача была разобрана в статье великого математика (работавшего в Петербурге) Леонарда Эйлера. Эту его статью считают первой работой по топологии. Эйлера спросили, можно ли совершить прогулку по Кёнигсбергу,[1] пройдя по

[1] Сейчас (2015) этот город называется Калининград и находится на территории Российской Федерации — но некоторые из эйлеровских мостов разрушены, и построены новые. В этом городе родился, жил, работал и умер философ Кант — в честь которого назван тамошний университет.

каждому мосту по одному разу и вернувшись в исходную точку. На рисунке (взятом из статьи Эйлера; её можно найти, например, в `https://math.dartmouth.edu/~euler/docs/originals/E053.pdf`) мосты обозначены буквами a, b, c, d, e, f, g.



Как решить задачу Эйлера? Заметим, что ходьба по суше никак не ограничивается — и поэтому каждый участок суши можно «стянуть в точку». На рисунке Эйлер отметил четыре таких участка A, B, C, D. Изобразим их точками и нарисуем (в виде линий) соединяющие их мосты, сохраняя обозначения Эйлера.



Теперь задачу Эйлера можно переформулировать так: можно ли пройти по каждой линии один раз и вернуться в исходную точку? Можно ли нарисовать эту картинку карандашом, не отрывая карандаша от бумаги, не проходя второй раз по уже нарисованным линиям и в конце вернувшись в исходную точку?

Глядя на картинку, легко понять, что это невозможно и в чём состоит препятствие. Посмотрим, например, на вершину D. В неё входят три линии. Обходя наш маршрут по циклу, мы должны пройти по каждой из этих трёх линий по одному разу. Но, войдя в точку по одной линии, мы должны выйти по другой, и снова войдя по третьей, мы уже не сможем выйти (если не хотим проходить по какому-то мосту

второй раз). Другими словами, необходимое условие разрешимости задачи обхода: в каждую точку входит чётное число линий (иначе они не поделятся на пары вход – выход). И в задаче Эйлера оно не выполнено.

Контрольный вопрос. Можно ли в этом рассуждении использовать вместо D другую точку? (Ответ: да, годится любая точка (в точках B и C тоже сходятся по три линии; в точке D их пять, но пять — тоже нечётное число.)

Задача. Можно ли пройти по каждому мосту по разу, если не требовать, чтобы маршрут был замкнут (начало и конец могут быть в разных местах)?

Решение. Нет: при этом нечётное число линий может быть в начале и конце пути, а во всех остальных вершинах должно быть по-прежнему чётное. А у нас во все четыре вершины входит нечётное число линий.

Задача для самостоятельного решения. Какое минимальное число мостов в задаче Эйлера нужно закрыть для прохода, чтобы для оставшихся мостов задача стала разрешимой?

Задача для самостоятельного решения. Как нарисовать распечатанный конверт, не отрывая карандаша от бумаги (левый рисунок)? Почему нельзя сделать того же для запечатанного конверта (правый рисунок)?

3.1.4 Рукопожатия

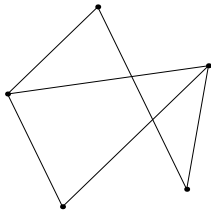
Задача. Перед началом встречи некоторые её участники пожали руки другим. Докажите, что количество участников, сделавших *нечётное* число рукопожатий, *чётно*.

(Рукопожатие симметрично, то есть учитывается для обоих участников.)

Решение. Вначале все участники сделали 0 рукопожатий, поэтому «нечётных участников» (тех, кто сделали нечётное число рукопожатий) в этот момент нет. Их количество равно нулю и потому чётно. (Нуль делится на 2 без остатка и является чётным числом.)

Что происходит при каждом рукопожатии? Количество «нечётных» либо уменьшается на 2 (два «нечётных» участника, пожав друг другу руки, становятся «чётными»), либо увеличивается на 2 (два «чётных» участника, пожав друг другу руки, становятся нечётными), либо остаётся прежним (если «чётный» участник жмёт руку «нечётному», то становится «нечётным», а другой — «чётным»). Задача решена.

Но можно решать задачу и иначе. Изобразим участников точками, а рукопожатия — линиями. (На плоскости линии могут пересекаться, но эти пересечения нас не интересуют).



Например, на нашем рисунке изображено 5 участников после 6 рукопожатий.

Контрольный вопрос. Кто из участников сделал больше всего рукопожатий? Сколько? (Ответ. Число рукопожатий, сделанных участником — это число линий, выходящих из изображающей его точки. Максимальное такое число равно 3, и таких наиболее общительных участников на рисунке двое.)

Подсчитаем для каждого участника число сделанных им рукопожатий. (На языке теории графов — подсчитаем для каждой вершины графа её степень, число выходящих из неё рёбер.) Все эти числа сложим. Что получится, как вы думаете?

Оказывается, *получится удвоенное число рукопожатий*. В самом деле, для каждой точки выпишем все линии, которые в неё входят. Число линий в этом общем списке будет равно интересующей нас сумме, а каждая линия будет учтена дважды (для двух её концов).

Почему наше наблюдение позволяет решить задачу? Почему, *если сумма нескольких целых чисел чётна, то число нечётных слагаемых нечётно*? Понятное дело: чётные слагаемые вообще на чётность суммы не влияют, а каждое нечётное слагаемое меняет чётность суммы, значит, их должно быть чётное число.

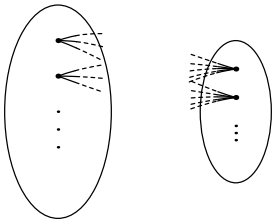
Задача для самостоятельного решения. Какое минимальное число дополнительных рукопожатий нужно сделать в описанной ситуации, чтобы все участники сделали по одинаковому числу рукопожатий?

3.1.5 Задачи и решения

Задача. На контрольной каждый из 20 школьников решил ровно 3 задачи, а каждую задачу решило ровно 5 человек. Сколько было задач?

Решение. Представим себе, что школьники писали решения задач на отдельных листках бумаги (каждое решение занимает ровно один листок). Тогда всего было $20 \times 3 = 60$ листков. С другой стороны, если разложить листки по задачам, то на каждую задачу придётся 5 листков, поэтому задач $60/5 = 12$.

Понятно ли, почему эта задача похожа не предыдущую? Нарисуем картинку, в которой школьники и задачи изображены точками, а решения изображены линиями (один конец — кто решил, другой — что решил). Чтобы не путаться, нарисуем школьников слева, а задачи справа. (Научно это называется «двудольный граф» со школьниками в левой доле и с задачами в правой).



На этом языке условие говорит, что в левой доле 20 точек («вершин»), и из каждой выходит по три линии («каждая вершина слева имеет степень 3»). А в правой доле неизвестно сколько точек, но в каждую приходит по 5 линий (от тех школьников, кто эту задачу решил). Сколько точек в правой доле? Решая задачу, мы подсчитали число линий двумя способами: считая их левые концы и их правые концы. Первый подсчёт даёт сумму степеней вершин слева, то есть $20 \times 3 = 60$, а второй даёт $5 \times$ (число задач). Поэтому задач $60/5 = 12$. Мы использовали такое соображение: сумма степеней левых вершин равна сумме степеней правых вершин (и равна числу рёбер).

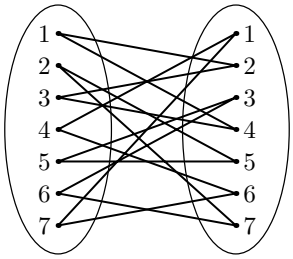
Есть ли какая-то польза от такой переформулировки? Вроде бы мы просто повторили то же самое рассуждение. Но в таком виде его может быть легче себе представить и легче применить похожие идеи в других задачах. Мы сейчас приведём такой пример — но задача эта более сложная, и строгое рассуждение в ней совсем не так просто придумать и понять (так что этот раздел можно пропустить, если не получается разобраться).

3.1.6 Разбор контрольной*

Задача. На контрольной каждый из N школьников решил ровно 2 задачи, а каждую задачу решило ровно 2 школьника. Докажите, что можно так огранизовать разбор задач, чтобы каждый школьник выступил по одному разу, рассказав одну из решённых им задач, и чтобы каждая задача была рассказана по одному разу.

Рассуждение из предыдущей задачи (всего $2N$ листков с решениями, по два для каждой задачи) показывает, что задач тоже было N. Видно, что по количеству всё сходится: задач столько же, сколько школьников. Но это ещё не значит, конечно, что можно организовать разбор.

Чтобы доказать, что это всегда возможно, нарисуем двудольный граф, о котором мы говорили. Слева и справа в нём по N вершин (слева школьники, справа задачи), и из каждой вершины (и слева, и справа) ведут по два ребра. Вот пример такого графа при $N = 7$.



Слева изображены школьники, справа — задачи. Например, школьник номер 1 решил задачи 2 и 4, а школьник номер 5 решил задачи 3 и 5.

Контрольный вопрос. Кто из школьников на рисунке решил задачу 4? Можно ли найти двух школьников, которые решили одинаковый набор задач? (Ответ: задачу 4 решили школьники 1 и 3. Школьники 1 и 3 решили одни и те же задачи 2 и 4; школьники 4 и 7 решили одни и те же задачи 1 и 7.)

Задача. Покажите, как для этого конкретного случая можно организовать разбор (каждый из семи школьников рассказывает одну из решённых им задач, и в итоге все задачи рассказаны).

Решение. Например, годится такой вариант: $1 \to 2$, $2 \to 5$, $3 \to 4$, $4 \to 6$, $5 \to 3$, $6 \to 7$, $7 \to 1$, где запись $i \to j$ означает, что школьник i рассказывает задачу j.

Задача для самостоятельного решения. Придумайте другой вариант разбора. Сколько таких вариантов существует?

Ну хорошо, для этого конкретного случая мы нашли искомое — но почему это возможно всегда? Сейчас мы это докажем — правда, не совсем строго. Представим себе рёбра графа в виде верёвок, соединяющих гвозди-вершины. В каждой вершине сходятся (по условию) две верёвки. Всего будет $2N$ верёвок, соединяющих N вершин слева и N вершин справа.) Давайте в каждой вершине свяжем концы верёвок друг с другом, отвязав от гвоздя и зацепив полученную петлю за тот же гвоздь. Свободных концов верёвок не остается, так что должны получиться (зацепленные друг за друга — это нам не важно) верёвочные кольца, надетые на гвозди.

Задача. Сколько колец какой длины получится на приведённом нами рисунке с 7 школьниками и 7 задачами?

Решение. Получатся кольца $1 \to 2 \leftarrow 3 \to 4 \leftarrow 1 \to \ldots$, $2 \to 5 \leftarrow 5 \to 3 \leftarrow 6 \to 7 \leftarrow 2 \to \ldots$, $4 \to 1 \leftarrow 7 \to 6 \leftarrow 4 \to \ldots$ из четырёх, шести и четырёх верёвок соответственно.

Задача. Почему каждое кольцо обязательно состоит из чётного числа верёвок?

Решение. В кольце чередуются вершины слева (школьники) и справа (задачи), поэтому нужно чётное число рёбер, чтобы вернуться в исходную долю.

То же рассуждение показывает, что в каждое кольцо входят поровну школьников и задач – они чередуются по схеме

$$\text{Ш}_1 - \text{З}_1 - \text{Ш}_2 - \text{З}_2 - \text{Ш}_k - \text{З}_k - \text{Ш}_1 \text{ (замыкая цикл)}$$

(Если в хороводе соседями мальчиков бывают только девочки и наоборот, то в нём поровну тех и других.)

Видно, что если в кольце оставить только каждую вторую связь, то получится как раз возможная схема разбора задач: школьник Ш$_1$ разбирает задачу З$_1$,..., школьник Ш$_k$ разбирает задачу З$_k$.

Контрольный вопрос. Какой второй вариант организации разбора для того же кольца? (Ответ: первую задачу разбирает второй школьник,..., $(k-1)$-ю задачу разбирает k-й школьник, k-ю задачу разбирает первый школьник.)

Объединяя такие разборы для всех колец, получаем схему разбора всех задач всеми школьниками. В самом деле, каждый школьник входит ровно в одно кольцо — в его вершине сходились две верёвки, которые и вошли в это кольцо. Аналогично для задач.

Задача для самостоятельного решения. Покажите, что число вариантов организации разбора всегда является степенью двойки. Какое максимальное число вариантов может быть, когда школьников и задач по N? (Ответ: $2^{\lceil N/2 \rceil}$.)

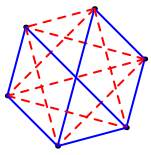
А что было бы, если бы каждый школьник решил ровно три задачи, и каждую задачу бы решило ровно три школьника? Оказывается, по-прежнему гарантирована возможность организовать разбор, но доказать это уже сильно сложнее. Один из способов состоит в использовании так называемой «теоремы Холла о представителях».

3.1.7 Знакомые и незнакомые

Задача. В компании из 6 человек некоторые знакомы друг с другом (это симметрично: если А знаком с Б, то и Б знаком с А). Докажите, что можно выбрать трёх человек, которые либо попарно знакомы (все три пары), либо попарно незнакомы (все три пары).

Как изобразить условие этой задачи? Соединим знакомых линией одного цвета (скажем, синей), а незнакомых линией другого (скажем, красной — на рисунке красные линии пунктирные). Получится граф из шести вершин, в котором каждая вершина соединена с каждой ребром. Такой граф называют «полным графом на шести вершинах». Его можно нарисовать в виде шестиугольника, в котором помимо сторон проведены все диагонали, и каждое ребро (рёбра = стороны и диагонали) либо красное, либо синее. (Пересечения диагоналей внутри шестиугольника не считаются вершинами, как обычно.)

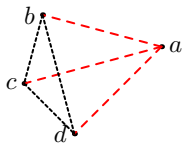
Цвета рёбер изображают ситуацию в задаче — а доказать надо, если говорить в этих терминах, что в таком графе обязательно найдётся либо целиком красный треугольник, либо целиком синий треугольник.



Контрольный вопрос. Найдите на рисунке одноцветный треугольник. Сколько их? (Ответ: синих треугольников нет, а красных треугольников треугольников два.)

Контрольный вопрос. Измените цвет одного ребра так, чтобы остался только один красный треугольник. Можно ли избежать при этом появление синего? (Ответ: можно перекрасить любое из шести рёбер двух красных треугольников, но тогда обязательно появится синий, иногда даже два

Осталось решить нашу задачу и доказать, что одноцветный треугольник обязательно найдётся. Как это сделать? Выберем произвольную вершину a. Она соединена с пятью другими пятью линиями. Эти линии двух цветов, поэтому есть как минимум три линии одного цвета (если линий каждого цвета было бы только две, то всего было бы 4, а не 5). Посмотрим на это подробнее: пусть из вершины a ведут три одинаковых (скажем, красных) ребра в какие-то три вершины b, c и d.



Понятно, как теперь завершить решение задачи? Достаточно посмотреть на цвета трёх рёбер в треугольнике b–c–d. Если среди них есть хотя бы одно красное, то оно образует красный треугольник с вершиной a. Если же все три ребра синие, так вот он и есть — синий треугольник. Утверждение доказано.

Что мы выиграли, говоря о красных и синих рёбрах вместо знакомых и незнакомых? Всего лишь одну (но важную) вещь: рассуждение теперь симметрично относительно замены цветов, что в первоначальной формулировке совсем не так явно видно (и потому рассуждение нужно было бы повторять дважды: когда у человека есть трое знакомых среди оставшихся и когда есть трое незнакомых).

Контрольный вопрос. Повторите рассуждение, избегая упоминаний рёбер и их цветов, а говоря только о знакомых и незнакомых.

Приведённое утверждение — лишь одно из серии утверждений, который называют утверждениями «рамсеевского типа» (в честь ??? Рамсея, который жил тогда-то). Среди них есть и сложные теоремы, и открытые проблемы, но мы приведём только несколько (не самых сложных) примеров.

Задача. Остаётся ли утверждение задачи верным, если в группе не шесть человек, а только пять?

Решение. Нет. Можно, например, нарисовать синий пятиугольник с красными диагоналями: ни сам пятиугольник, ни красная звезда внутри него не содержат треугольников.

Задача*. Докажите, что в группе из десяти человек можно выбрать либо четверых попарно знакомых (всего шесть знакомств между ними), либо трёх попарно незнакомых.

В терминах графов: в полном графе из десяти вершин с красными и синими рёбрами можно найти либо полный синий граф с четырьмя вершинами (=синий четырёхугольник с диагоналями), либо полный красный граф с тремя вершинами (красный треугольник).

Решение. Выберем одну вершину a и посмотрим на 9 рёбер, из неё выходящих. Там есть либо 6 синих, либо 4 красных, иначе всего было бы только $5 + 3 = 8$ вершин вместо 9. (Теперь у нас уже цвета не симметричны, поэтому и границы нужны разные, но схема рассуждения та же.)

Пусть есть 6 синих. Посмотрим (временно) только на них и соединяющие их рёбра. Мы уже знаем, что там можно найти либо синий, либо красный треугольник. И то, и другое нам годится: красный треугольник сам по себе хорош, а синий треугольник вместе с ведущими из a синими рёбрами даёт синий четырёхугольник с диагоналями.

Теперь второй случай: пусть есть 4 красных. Если эти четыре точки соединяет хотя бы одно красное ребро, то есть красный треугольник, если же нет — то есть синий чётырёхугольник с диагоналями. Задача решена.

Задача для самостоятельного решения*. Пусть $R(m, n)$ — наименьшее число людей в группе, которое гарантирует наличие m попарно знакомых или n попарно незнакомых. Эти числа определены при целых положительных m, n, при этом считаем $R(m, 1) = R(1, n) = 1$ (любой человек считается за одноэлементную группу попарно знакомых, а также попарно незнакомых). Рассмотренные нами задачи показывает, что $R(3, 3) = 6$ и что $R(4, 3) \leq 10$ (предыдущая задача).

Чему равно $R(2, 2)$? (Ответ: 2)

Чему равно $R(2, 3)$? (Ответ: 3)

Чему равно $R(2, m)$? (Ответ: m).

Докажите, что $R(m, n) \leq R(m - 1, n) + R(m, n - 1)$.

3.2 Неориентированные графы

Пора уже дать точное определения графа, не ограничиваясь примерами и неформальными разговорами о картинках. Это особенно важно, потому что в разных задачах приходится использовать графы разных видов (с разными определениями), и надо внимательно следить за тем, какое именно определение мы используем. Начнём с того, что называется неориентированными графами. (Более точное название было бы «неориентированные графы без петель и кратных рёбер», но мы будем опускать это уточнение.)

3.2.1 Определение

Чтобы задать *неориентированный граф*, нужно:

- Указать конечное множество V, элементы которого называются *вершинами* графа. (Другими словами, нужно перечислить вершины графа, при этом порядок перечисления не важен.)
- Для каждой пары различных вершин v и v' из V указать, соединены они ребром или нет.

Например, граф из раздела 3.1.1 имеет пять вершин A, B, C, D, E. Это записывают так: $V = \{A, B, C, D, E\}$ (порядок перечисления мог бы быть и другим). Рёбрами этого графа являются пары вершин (a, b), (a, c), (a, e), (b, d), (c, d), (c, e). Здесь речь идёт о *неупорядоченных* парах, то есть мы не различаем, скажем, (a, b) и (b, a). Обозначая через E множество рёбер, можно написать, что $E = \{(a, b), (a, c), (a, e), (b, d), (c, d), (c, e)\}$. И здесь порядок перечисления рёбер (и порядок указания концов рёбер) не имеет значения.

Традиция использовать букву V для вершин и E для рёбер происходит от английских слов «vertex» (вершина) и «edge» (ребро).

При компьютерном представлении графа с n вершинами можно как-то занумеровать его вершины числами от 0 до $n - 1$, а биты, кодирующие наличие рёбер, записать в таблицу $edge[i, j : 0..n - 1]$: `Boolean`. Если $edge[i, j]$ истинно, то в графе есть ребро между вершинами i и j. Эта таблица должна быть симметричной: $edge[i, j] = edge[j, i]$, поскольку тут речь идёт об одном и том же ребре.[2] Ещё можно заметить, что значения $edge[i, i]$ не соответствуют никаким рёбрам, так как мы не разрешаем «петли» — рёбра из вершины в саму себя. На практике бывает удобно считать, что эти места таблицы содержат значение `false`. Слова о том, что «в графе нет кратных рёбер» означают, что не может быть нескольких рёбер, соединяющих одну и ту же пару вершин.[3]

Задача для самостоятельного решения. При больших n имеет смысл хранить информацию о рёбрах графа более экономно. Предложите такой способ и оцените, во сколько раз тут выигрыш. (Указание. Если хранить только элементы таблицы над диагональю, скажем, в одномерном массиве типа `Boolean`, можно вдвое уменьшить расход памяти. Правда, чтение тогда будет более дорогой операцией.)

На математическом языке говорят, что *графом* называется пара (E, V), где E — некоторое конечное множество, элементы которого называются *вершинами* графа, а V — конечное множество, элементы которого называются *рёбрами* и являются неупорядоченными парами вершин (двухэлементными подмножествами множества E).

3.2.2 Соседи. Степени вершин

Вершины $v, v' \in V$ называются *соседями*, если в графе есть соединяющее их ребро (другими словами, если $edge[v, v'] =$ `true`). Напомним, что мы не разрешаем петли, поэтому сама вершина *не* является своим соседом.

Число соседей вершины (другими словами, число выходящих из неё рёбер) называется *степенью* этой вершины. В нашем компьютерном представлении, если принять `false` $= 0$ и `true` $= 1$, степень вершины v можно записать как $\sum_{w \in V} edge[v, w]$.

[2] Напомним, что в этом разделе мы рассматриваем неориентированные графы. Для ориентированных графов всё не так, см. ниже раздел 3.3.

[3] Можно было бы это разрешить, и тогда в таблице вместо булевых значений надо было бы хранить неотрицательные целые числа, «кратности» рёбер. Иногда это полезно, но мы такие графы рассматривать не будем.

Здесь существенно, что таблица симметрична и что диагональ $edge[v, v]$ мы считаем нулевой.

Теперь мы уже можем точно сформулировать утверждение о сумме степеней.

Теорема. В неориентированном графе сумма степень всех вершин равна удвоенному числу рёбер.

(Отсюда следует, в частности, что эта сумма чётна — и потому количество вершин нечётной степени нечётно, как мы видели.)

Доказательство. Будем складывать все элементы таблицы $edge[i, j]$. В строке i сумма элементов таблицы равна степени вершины i. (Напомним, что мы считаем $edge[i, i] = 0$.) Значит, сумма всех чисел в таблице равна сумме степеней вершин. С другой стороны, каждое ребро графа вносит вклад 2: ребро между i и j встречается как $edge[i, j]$ и $edge[j, i]$. Значит, сумма всех чисел в таблице равна удвоенному числу рёбер.

Контрольный вопрос. Чему равна сумма чисел в *столбце* таблицы *edge*? (Ответ: тоже степени соответствующей вершины, таблица симметрична.)

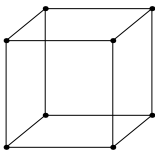
Пример. *Граф–путь* P_n имеет n вершин $v_1, \ldots v_n$. Рёбрами связаны пары вершин v_i и v_{i+1} $(1 \leqslant n - 1)$. Таким образом, в графе–пути $n - 1$ ребро. Говорят, что длина пути равна $n - 1$.

Контрольный вопрос. Какие степени вершин в графе P_n?

Пример. *Граф–цикл* C_n имеет n вершин $v_1, \ldots v_n$. Рёбрами связаны пары вершин v_i и v_{i+1} $(1 \leqslant n - 1)$ и пара вершин v_n и v_1. Таким образом, в графе–цикле n рёбер. Говорят, что длина цикла равна n.

Задача. Нарисуйте граф из 8 вершин, в котором степень каждой вершины равна 3.

Решение. Степень всех вершин одинакова, поэтому естественно стараться выбрать граф каким-нибудь симметричным. Можно сообразить, что годится каркас куба: вершины графа — это вершины куба, а рёбра графа — рёбра куба (отрезки, соединяющие соседние вершины).



На этом рисунке пересечения рёбер вне вершин куба, как всегда, не считаются. Но этих пересечений можно было бы и избежать.

Можно рассмотреть аналогичные графы («булевы кубы») больших размерностей.

Контрольный вопрос. Вершинами графа являются битовые строки длины n, а соседними считаются вершины, отличающиеся ровно в одной позиции (один из битов изменяется). Сколько вершин и рёбер в этом графе? Сколько соседей у каждой вершины? (Ответ: 2^n вершин, у каждой n соседей, всего рёбер $2^n n/2$.)

Контрольный вопрос. Рассмотрим *полный граф с n вершинами*, в котором любая пара вершин соединена ребром. Сколько в нём рёбер? (Ответ: степень каждой

вершины $n-1$, всего рёбер $n(n-1)/2$.)

Контрольный вопрос. Приведите пример графа с n вершинами и n рёбрами, в котором каждая вершина имеет степень 2. (Ответ: вершины и стороны n-угольника.)

Задача для самостоятельного решения*. Как нарисовать куб на плоскости, чтобы рёбра его не пересекались? Почему это можно сделать для любого выпуклого многогранника?

Задача для самостоятельного решения*. Нарисуйте граф из 12 вершин, каждая из которых имеет степень 5. (Указание. Какие вы знаете правильные многогранники?)

Задача для самостоятельного решения*. В предыдущих задачах, имея выпуклый многогранник, мы строили граф, в котором вершинами и рёбрами были как раз вершины и рёбра многогранника. Но можно построить и граф, в котором вершинам будут соответствовать *грани* многогранника, а рёбра — по-прежнему рёбра многогранника. Как?

Задача для самостоятельного решения*. Бывает ли выпуклый многогранник, в котором все его 15 граней — треугольники?

Задача. Нарисуйте граф из 9 вершин, в котором степень каждой вершины равна 3.

Решение. Тут есть подвох: такого графа не существует. В самом деле, по доказанному в нём должно было бы быть $3 \times 9/2 = 13{,}5$ рёбер, а их может быть лишь целое число.

На кружках для школьников эту задачу часто излагают так: в группе из 9 человек каждый человек имеет трёх друзей. Покажите, что либо кто-то учитывает самого себя в числе друзей, либо дружба не всегда взаимна.

Задача. При каких n существует граф из n вершин, в котором каждая вершина имеет степень 3?

Решение. Мы уже знаем, что при нечётных n это невозможно. Кроме того, это невозможно и при $n = 2$ (почему)? При чётных n, начиная с 4, такой граф можно построить. Достаточно расположить вершины по кругу в вершинах правильного многоугольника, соединив каждую вершину с двумя соседними по кругу, а также с диаметрально противоположной, получится три ребра.

Контрольный вопрос. Где мы использовали, что n чётно?

Задача. Докажите, что в любом графе с более чем одной вершиной есть две вершины одинаковой степени.

Решение. Пусть в графе N вершин. Их степени могут быть числами от 0 до $N-1$. Таких чисел как раз N. Значит, если степени всех вершин разные, то использованы все варианты от 0 до $N-1$. Но степени 0 и $N-1$ не могут встречаться одновременно: первое означает, что из какой-то вершины не выходит ни одного ребра, второе — что какая-то другая вершина соединена рёбрами во всеми остальными. Но тогда спрашивается, соединены ли эти две вершины ребром?

Контрольный вопрос. Где в этом решении используется, что $N > 1$? (Ответ: в противном случае получаются не две вершины, а одна.)

Задача. В любой момент футбольного турнира, проходящего в один круг — каж-

дая команда играет с каждой по одному разу — есть две команды, сыгравшие одинаковое число матчей.

Решение. В каждый момент турнира можно составить граф, вершинами которого являются команды, а рёбрами — уже сыгранные матчи. Степень вершины (команды) тогда равна числу сыгранных ей матчей, и можно применить предыдущую задачу.

Задача для самостоятельного решения. Где в этом рассуждение использовано, что турнир в один круг? Верно ли аналогичное утверждение для произвольного турнира?

Задача для самостоятельного решения. Докажите, что в выпуклом многограннике всегда найдутся две многоугольные грани с одинаковым числом сторон.

Задача для самостоятельного решения*. Как надо было бы определить граф с петлями и кратными рёбрами, и степени его вершин, чтобы сумма степеней вершин по-прежнему равнялась удвоенному числу рёбер?

3.2.3 Связные компоненты

Вспомним задачу об авиарейсах. В ней вершины графа — это города (точнее было бы говорить об аэропортах), а ребро между вершинами v и v' означает наличие авиасообщения между v и v' (двустороннего). В таком графе про любые два города возникает вопрос: можно ли добраться из одного в другой, пусть с пересадками. Все города (вершины) разбиваются на «связные компоненты»: если два города в одной связной компоненте, то добраться можно, а если в разных, то нельзя.[4]

Теперь хотелось бы определить понятие связности графа (и связной компоненты) более формально. Ничего сложного, неожиданного или красивого при этом не будет, но этим формальным языком надо уметь пользоваться.

Формальное определение

Будем называть *путём* в графе последовательность вершин $v_1, v_2, v_3, \ldots, v_k$, в которой стоящие рядом члены (вершины v_i и v_{i+1} при всех i) соединены ребром. Вершина v_1 называется *началом* пути, вершина v_k — его *концом*. *Длиной* пути будем называть число рёбер, то есть $k-1$. (Предупреждение: длина пути на единицу меньше числа вершин в нём!). Мы будем разрешать также и пути длины 0, то есть последовательности из одной вершины. У такого пути начало совпадает с концом. Рёбер в таком пути нет, но вершина (одна) есть.

Вершины v и v' называются *связанными*, если существует путь с началом в v и концом v'. Граф называется *связным*, если любые две его вершины связаны.

Контрольный вопрос. Связана ли вершина с самой собой? (Ответ: да, поскольку мы разрешаем пути длины 0.)

Бдительные читатели, наверно, уже заметили, что наша терминология (связанные вершины) симметрична: мы неявно подразумеваем, что *если есть путь с на-*

[4] Похожие вопросы возникают и для рельсового транспорта: например, в московской трамвайной сети приходится перевозить трамваи из одной связной компоненты в другую автотранспортом.

чалом v и концом v', то есть и путь с началом v' и концом v. Почему это действительно так? Достаточно «обратить» путь, то есть записать его вершины в обратном порядке. Обращение пути $v_1, \ldots, v_k$ даёт путь $v_k, \ldots, v_1$. Это по-прежнему путь (граф неориентированный), и начало и конец в нём поменялись местами.

Ещё одно важное свойство называется *транзитивностью* отношения связанности: *если вершина u связана с вершиной v, а вершина v связана с вершиной w, то вершина u связана с вершиной w.* Свойство это выглядит очевидным (если из u можно добраться в v, пусть даже с пересадками, и из v можно добраться в w, то из u можно добраться в w), но если хотеть изложить его доказательство более формально, это можно сделать так.

По предположению вершина u связана с v: есть путь $s_1, \ldots, s_k$, в котором $s_1 = u$, а $s_k = v$. Есть и путь $t_1, \ldots, t_l$, в котором $t_1 = v$, а $t_l = w$. Их можно соединить в один путь

$$s_1, s_2, \ldots, s_{k-1}, s_k = t_1, t_2, \ldots, t_{l-1}, t_l,$$

поскольку конец первого пути совпадает с началом второго. Ясно, что это снова путь (любая пара рядом стоящих членов была в одном из путей, так что соединена ребром по предположению). Начало этого пути $s_1 = u$, а конец $t_l = w$, так что вершины u и w связаны.

Теперь можно точно сформулировать обсуждавшееся свойство графов:

Теорема. Вершины неориентированного графа можно разбить на непересекающиеся группы, называемые *связными компонентами*, при этом:

- каждая вершина графа попадает ровно в одну группу;
- любые две вершины из одной группы связаны;
- любые две вершины из двух разных групп не связаны.

Такая компонента может быть одна — это значит, что любые две вершины связаны (и, как мы уже говорили, в этом случае граф называют связным).

Доказательство. Для каждой вершины v составим группу из всех связанных с ней вершин. Обозначим эту группу $C(v)$ — от слова «connected» (*англ.* связаны). Наблюдение: для любых двух вершин v и w группы $C(v)$ и $C(w)$ либо совпадают (содержат одни и те же вершины), либо не пересекаются. В самом деле, если v и w связаны, то по транзитивности всякая вершина, связанная с одной, связана и с другой, так что $C(v) = C(w)$. Если же v и w не связаны, то докажем, что $C(v)$ и $C(w)$ не пересекаются (не могут иметь общих вершин). В самом деле, если u было бы общей вершиной ($u \in C(v) \cap C(w)$), то v и u связаны (по определению $C(w)$), а также w и u связаны (по определению $C(w)$). Транзитивность гарантирует, что v связано с w, вопреки предположению.

Мы проверили, что получилось разбиение на непересекающиеся группы. Каждая вершина v попадает в группу $C(v)$. Вершины из одной группы связаны: если $v, w \in C(u)$, то v и w связаны с u, и по транзитивности v связано с w. (Мы повторяем уже использованное рассуждение). Вершины из разных групп не связаны: если вершина

$p \in C(u)$ связана с вершиной $q \in C(v)$, то получается цепочка: u связано с p, затем p связано с q, наконец, v связано с q. Дважды применяя транзитивность, заключаем, что u связано с v и потому $C(u)$ и $C(v)$ — не разные группы, а одна и та же. Теорема доказана.

Контрольный вопрос. К каким тройкам вершин мы применяли транзитивность в первый и второй раз? (Ответ: к u, p, q и к u, q, v.)

Подобные формальные рассуждения, конечно, удручают — мы долго и скучно доказываем нечто совершенно понятное (по крайней мере в ситуации с рейсами). Тем не менее это надо уметь — в более сложных случаях это позволит избежать ошибок.

Задача для самостоятельного решения*. Рассмотрим граф, вершинами которого являются трёхбуквенные слова русского языка, а рёбра соединяют два слова, которые отличаются в одной букве (получаются одно из другого заменой одной буквы). Находятся ли слова ТОК и СЫЧ в одной связной компоненте?

Нижняя оценка для числа связных компонент

Теперь мы всё-таки сформулируем и докажем какой-то менее очевидный (хотя и простой) результат.

Теорема. Если граф имеет V вершин и E рёбер, то в нём не меньше $V-E$ связных компонент. В частности, если граф связен, что $E \geq V - 1$.

(Часто буквой V и E обозначают не только множества вершин и рёбер, но и количества тех и других. Мы тоже так будем делать, если нет оснований опасаться путаницы.)

Контрольный вопрос. Что утверждает теорема при $E = 0$? (Ответ: что в графе без рёбер не меньше V связных компонент — а именно, каждая вершина образует свою компоненту.)

Контрольный вопрос. Почему второе утверждение следует из первого (и мы имеем право сказать «в частности»)? (Ответ: в связном графе одна компонента, и теорема даёт $V - E \leq 1$, то есть $V - 1 \leq E$.)

Доказательство. Коротко говоря, добавление одного ребра уменьшает число связных компонент, первоначально равное V, не более чем на 1 (уменьшение происходит, если ребро соединяет разные компоненты).

Более формально это доказательство можно оформить как индукцию по числу рёбер.

Базис индукции: в графе нет рёбер, и теорема утверждает очевидную вещь: в таком графе число связных компонент не меньше числа вершин (на самом деле тут равенство).

Шаг индукции. Пусть имеется граф G с V вершинами и $E > 0$ рёбрами. Временно удалим одно ребро. Получится граф G' с V вершинами и $E - 1$ рёбрами. В нём, по предположению индукции, не менее $V - E + 1$ компонент. Нам надо доказать, что возращение удалённого ребра уменьшает число компонент не более чем на 1. Пусть это удалённое ребро e соединяло вершины v и w.

Случай 1. В графе G' вершины v и w лежали в одной компоненте. Тогда в G будут те же компоненты, что и в G'. В самом деле, любой путь в G' будет путём в G, и остаётся проверить, что если вершины соединены в G, то они соединены и в G'. Соединяющий их путь может не использовать ребра e, и тогда он является путём в G. Если же он использует ребро e, то вместо этого ребра нужно вставить путь из v в w, который по предположению имелся в G'.

Случай 2. В графе G' вершины v и w лежали в разных компонентах. Теперь (в G) они соединены ребром, то есть попали в одну компоненту. Покажем, что все остальные компоненты, кроме этих двух, остались без изменений. В самом деле, если вершина z в графе G' лежала в какой-то третьей компоненте, то ни v, ни w из ней доступны не были — а значит, появление ребра v–w не увеличило множество доступных из z вершин (прежде чем в первый раз воспользоваться ребром v–w, нам надо дойти либо в v, либо в w, а это невозможно).

На этом доказательство шага индукции (и всей теоремы) завершается.

Применения нижней оценки*

Что даёт нижняя оценка на число связных компонент? Можно, конечно, сказать, что для связи между n городами нужно как минимум $n-1$ авиарейсов — это просто бытовая переформулировка доказанного. Но есть и более интересные применения. Сейчас мы рассмотрим два таких примера.

Пример 1. *Самый тяжёлый камень.* Есть n различных по весу камней. Эксперт, знающий веса камней, выбрал самый тяжёлый камень и хочет доказать суду, что он действительно самый тяжёлый. Для этого он в присутствии суда выполняет взвешивания на чашечных весах без гирь. На каждую чашку этих весов помещается по камню, и весы показывают, какой камень тяжелее. Какое минимальное число взвешиваний придётся сделать эксперту?

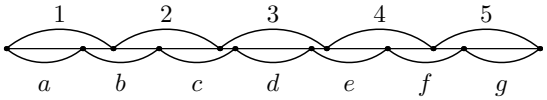
Ясное дело, что достаточно $n-1$ взвешиваний: эксперт может публично сравнить известный ему самый тяжёлый камень со всеми остальными — их как раз ровно $n-1$. Это не единственный возможный способ: например, эксперт может заранее упорядочить камни по весам $a_1 < a_2 < \ldots < a_n$ и продемонстрировать взвешивания $a_1 < a_2$, $a_2 < a_3$,…, $a_{n-1} < a_n$. После этого порядок камней станет ясен не только эксперту, но и суду. Второй способ требует тоже $n-1$ сравнений. Но нельзя ли обойтись меньшим числом сравнений?

Наша оценка на число связных компонент позволяет доказать, что нельзя. В самом деле, представим себе неориентированный граф, где вершины — это камни, а рёбра — это выполненные экспертом сравнения камней. Заметим, что мы даже не интересуемся тем, каковы были результаты сравнения, а просто фиксируем факт его проведения. Теперь ключевое соображение: если сравнений было меньше $n-1$, то этот граф не связен, в нём несколько связных компонент. В этом случае по результатам сравнений нельзя судить о том, какой камень самый тяжёлый. Понятно ли, почему?

В самом деле, поскольку взвешивания происходят только внутри связных компонент, то изменение весов всех камней внутри одной компоненты на одно и то же

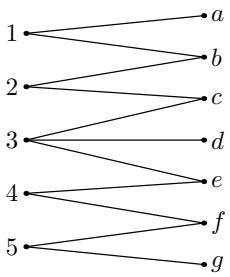
число (добавление константы) не изменит результатов взвешиваний ни в этой компоненте, ни (тем более) в других. Значит, суд увидит то же самое, что было бы и без изменения. А между тем такое изменение может увеличить вес камней в любой из компонент настолько, что самый тяжёлый камень окажется там — так что результаты показанных суду взвешиваний ничего не доказывают.

Пример 2. *Деление торта.* Допустим, что нам надо заранее разрезать торт на несколько кусков, готовясь к приходу гостей — так, чтобы его можно было раздать поровну m людям, а также чтобы его можно было (с теми же разрезами, только перегруппировав куски) раздать поровну n людям. Какое минимальное число кусков понадобится? Пусть, скажем, $m = 5$ и $n = 7$; сколько надо кусков, чтобы можно было раздать торт поровну и пяти, и семи людям? Наивный способ состоит в том, чтобы разделить торт на части по $1/35$ и потом группировать куски по 5 и по 7. Но легко понять, что это не оптимально (с точки зрения числа кусков) — представим себе торт отрезком и наметим точки разреза на 5 и на 7 равных частей. В первом случае будет 4 точки разреза, во втором 6, всего 10 точек разреза, то есть 11 кусков. Сделав все эти разрезы заранее, мы решим задачу, то есть 11 кусков достаточно. (В общем случае достаточно $m + n - 1$ кусков.)



Как доказать, что это оптимальный способ, то есть что 10 кусков недостаточно? Можно разбирать разные варианты и случаи и это установить, но есть такое рассуждение (которое легко обобщается на произвольные взаимно простые m и n).

Нарисуем схему раздачи в виде графа. Слева изобразим 5 гостей одного варианта, назовём их 1, 2, 3, 4, 5. Справа изобразим 7 гостей другого, назовём их a, b, c, d, e, f. Куски изобразим рёбрами, соединяющими тех, кому они попали при том и другом варианте. Получится граф, рёбра которого соотвествуют кускам: ребро соединяет тех людей (в первом и втором варианте), которым соответствующий кусок достанется.[5] Вот граф, изображающий ту же самую раздачу, что на предыдущем рисунке, но, разумеется, возможны и другие варианты раздач.



[5]Тут на самом деле есть проблема: вообще говоря, могут быть два куска, которые попадают одним и тем же людям в обоих вариантах, чего в графе быть не может. Но в этом случае эти два куска можно объединить в один, уменьшив число кусков.

Нам надо доказать, что в этом графе не менее 11 рёбер. Заметим, что в нём 12 вершин. Значит, если рёбер меньше 11, то граф будет несвязен (вот где нужна предыдущая теорема). Связная компонента состоит из некоторых вершин; разобьём и рёбра в соответствии с тем, в какую компоненту они попали (концы любого ребра попадают в одну компоненту, так как они связаны). Таким образом, куски разбиваются на несколько групп. Спросим себя, какой может быть суммарный вес кусков в одной из групп. Все эти куски в каждом из вариантов идут каким-то гостям, и больше эти гости ничего не получают (иначе эти новые куски вошли бы в ту же связную компоненту). Значит, общий вес кусков в группе должен быть кратен $1/5$ торта и $1/7$ торта одновременно, а это невозможно, кроме того случая, когда общий вес равен 1, то есть группа только одна. Получаем искомое противоречие.

Нахождение связных компонент: on-line алгоритм

В качестве отступления разберём красивый алгоритм нахождения связных компонент в графе. Представим себе граф с n вершинами, пронумерованными от 0 до $n-1$. Изначально в этом графе нет рёбер. Их постепенно добавляют, сообщая нам, какие две вершины i и j связывает очередное ребро. Помимо этого, нас время от времени спрашивают, связана ли такая-то вершина с такой-то (находятся ли они в одной компоненте на данный момент). В каком виде нам надо хранить информацию о текущем состоянии графа, чтобы обрабатывать запросы этих двух видов по возможности быстро? (Попытайтесь сами придумать какой-то способ, прежде чем читать дальнейшее.)

Другими словами, мы должны написать модуль, реализующий такие функции:

- *Initialize*(n : `integer`): создать граф с n вершинами без рёбер;
- *AddEdge*(i, j : `integer`): добавить ребро, соединяющее вершины i и j.
- *Connected*(i, j : `integer`) : `Boolean`: связаны ли вершины i и j в текущем графе (находятся ли они в одной связной компоненте).

Сначала реализуем совсем простую идею. Вспомним доказательство нижней оценки для числа компонент. Нам надо при каждом добавлении ребра узнавать, лежат ли концы его ребра в одной компоненте. Если лежат, то добавление этого ребра ничего не меняет; если нет, то надо слить две компоненты в одну.

Договоримся, что в каждый момент в каждой компоненте выбрана одна из вершин, которую мы будем называть *представителем* этой компоненты. В остальных вершинах компоненты хранится ссылка на этот самый представитель — или, по крайней мере, ссылка на вершину, где есть ссылка на этот представитель, или на вершину, где есть ссылка на вершину, где есть ссылка на него, и т.д. Более точно, мы заводим массив $ref[0..n-1]$ с таким свойством: если i — представитель какой-то компоненты, то $ref[i] = i$, а если нет, то цепочка

$$i, ref[i], ref[ref[i]], \ldots$$

рано или поздно стабилизируется на представителе связной компоненты, в которую попадает i.

Этот инвариант несложно поддерживать и использовать. При инициализации мы полагаем $ref[i] = i$: каждая вершина образует отдельную компоненту и является её представителем.

В любой момент можно найти представитель данной вершины, пройдя по цепочке:

Representative (i : `integer`) :

`while` $ref[i] \neq i$: $i \leftarrow ref[i]$

`return` $ref[i]$

Попадание в одну компоненту означает совпадение представителей:

Connected(i, j) : `return` (*Representative*$[i]$ = *Representative*$[j]$)

Соединение двух компонент в одну теперь можно реализовать так:

AddEdge(i, j) :

`if not` *Connected*(i, j) : $ref[$*Representative*$(i)] \leftarrow$ *Representative*(j)

Представителем объединённой компоненты становится представитель второй из объединяемых, а в цепочках для первой компоненты добавляется ещё один член. (Можно было бы сделать и наоборот, взяв представителя из первой компоненты.)

Этот алгоритм можно сильно ускорить с помощью двух оптимизаций. Первая состоит в том, что, проходя по цепочке, её можно оптимизировать, записав везде в *ref* соответствующего представителя. Это проще написать в рекурсивной форме:

Representative (i : `integer`) :

`if` $ref[i] = i$: `return` i;

`if` $ref[i] \neq i$: $rep \leftarrow$ *Representative*$(ref[i])$; $ref[i] \leftarrow rep$; `return` rep

Вторая оптимизация, про которую мы подробно говорить не будем, состоит в том, что мы стараемся использовать в качестве представителя объединения «более авторитетного» из двух представителей — того, у кого выше специальный параметр, называемый «рангом»; при таком присоединении ранг этого более авторитетного представителя не меняется (а ранг другого более не имеет значения). При равенстве рангов мы берём любого из двух, и его ранг увеличиваем на 1.

Подробнее про этот алгоритм и оценку времени его работы можно прочесть в главе 22 книги «Построение и анализ алгоритмов» (Кормен, Лейзерсон, Ривест, М.: МЦНМО, 2000), или в главе 5 книги «Алгоритмы» (Дасгупта, Пападимитриу, Вазирани, М.: МЦНМО, 2014). Можно также найти много информации в интернете по ключевым словам “disjoint-set data structure” или “Union–Find data structure”.

3.2.4 Расстояния. Простые пути.

Обычно, покупая билет на самолёт, мы при прочих равных стараемся выбрать маршрут с наименьшим числом пересадок. Аналогичным образом, имея вершины u и v в графе, можно искать путь минимальной длины с началом в u и концом в v. Длина этого пути называется *расстоянием* от u до v в графе. Будем обозначать его $d(u, v)$.

Это определение расстояния имеет смысл, если вообще есть путь из u в v, то есть если u и v связаны. Если же они лежат в разных компонентах, то удобно считать $d(u, v) = +\infty$.

Педантичные читатели заметили бы, что мы не обосновали, почему кратчайший путь существует (если вершины u и v связаны). Педантичные авторы ответили бы, что расстояния представляют собой натуральные числа, и любое множество натуральных чисел имеет наименьший элемент.

Контрольный вопрос. Чему равно расстояние от u до u? (Ответ: нулю — мы считаем, что пути, состоящие из единственной вершины, имеют нулевую длину.)

Контрольный вопрос. Для каких пар вершин u, v расстояние $d(u, v)$ равно 1? (Ответ: для вершин, соединённых ребром.)

Контрольный вопрос. Закончите предложение: «$d(u, v) \geq k$ означает, что в графе нет пути. . . ». (Ответ: «. . . из u в v, имеющего длину меньше k».)

Задача. Докажите, что $d(u, v) = d(v, u)$.

Решение. Обращая путь из u в v, получим путь той же длины из v в u. Поэтому $d(v, u) \leq d(u, v)$. Меняя обозначения, замечаем, что $d(u, v) \leq d(v, u)$, поэтому $d(u, v) = d(v, u)$.

Задача. Докажите *неравенство треугольника* для расстояний в графе:

$$d(u, w) \leq d(u, v) + d(v, u).$$

(Название объясняется тем, что для точек плоскости аналогичное неравенство говорит, что сторона uw треугольника uvw не превосходит суммы двух его других сторон uv и vw.)

Решение. Пусть $d(u, v) = k$ и $d(v, w) = l$. Тогда существует путь из u в v длины k и путь из v в w длины l. Соединяя их, получаем, что существует путь длины $k + l$ из u в w, так что $d(u, v) \leq k + l = d(u, v) + d(v, w)$.

Контрольный вопрос. Можно ли утверждать в последнем рассуждении, что $d(u, w) = k + l$? (Ответ: нет, построенный путь из u в w не обязан быть кратчайшим.)

Задача для самостоятельного решения*. Докажите, что для любых трёх вершин u, v, w выполнено неравенство $|d(u, v) - d(u, w)| \leq d(v, w)$.

Может быть, вы слышали от знакомых математиков про «число Эрдёша» — названное по имени венгерского математика, доказавшего много разных вещей, в том числе про графы, и имевшего много работ в соавторстве. Рассмотрим граф, вершинами которого являются авторы математических работ. Авторы u и v соединены в этом графе ребром, если у них есть совместная публикация (работа, где оба они являются авторами). Получается некоторый большой (но конечный) граф, и Эрдёш

является одной из его вершин. Расстояние до Эрдёша в этом графе (длина кратчайшего пути, то есть кратчайшей цепочки, в которой рядом стоят соавторы) называют «числом Эрдёша» автора.

Другой похожий пример — «теория шести рукопожатий». В ней говорится о графе, вершинами которого являются живущие сейчас люди, а ребро u – v означает, что u и v знакомы (= пожимали друг другу руки). Теория эта предполагает, что расстояние между любыми двумя вершинами в этом графе не превышает 6. Вряд ли она верна уж совсем буквально (наверно, бывают отшельники, ни с кем не знакомые?), но говорят, что это довольно близко к действительности.

Вообще для связного графа определяют его *диаметр* как наибольшее расстояние между какими-то его двумя вершинами. Теория шести рукопожатий утверждает, таким образом, что диаметр графа знакомств не превышает 6.

Все эти разговоры не выглядят серьёзно (и не претендуют на серьёзность), но статистический анализ больших графов, возникающих в жизни (скажем, графов знакомств в социальных сетях) — это популярная в наше время тема для статистических исследований и построения каких-то вероятностных моделей формирования такого рода графов.

Алгоритмы вычисления расстояний (и отыскания кратчайших путей) в графах — важный раздел алгоритмической теории графов; важно это и на практике, где обычно приписывают каждому ребру графа некоторую длину и длиной пути считают не число рёбер, а сумму длин этих рёбер. С такими алгоритмами сталкивались все, кто видели работу навигатора для автомобиля, определяюмщего кратчайший маршрут к цели. Можно только поразиться, до чего дошла наука о быстрых алгоритмах на графах — сравнительно слабые процессоры в смартфонах без особенного труда отыскивают кратчайшие пути на картах с огромным числом населённых пунктов и дорог. Но это отдельный разговор.

Понятие расстояния позволяет легко доказать следующее утверждение.

Теорема. Если вершины u и v в графе связаны, то существует соединяющий их путь, в котором все вершины различны (нет повторяющихся вершин).

Доказательство. Рассмотрим кратчайший путь из u в v. Если бы в него дважды входила некоторая вершина w, то участок между этими вхождениями можно было бы выбросить, и получился бы более короткий путь из u в v, вопреки предположению.

Пути без повторяющихся вершин называют иногда *простыми путями*. (К сожалению, терминология тут не вполне установилась.)

Контрольный вопрос. Приведите пример графа и простого пути в нём, не являющегося кратчайшим. (Ответ: возьмём граф в форме кольца и простой путь, занимающий больше половины этого кольца.)

Задача. Путь a_1–a_2–…–a_n является кратчайшим путём между a_1 и a_n. Докажите, что любой его участок a_k–a_{k+1}–…–a_l является кратчайшим путём между a_k и a_l.

Решение. Если этот участок — не кратчайший, то в исходном пути его можно заменить на более короткий, а это невозможно (исходный путь по предположению

был кратчайшим).

Задача для самостоятельного решения. Расстояние $d(u, v)$ между вершинами u и v в графе G равно 7. Докажите, что найдётся вершина w, для которой $d(u, w) = 4$ и $d(v, w) = 3$.

Задача. Имеется связный граф. Докажите, что в нём можно выбрать одну из вершин так, чтобы после её удаления вместе со всеми ведущими из неё рёбрами останется связный граф.

Решение. Выберем произвольную вершину u, назовём её «началом». Теперь выберем наиболее удалённую от начала вершину v (одну из таких, если их несколько). Докажем, что после удаления вершины v и её рёбер граф останется связным, а именно, что любая оставшаяся вершина w по прежнему связана с началом u. В самом деле, кратчайший путь из u в w в исходном графе не мог проходить через v, потому что в этом случае w было бы строго дальше от u, чем v, а v была одной из наиболее далёких от начала вершин.

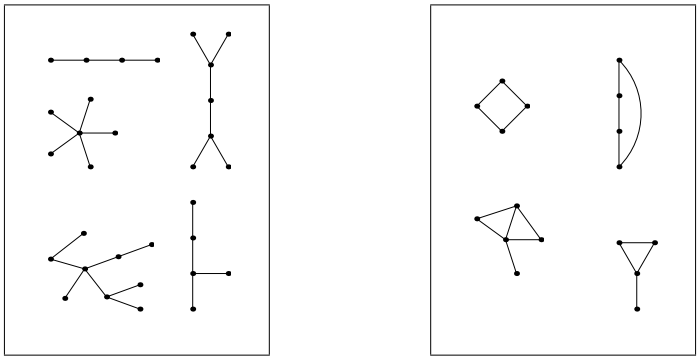
Задача. Вершинами булева куба являются битовые строки длины n, а соседями являются строки, отличающиеся в одной позиции. Чему равно расстояние между двумя произвольными вершинами? Чему равен диаметр этого графа? Сколько существует кратчайших путей между вершинами $00\ldots0$ и $11\ldots1$?

Решение. По определению расстояние — это минимальная длина пути по рёбрам, то есть минимальное число замен битов, позволяющих получить одну строку из другой. Оно равно числу различающихся битов и максимально для строки и её побитового отрицания, так что диаметр равне n. Кратчайший путь из $00\ldots0$ в $11\ldots1$ изменяет все биты в каком-то порядке (никакой бит не меняется дважды), и существует $n!$ таких порядков.

3.2.5 Деревья

Есть такая игра, называется «распознавание образов» или «машинное обучение» — человеку (или программе) дают два набора картинок, и требуется сформулировать правило, которое отличается объекты этих двух групп. Обычно при этом часть картинок при обучении не показывают, приберегая для «экзамена», где проверяется, насколько правило, сформулированное на основе учебных картинок, хорошо работает для экзаменационных.

Попробуйте сделать нечто подобное: не читая текста дальше, сформулируйте, чем отличаются графы левой и правой групп. (Все графы связны, так что мы не разделяем на рисунке их друг от друга — и так понятно, где кончается один и начинается другой.)



Картинки слева называются *деревьями*.

Контрольный вопрос. Расклассифицируем несколько первых букв алфавита на деревья (Г, Е, Ж) и не деревья (А, Б, В, Д).

ГЕЖ АБВД

Что будет с остальными буквами алфавита в этой классификации? (Буквы Ё, Й и Ы не связны, так что их мы не рассматриваем.) (Ответ: налево пойдут буквы З, И, К, Л, М, Н, П, С, Т, У, Х, Ц, Ч, Ш, Щ, Э, направо пойдут буквы О, Р, Ф, Ъ, Ь, Ю, Я.)

Наверно, вы уже придумали, как объяснить разницу между деревьями и недеревьями. Вот несколько вариантов описания свойства, отличающего графы слева от графов справа:

- связный граф, где нельзя удалить ни одного ребра без нарушения связности;
- связный граф, где число рёбер на единицу меньше числа вершин;
- связный граф, где для любых двух вершин u, v существует единственный простой путь из u в v;
- связный граф, где нет простых циклов длины больше 2.

В последнем пункте под *простым циклом* мы понимаем путь $a_1, a_2, \ldots, a_n, a_1$ (начало совпадает с концом), в котором все вершины $a_1, \ldots, a_n$ различны.

Теорема. Все четыре указанных свойства равносильны: граф, обладающий одним из них, обладает и всеми остальными.

Тем самым эти свойства определяют разными способами один и тот же класс графов; графы этого класса называются *деревьями*.

Доказательство. Для удобства пронумеруем эти свойства сверху вниз от (1) до (4) и начнём с очевидных наблюдений.

(2) ⇒ (1). Если в графе число рёбер на единицу меньше числа вершин, то после удаления ребра их будет на 2 меньше, а мы уже знаем, что в этом случае связность нарушится.

(1) ⇒ (2). Вспомним, как число связных компонент уменьшалось при добавлении рёбер. Если ни одно из рёбер нельзя удалить, значит, каждый раз новое ребро соединяло вершины из разных компонент. (Ребро, соединяющее вершины одной компоненты, можно всюду заменять путём, который соединял эти вершины до появления ребра — и удаление его не нарушит связность.) Поэтому к моменту, когда останется одна компонента, было добавлено как раз $V - 1$ рёбер, где V — число вершин.

Итак, (1) и (2) равносильны. Теперь докажем, что из этих свойств следует (3), то есть единственность простых путей. Здесь снова нужно будет смотреть за процессом добавления рёбер — мы знаем, что добавляемое ребро всегда соединяет две разные компоненты — и по индукции доказывать, что нет двух простых путей с одинаковыми началом и концом.

Итак, предположим, что это свойство выполнено для графа из нескольких связных компонент, и мы добавляем туда ребро p–q, соединяя две различные компоненты P и Q (содержащие p и q соответственно). Нам надо доказать, что и для нового графа это свойство выполнено. Пусть это не так, и есть два простых пути с одинаковым началом u и одинаковым концом v. Как минимум один из этих путей должен включать новое ребро p–q, поскольку до его добавления свойство единственности выполнялось (предположение индукции). Ребро это может входить только один раз, поскольку путь простой. Значит, начало пути лежит в одной из компонент P или Q, а конец в другой (остальные рёбра старые и не переводят в другую компоненту). Пусть, например, u лежит в P, а v лежит в Q. Тогда путь состоит из трёх частей: простой путь от u до p, ребро p–q, и простой путь от v до q. Посмотрим теперь на другой путь из u в v и убедимся, что он совпадает с первым. Он тоже должен использовать ребро p–q, поскольку до этого ребра вершины u и v лежали в разных компонентах. Тогда по тем же причинам он разбивается на часть от u до p, ребро p–q, и часть от q до v. Остаётся воспользоваться предположением индукции (единственность пути в старом графе от u до p, и от q до v) и увидеть, что два рассматриваемых пути совпадают.

Из (3) легко следует (4). В самом деле, если есть простой цикл a_1–a_2–…–a_n–a_1, где $n > 2$ и все $a_1, \ldots, a_n$ различны, то есть два простых пути из a_1 в a_n, а именно путь a_1–a_2–…–a_n и путь из одного ребра a_1–a_n.

Наконец, надо убедиться, что из (4) следует (1) и (2). Снова нужно вспомнить процесс слияния компонент при добавлении рёбер. Если (1) и (2) неверны, то в какой-то момент добавляется ребро, соединяющее уже связанные друг с другом вершины. Эти вершины связаны простым путём (как мы доказали), и вместе с добавленным ребром получится цикл. При этом длина цикла по крайней мере 3, потому что имевшийся простой путь был с пересадками (ребра-то раньше не было).

Теорема об эквивалентности четырёх свойств доказана.

Задача. Покажите, что из любого связного графа можно удалить часть рёбер

таким образом, чтобы оставшийся граф был деревом.

Решение. Здесь удобно пользоваться определением (2): если связный граф ещё не дерево, то есть ребро, которое можно удалить без нарушения связности. Если то, что останется — снова не дерево, удалим ещё одно ребро без нарушения связности, и так далее, пока не останется дерево.

Остовным деревом в графе называется дерево, которое получается удалением части рёбер. В предыдущей задаче утверждается, что в любом связном графе есть остовное дерево. Остовных деревьев может быть много. Например, в полном графе на n вершинах есть n^{n-2} остовных дерева. Мы не приводим доказательство этого факта.

Задача. Покажите, что в дереве из более чем одной вершины всегда есть вершина степени 1.

Решение. Вершин степени нуль в нём нет, иначе оно не будет связным. Если все вершины степени 2 или более, то сумма степеней не меньше удвоенного числа вершин, так что число рёбер (равное, как мы знаем, половине суммы степеней) не меньше числа вершин, а в дереве оно меньше (на единицу).

Задача для самостоятельного решения*. Покажите, что если граф является деревом, то его вершины можно так разместить на плоскости, чтобы соединяющие их рёбра (отрезки прямых) не пересекались. (Указание: можно использовать предыдущую задачу.)

Задача для самостоятельного решения* Докажите, что если в графе нет простых циклов длины 3 или больше, то можно добавить к нему рёбра так, чтобы получилось дерево.

Задача для самостоятельного решения*. Связность и деревья можно описать в терминах линейной алгебры. Для данного графа G с множеством вершин V рассмотрим векторное пространство ${\mathbb{F}_2}^V$ над полем $\mathbb{F}_2 = \{0, 1\}$ из двух элементов, состоящее из формальных комбинаций вершин с коэффициентами 0/1 (другими словами, его элементы — это функции $V \to \mathbb{F}_2$, или битовые строки длины $|V|$). Каждому ребру $e = (v, v')$ поставим в соответствие сумму его концов $v + v'$, то есть функцию, равную 1 на v и v' и 0 в остальных местах. Все рёбра лежат в подпространстве Z, образованном строками с нулевой (=чётной) суммой коэффициентов.

1. Покажите, что граф связен тогда и только тогда, когда векторы, соответствующие его рёбрам, порождают всё подпространство Z.

2. Покажите, что граф не имеет циклов тогда и только тогда, когда векторы, соответствующие рёбрам, линейно независимы.

(Указание. Если граф не связен, то рёбра лежат в меньшем подпространстве, соответствующем нулевой сумме во всех связных компонентах. Если вершины u и v связаны, то $u + v$ лежит в подпространстве, порождаемом рёбрами. Сумма рёбер, соответствующих простому циклу, равна нулю. Если цикла нет, то удаление ребра u–v разводит вершины u и v в разные связные компоненты, и для всех остальных рёбер сумма коэффициентов в каждой компоненте равна нулю, в отличие от выбранного ребра u–v, так что это ребро не выражается через остальные.)

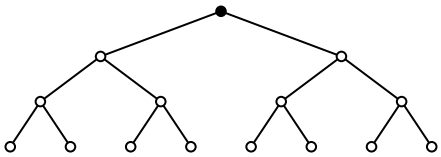
3.2.6 Полное бинарное берево

Рассмотрим граф, вершины которого являются двоичными словами длины $\leqslant n$, а рёбра имеют вид $\{u, u0\}$ или $\{u, u1\}$. Такой граф является деревом. Действительно, он связный: от любого слова можно перейти к любому, стирая последний символ или дописывая символ к концу слова.

Посчитаем количество ребер. Для этого удобно ввести на рёбрах ориентацию и считать, что более короткий конец ребра является его началом, а более длинный — его концом. Начал рёбер столько же, сколько всего рёбер. Это количество равно удвоенному числу двоичных слов длины меньше n (каждое такое слово является началом двух рёбер), т.е.

$$2 \cdot (1 + 2 + \cdots + 2^{n-1}) = 2^{n+1} - 2.$$

Количество вершин в таком графе равно количеству слов длины не больше n, т.е. $2^{n+1} - 1$. Таким образом, этот граф является деревом. Его полное название: *полное бинарное дерево глубины* n. На рисунке изображено полное бинарное дерево глубины 3. (Подумайте, каким вершинам какие слова соответствуют.)



На это рисунке выделена одна из вершин. Это пустое слово.

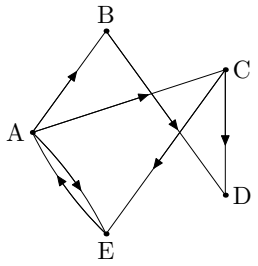
Если в дереве выделена особая вершина, она обычно называется *корнем дерева*. В этом случае вершины степени 1 (*висячие вершины*) называют *листьями дерева*.

Количество листьев в полном бинарном дереве глубины n равно 2^n — это количество двоичных слов длины n.

3.3 Ориентированные графы

3.3.1 Определение

Часто полезно рассматривать графы, в которых рёбра имеют ориентацию (тогда их изображают линиями со стрелками, а не просто линиями). Например, если в разделе 3.1.1 допустить односторонние рейсы (самолёт возвращается без пассажиров или летит в какое-то другое место), то такой односторонний рейс будет изображаться линией со стрелкой — а двусторонний рейс можно будет изобразить двумя линиями со стрелками туда и сюда. Изучая, можно ли теперь из города X попасть в город Y, мы должны будем искать путь из X в Y по стрелкам. Это отношение достижимости уже не обязано быть симметричным: можно случиться так, что из X в Y проехать можно, а обратно — нет. Скажем, в ориентированном графе на рисунке видно, что из A можно попасть в D (с пересадками), а обратно проехать нельзя никак.



Формально говоря, мы рассматриваем *ориентированные графы*,[6] в которых для каждой пары различных вершин v, v' известно, есть ли ребро из v в v'. Про такое ребро говорят, что v является его *началом*, а v' — его *концом*. При этом возможно, что ребро из v в v' есть, а ребра из v' в v нет (скажем, на рисунке есть ребро AB, но нет ребра BA). Может быть и так, что рёбра есть в обе стороны (скажем, на рисунке есть ребро AE и есть ребро EA) — или что ни в одну сторону нет рёбер (скажем, на рисунке нет ни ребра AD, ни ребра DA). Но «кратные рёбра» (несколько рёбер с общими началом и концом) и «петли» (рёбра, ведущие из вершины в неё же) по-прежнему мы не разрешаем.

Занумеровав вершины ориентированного графа числами от 0 до $n-1$, можно закодировать информацию о её рёбрах в виде таблицы $edge[i,j]$, в которой i, j принимают значения в интервале от 0 до $n-1$, и $edge[i,j]$ истинно, когда есть ребро из i в j (и ложно в противном случае). Теперь уже не требуется, чтобы таблица была симметрична. Как и раньше, значения $edge[i,i]$ не имеют смысла, и удобно договориться, что в таблице $edge[i,i] =$ `false`.

На математическом языке говорят, что *ориентированным графом* называется пара (E,V), где E — некоторое конечное множество, элементы которого называются *вершинами* графа, а V — конечное множество, элементы которого называются *рёбрами* и являются упорядоченными парами вершин (другими словами, $V \subset E \times E$).

Неориентированные графы (те, которые мы рассматривали раньше) можно отождествить с ориентированными графами, в которых для каждого ребра существует и обратное.

3.3.2 Степени вершин

В неориентированных графах мы определяли степень как число рёбер, выходящих из этой вершины. Теперь надо различать выходящие и входящие рёбра, поэтому вместо одного числа получаются два. Их называют *исходящей степенью* и *входящей степенью*.

Раньше у нас была теорема о том, что сумма степеней вершин равна удвоенному числу рёбер. Как её надо переделать для случая ориентированных графов? Наверное, вы уже догадались, как:

[6] Иногда для краткости говорят даже «орграфы» (по-английски ‘directed’ graphs сокращают до ‘digraphs’).

Теорема. Сумма исходящих степеней всех вершин равна сумме входящих степеней всех вершин: обе суммы равны числу рёбер графа.

Доказательство. Каждое ребро имеет одно начало (выходит из какой-то вершины), и поэтому учитывается по разу, когда мы складываем исходящие степени всех вершин. Аналогично для концов рёбер.

3.3.3 Пути и достижимость.

Определяя пути в ориентированных графах, мы должны учитывать направления рёбер (идти только по стрелкам на рисунках, а не против). *Путём* в ориентированном графе будет последовательность вершин $v_1, v_2, \ldots, v_k$, в которой соседние члены соединены ребром в нужную сторону, то есть в графе есть ребро с началом v_i и концом v_{i+1} (при всех $i = 1, 2, \ldots, k-1$). Длиной пути по-прежнему считают число рёбер, то есть $k-1$. Путь из одной вершины имеет длину 0. Вершину v_1 называют началом пути, а вершину v_k — его концом.

В отличие от неориентированных графов, пути обращать нельзя (точнее, не всегда можно обращать), и вместо определения связанных вершин мы должны дать другое. Говорят, что вершина v *достижима* из вершины u, если существует путь с началом u и концом v. В частности, каждая вершина достижима сама из себя, так как мы разрешаем пути длины 0. Это отношение уже не обязано быть симметрично: возможно, что вершина v достижима из вершины u, но не наоборот. Но оно по-прежнему транзитивно: если вершина v достижима из u, а w достижима из v, то w достижима из u (соединяем пути).

Как и раньше, *простой путь* в ориентированном графе — это путь, не проходящий дважды через одну вершину. Достижимость, как и раньше, можно определять с помощью простых путей: если v достижима из u, то есть простой путь из u в v. В самом деле, если какая-то вершина встречается дважды, то участок пути между её вхождениями можно вырезать, получив более короткий путь, так что путь минимальной длины всегда будет простым (а он существует).

В неориентированном графе мы определяли расстояние между связанными вершинами как длину кратчайшего пути. В ориентированном графе, если вершина v достижима из u, можно тоже рассмотреть длину кратчайшего пути из u в v. Но называть это «расстоянием» не стоит: обычно от расстояния требуют симметричности, а здесь она очевидным образом нарушается. (Если по кольцевой дороге автобус ходит в одну сторону, и мы проехали свою остановку, то нам придётся почти что объехать кольцо, чтобы вернуться.)

Контрольный вопрос: выполнено ли для этого «несимметричного расстояния» $d(u, v)$ свойство транзитивности

$$d(u, w) \leq d(u, v) + d(v, w)?$$

3.3.4 Достижимость и разрезы

Сейчас мы на примере достижимости попытаемся проиллюстрировать важную идею, которую называют *двойственностью*. Представьте себе, что вам на экзамене дали

граф, указали две вершины s и t и просят доказать, что вершина t достижима из вершины t. Что вы напишете в качестве доказательства? Согласно определению, для доказательства достаточно предъявить путь из s в t. Этот путь можно выбрать простым, и тогда его длина меньше числа вершин, так что он не такой уж и длинный (по сравнению с размером графа).

Теперь более сложный вопрос: пусть вам надо доказать, что вершина t *недостижима* из вершины s, то есть что такого пути нет. Что же тогда можно предъявить в качестве доказательства?[7] Можно, конечно, перечислить все пути из вершины s, длина которых меньше числа вершин графа, и убедиться, что ни один из них не кончается в t. В этом случае, как мы знаем, и более длинных путей из s в t нет. Но путей даже и ограниченной длины вообще-то очень много (как говорят, их число экспоненциально растёт с ростом графа). Нельзя ли предъявить какое-то более короткое доказательство?

Таким доказательством может служить то, что называют *разрезом* графа. Представим себе, что все вершины графа разбиты на две категории (каждая вершина отнесена ровно к одной из двух), названные S и T. При этом вершина s отнесена к категории S, а вершина t отнесена к категории T. Такое деление называют *разрезом* графа (можно представить себе, что бумажку, где граф был нарисован, аккуратно разрезали на две части). Если при этом оказалось, что *в графе нет ни одного ребра, ведущего из S в T*, то ясно, что t недостижима из s. (Чтобы из страны можно было улететь, нужно иметь хотя бы один рейс изнутри наружу.) Заметим, что указать разрез не так сложно (у каждой вершины надо написать, в S она или в T) — и после этого можно убедиться, проверив все рёбра по очереди, что ни одно из них не ведёт из S в T. Так что никаких объектов экспоненциального размера в таком доказательстве недостижимости не появляется. Следующая «теорема двойственности» говорит, что это универсальный способ доказательства.

Теорема. Если в ориентированном графе вершина t недостижима из s, то существует разрез (S, T), это устанавливающий ($s \in S$, $t \in T$, и из S в T не ведёт ни одного ребра).

Доказательство. Пусть S — множество всех вершин, достижимых из s (соответственно, T — множество всех вершин, недостижимых из s). Тогда $s \in S$ по определению, $t \in T$ по предположению, и остаётся проверить, что нет рёбер из достижимых вершин (S) в недостижимые (T). Но это очевидно: если из достижимой (из s) вершины куда-то ведёт ребро, то и конец этого ребра достижим (из s), надо просто добавить это ребро в конец пути.

Формально говоря, из этого доказательства не видно, как реально найти это самое множество всех достижимых из s вершин (за разумное время — не перебирая все пути). Но по существу из него можно извлечь алгоритм:

[7] Рассказывают, что много лет назад (при советской власти) при посадке в поезд с доской для сёрфинга начальник поезда потребовал справку, что эта доска *не принадлежит* никакой государственной организации. Но задумался и отстал, когда его спросили, какой организацией должна быть выдана справка.

$S \leftarrow \{s\}$

while есть ребро, ведущее из S вовне S **do**

    добавить конец этого ребра к S

od

Контрольный вопрос для программистов: что надо добавить в этот алгоритм, чтобы в случае достижимости найти путь из s в t? Для опытных программистов: как реализовать этот алгоритм так, чтобы время его работы было пропорционально числу рёбер (если граф представлен списками исходящих рёбер для каждой вершины)?

Контрольный вопрос. Доказательство теоремы несимметрично: вершина s в нём играет более центральную роль, чем вершина t. Но можно сделать и наоборот, отдав центральную роль вершине t. Какой разрез при этом получится?

Как говорят в теоретической информатике, мы сначала доказали, что достижимость лежит в классе NP (есть короткие доказательства достижимости), потом что недостижимость лежит в классе NP (есть короткие доказательства недостижимости), и наконец заметили, что достижимость лежит в классе P, объяснив, как найти короткое доказательство того или другого.

3.3.5 Компоненты сильной связности и ациклические графы

Отношение достижимости несимметрично, но можно определить симметричное отношение «достижимости в обе стороны». Будем говорить, что вершина u *сильно связана* с вершиной v, если v достижима из u и наоборот, то есть если есть путь из u в v, а также путь из v в u.

Теорема. Вершины ориентированного графа можно разбить на непересекающиеся группы, называемые *сильно связными компонентами*, при этом:

- каждая вершина графа попадает ровно в одну группу;
- любые две вершины из одной группы сильно связаны (есть пути в обе стороны);
- любые две вершины из двух разных групп не являются сильно связанными (в одну из сторон — или даже в обе — пути нет).

Для неориентированных графов аналогичное утверждение было доказано в разделе 3.2.3.

Доказательство повторяет рассуждение для неориентированных графов: для каждой вершины v мы составляем группу $C(v)$ из всех сильно связанных с ней вершин и доказываем, что получается искомое разбиение.

На самом деле это общее рассуждение: свойства рефлексивности, симметричности и транзитивности для отношения (в данном случае для отношения сильной связанности) гарантируют возможность разбиения на классы. Такие отношения называют *отношениями эквивалентности.*

Два крайних случая:

(1) в графе любые две вершины сильно связаны (из любой вершины можно попасть в любую, сильно связанная компонента только одна). Такие графы называют *сильно связными*.

(2) никакие две различные вершины не являются сильно связанными (для любой пары хотя бы в одну сторону пути нет, сильно связные компоненты одноэлементные). Такие графы называются *ациклическими* — название объясняется следующей теоремой.

Теорема. Следующие свойства ориентированного графа равносильны:

- Сильно связные компоненты одноэлементны.
- В графе нет циклов (путей вида $a_1, \ldots, a_n$, где начало совпадает с концом, то есть $a_1 = a_n$; при $n = 1$ путь из единственной вершины, имеющий нулевую длину, циклом не считается).
- Вершины графа можно пронумеровать натуральными числами таким образом, чтобы все рёбра вели «вверх»: из вершины с меньшим номером в вершину с большим.

Третье условие удобно представлять себе так: если выписать вершины в порядке номеров, то рёбра графа будут идти слева направо.

Доказательство. Начнём с очевидных утверждений. Если в графе есть цикл с $n > 1$ вершинами, то вершины этого цикла сильно связаны (из любой можно попасть в любую по циклу), так что из первого свойства следует второе. Наоборот, если различные вершины a, b сильно связаны, то существуют пути из a в b и из b в a, и из этих двух путей можно составить цикл. Наконец, если возможна нумерация вершин, при которой рёбра ведут «вверх», то цикла нет: вдоль него номера вершин должны расти, и вернуться в начало мы не сможем.

Осталось доказать единственное нетривиальное утверждение: если в ориентированном графе нет циклов, то его вершины можно пронумеровать требуемым способом.

Лемма. В ориентированном графе без циклов есть вершина, из которой не выходит ни одного ребра.

Доказательство леммы. Пусть это не так и из любой вершины выходит хоть одно ребро. Возьмём какую-то вершину a_1, из неё идёт ребро в какую-то другую вершину a_2, из неё идёт ребро ещё куда-то, и так далее. Поскольку граф конечен, то рано или поздно мы придём второй раз в какую-то вершину a_i, где уже были, и участок пути после a_i свернётся в цикл — вопреки предположению, что циклов нет.

Теперь можно закончить доказательство теоремы. Выберем вершину, из которой не ведёт ни одного ребра. Ей можно без опаски присвоить номер, бóльший всех остальных номеров. Поэтому рассуждаем по индукции: удалив эту вершину (и все входящие в неё рёбра) из графа, получим граф без циклов. (Циклы в нём были бы циклами и в исходном графе.) Пронумеруем его вершины от 1 до какого-то числа n (число оставшихся вершин) с соблюдением условия (рёбра ведут из меньшего к

большему), и после этого вернём выброшенную вершину под номером $n+1$. (Базис индукции — граф с одной вершиной — очевиден.)

Эта теорему можно объяснить так. Представим себе, что есть какие-то бумаги (справки, разрешения и пр.), которые надо получать в определённом порядке: налоговая инспекция требует счёта в банке, банк требует регистрации где-нибудь ещё, и пр. Эту зависимость можно представить в виде графа: вершины это нужные бумаги, а ребро $u \to v$ показывает зависимость v от u (чтобы получить v, надо уже иметь v).[8] Доказательство тоже легко объяснить в этих терминах: если нет ни одной бумаги, которую можно получить просто так (не имея каких-то других), и из каждой инстанции нас посылают в какую-то другую, то рано или поздно мы совершим цикл и придём в инстанцию, где уже были. Если же такая простая для получения бумага есть, то начнём с того, что получим её — и сведём задачу к меньшей (которую, по предположении индукции, можно считать уже решённой).

Два заключительных замечания:

(1) Доказанную теорему тоже можно воспринимать как утверждение типа двойственности: если цикл есть, то можно его предъявить, а если цикла нет, то в качестве доказательства того, что его быть не может, можно предъявить нумерацию.

(2) Доказательство теоремы по существу даёт алгоритм поиска цикла или нумерации (что найдётся): сначала вызываем процедуру, соответствующую лемме, которая либо находит цикл, либо находит вершину, из которой не выходят рёбра. В первом случае задача решена, во втором случае мы рекурсивно вызываем тот же алгоритм для остальных вершин. Этот алгоритм не очень долгий (время работы полиномиально зависит от числа вершин), но не оптимальный: если нам нужно искать такую нумерацию на практике (это называют «топологической сортировкой»), то есть более эффективные алгоритмы (время их работы пропорционально числу рёбер, если граф представлен списками рёбер).

3.3.6 Графы преобразований

Рассмотрим для примера такую задачу: с числом 1 тысячу раз проделали операцию «возведение в квадрат и прибавление единицы»: после первой операции получилось 2, после второй 5, после третьей 26 и так далее. Какая будет последняя цифра у получившегося после тысячи операций числа?

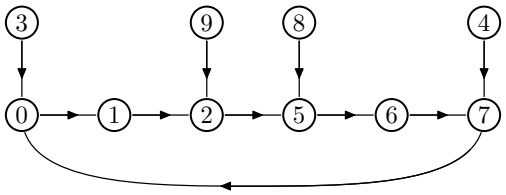
Формально: $a_0 = 1$, $a_{n+1} = a_n^2 + 1$; надо найти последнюю цифру a_{1000}.

Хотя на вид задача выглядит пугающе — число астрономически большое — на самом деле она совсем не сложная. Для начала заметим, что последняя цифра квадрата числа определяется последней цифрой самого числа, предыдущих цифр знать не надо. (Почему? вспомните, как происходит умножение в столбик.) То же самое и с прибавлением единицы. Соответственно, можно сказать, какая будет последняя цифра числа $a^2 + 1$ в зависимости от последней цифры числа a:

[8] Те счастливые люди, которые никогда не собирали документов, могут использовать другую метафору: ребро (носок $\to$ ботинок) означает, что ботинок нужно надевать, когда носок уже одет.

a	0	1	2	3	4	5	6	7	8	9
a^2+1	1	2	5	0	7	6	7	0	5	2

Для удобства представим информацию в этой таблице в виде ориентированного графа: первая колонка изображается стрелкой $0 \to 1$, вторая — стрелкой $1 \to 2$, третья — стрелкой $2 \to 5$ и так далее.



Теперь задачу можно переформулировать так: *выйдя из точки* 1*, мы прошли тысячу раз по стрелкам, куда мы пришли*? В этом виде её решить совсем просто: через шесть шагов мы возвращаемся в точку 1, значит, через $996 = 166 \cdot 6$ шагов мы тоже вернёмся в точку 1, останется четыре шага, которые приведут нас в точку 7. Ответ: число a_{1000} оканчивается на 7.

Контрольные вопросы. На какую последнюю цифру оканчивается число a_{2000}? На какую последнюю цифру оканчивается b_{1000}, если $b_0 = 4$, а $b_{n+1} = b_n^2 + 1$? На какую последнюю цифру оканчивается c_{1000}, если $c_0 = 1$, $c_{n+1} = c_n^2 + 2$?

В этом примере из каждой вершины выходила ровно одна стрелка,[9] поэтому движение по стрелкам было предопределено (нет выбора, куда идти). Начав двигаться из некоторой точки, мы в какой-то момент попадаем в уже пройденную точку (граф конечен), и после этого зацикливаемся. Это позволяет легко рассчитать, в какой точке мы окажемся после данного числа шагов.

Задача. Предположим, что последовательность чисел задана соотношением $a_{n+1} = f(a_n)$, где f — некоторая функция (определённая на всех числах). Покажите, что либо все члены последовательности различны, либо она периодична: после некоторого начала (предпериода) числа начинают повторяться (период).

Задача (Продолжение) Покажите, что второй случай имеет место тогда и только тогда, когда $a_{2n} = a_n$ при некотором n.

Задача. Имеется n людей, которые хотят обменяться квартирами (каждый хочет переехать в квартиру кого-то другого, при этом в каждую квартиру хочет переехать только один человек). За один день несколько пар жильцов могут поменяться квартирами (в каждой паре жилец переезжает на место другого). Докажите, что весь переезд можно провести за два дня (в конце первого дня все жильцы размещены в каких-то квартирах — возможно, не в тех, откуда они выезжают и не в тех, куда они переезжают).

Задача для тех, кто видел кубик Рубика. Если повернуть одну (скажем, правую) грань на 90° по часовой стрелке, и сделать так четыре раза, то кубик вернётся в исходное положение. Назовём это действие A. Аналогичным образом, если повернуть

[9] Такие графы называют графами функций (соответствующая функция отображает начало стрелки в её конец, так что стрелка направлена от аргумента к значению).

верхнюю грань на 90° по часовой стрелке (назовём это преобразование B), то кубик тоже вернётся в исходное положение. А теперь будем чередовать A и B, выполняя их в последовательности $ABABAB\ldots$. Будет ли момент, когда кубик вернётся в исходное положение?

3.4 Эйлеровы циклы

3.4.1 Определение

В этом разделе мы вернёмся к уже обсуждавшимся мостам в Кёнигсберге (см. раздел 3.1.3) и сформулируем и докажем общее утверждение. Оно существует в двух вариантах: для неориентированных и ориентированных графов.

Мы уже говорили, что *циклом* в графе называется путь, в котором начало и конец совпадают, то есть последовательность вершин $a_1, a_2, \ldots, a_n$, в которой (a_i, a_{i+1}) является ребром графа (при всех $i = 1, 2 \ldots, n-1$) и начало совпадает с концом: $a_1 = a_n$.[10]

Это определение относится и к неориентированным, и к ориентированным графам. Во втором случае имеется в виду, что ребро (a_i, a_{i+1}) ведёт из a_i в a_{i+1}, то есть a_i является началом, а a_{i+1} —концом ребра.

Цикл называется *эйлеровым*, если он проходит по всем рёбрам графа по одному разу (любое ребро входит в цикл, и никакое ребро не входит дважды). Для неориентированных графов мы имели в виду именно это, когда просили нарисовать граф не отрывая карандаша от бумаги и вернуться в исходную точку (в разделе 3.1.3).

3.4.2 Критерий существования

Теперь мы можем сформулировать обещанный критерий, сначала для неориентированных графов, потом для ориентированных.

Теорема. Неориентированный граф без вершин нулевой степени содержит эйлеров цикл тогда и только тогда, когда он связен и степени всех вершин чётны.

Теорема. Ориентированный граф без вершин нулевой степени (в которые не входит и из которых не выходит рёбер) содержит эйлеров цикл тогда и только тогда, когда он сильно связен и у любой вершины входящая степень равна исходящей.

Оговорка про вершины нулевой степени необходима: они нарушают связность графа, но никак не мешают эйлерову циклу (их можно удалить, и с точки зрения поиска эйлерова цикла это ничего не меняет).[11]

[10] Можно не выделять в цикле точку начала и конца, то есть считать, скажем, $a \to b \to c \to a$ и $b \to c \to a \to b$ одним и тем же циклом. Мы будем придерживаться исходного определения, с выделенным началом и концом, если противное не оговорено явно.

[11] Педантичные читатели заметили бы, что мы не оговорили особый случай графа (ориентированного графа) совсем без вершин. В нём нет вершин нулевой степени. Он связен (соответственно сильно связен), и можно считать, что в нём есть пустой эйлеров цикл — ну, или отдельно потребовать, чтобы в графе были вершины.

Доказательство. Будем доказывать параллельно оба варианта теоремы. Пусть сначала эйлеров цикл есть. Тогда он проходит через все вершины (поскольку они имеют ненулевую степень), и по нему можно дойти от любой вершины до любой. Значит, граф связен (сильно связен в ориентированном случае).

Теперь про степени. Возьмём какую-то вершину v, пусть она встречается в цикле k раз. Идя по циклу, мы приходим в неё k раз и уходим k раз, значит, использовали k входящих и k исходящих рёбер. При этом, раз цикл эйлеров, других рёбер у этой вершины нет, так что в ориентированном графе её входящая и исходящая степени равны k, а в неориентированном графе её степень равна $2k$. Таким образом, в одну сторону критерий доказан.

Рассуждение в обратную сторону чуть сложнее. Будем рассматривать пути, которые не проходят дважды по одному ребру. (Таков, например, путь из одного ребра.) Выберем среди них самый длинный путь

$$a_1 \to a_2 \to a_3 \to \ldots \to a_{n-1} \to a_n$$

и покажем, что он является искомым циклом, то есть что $a_1 = a_n$ и что он содержит все рёбра.

В самом деле, если он самый длинный, то добавить к нему ребро $a_n \to a_{n+1}$ уже нельзя, то есть все выходящие из a_n рёбра уже использованы. Это возможно, лишь если $a_1 = a_n$. В самом деле, если вершина a_n встречалась только внутри пути (пусть она входит k раз внутри пути и ещё раз в конце пути), то мы использовали $k+1$ входящих рёбер и k выходящих, и больше выходящих нет. Это противоречит равенству входящей и исходящей степени (в ориентированном случае) или чётности степени (в неориентированном случае).

Итак, мы имеем цикл, и осталось доказать, что в него входят все рёбра. В самом деле, если во всех вершинах цикла использованы все рёбра, то из вершин этого цикла нельзя попасть вне цикла, то есть использованы все вершины (мы предполагаем, что граф связен или сильно связен) и, следовательно, все рёбра. С другой стороны, если из какой-то вершины a_i выходит ребро $a_i \to v$, то путь можно удлинить до

$$a_i \to a_{i+1} \to \ldots \to a_n = a_1 \to a_2 \to \ldots \to a_i \to v,$$

вопреки нашему выбору (самого длинного пути). Аналогично можно получить противоречие и для входящего ребра $v \to a_i$, добавив его в начало. (А можно заметить, что если есть неиспользованное входящее ребро, то есть и неиспользованное выходящее.) Это рассуждение было для ориентированного случая, но в неориентированном всё аналогично. Теорема доказана.

Помимо эйлеровых циклов, можно рассматривать *эйлеровы пути* — пути в графе, которые проходят один раз по каждому ребру. (Для неориентированных графов: рисуем картинку, не отрывая карандаша от бумаги, но не обязаны вернуться в исходную точку.) Для них тоже есть критерий: в неориентированном случае нужно, чтобы граф был связен и было не более двух вершин нечётной степени.

Контрольный вопрос. Может ли такая вершина быть только одна?

Задача для самостоятельного решения. Как сформулировать аналогичный критерий существования эйлеровых путей в ориентированном графе?

3.4.3 Последовательности де Брёйна

Вот любопытное применение этой теоремы. Пусть мы хотим написать цифры по кругу — скажем, по окружности круглого стола. Идя вокруг стола, мы читаем двузначные числа (образованные парами соседних цифр). Требуется, чтобы каждое число от 00 до 99 встретилось при этом ровно по одному разу.

Попробуем это доказать так (предупреждение: в этом рассуждении есть ошибка, попробуйте её заметить, не читая следующего абзаца). Рассмотрим ориентированный граф с вершинами $0, 1, 2, \ldots, 9$, в котором есть все возможные рёбра (любые две вершины соединены рёбрами в обе стороны). Он, очевидно, сильно связен, и из любой вершины входит столько же рёбер, сколько и выходит. Значит, по доказанной теореме в нём есть эйлеров цикл, и это как раз то, что нам надо: написав вершины этого цикла по окружности, мы прочтём вдоль неё каждое двузначное число по одному разу.

Заметили ошибку? В нашем определении графа не разрешались петли, ведущие из вершины в себя. Поэтому числа $00, 11, \ldots, 99$ по кругу не встретятся. Но их можно добавить: для каждой цифры нужно выбрать какое-то место, где её удвоить.

С точки зрения задачи об эйлеровом цикле петли (рёбра из вершины в неё же) и кратные рёбра (несколько рёбер с общими началами и концами) вполне имеют смысл. Можно было бы изменить определение графа, разрешив их.[12] Например, можно считать, что граф задаётся множествами E («вершин») и V («рёбер») и двумя отображениями множества E в V, которые ставят в соответствие каждому ребру его начало и конец. (Как тогда следует определить путь?) После этого можно доказать критерий эйлеровости для таких графов, и обойтись без сделанной нами оговорки, но мы эту тему развивать не будем.

Задача для самостоятельного решения. Докажите, что по кругу можно написать тысячу цифр так, чтобы вдоль круга можно было бы прочесть все трёхзначные числа от 000 до 999 по одному разу.

Из этой задачи следует, что если для открытия кодового замка нужно нажать три цифры кода в правильном порядке (и не важно, какие цифры нажимались до этого), то его можно открыть, сделав не более чем 1002 нажатий. (Почему нельзя обойтись меньшим числом?)

Построенные таким образом последовательности называются последовательностями де Брёйна (1918–2012), хотя частный случай последовательностей из двух элементов был рассмотрен ещё в XIX веке. Мы привели этот пример, поскольку в нём простой критерий существования эйлерова цикла при правильном взгляде на вещи сразу же даёт требуемый результат — совсем не очевидный, если не догадаться про интерпретацию в терминах графов.

3.4.4 Гамильтоновы циклы

В этом разделе мы ничего не будем доказывать, а только дадим определение: *гамильтонов цикл* — это цикл, который проходит через каждую вершину ровно по

[12] Собственно говоря, граф мостов в Кёнигсберге как раз содержит рёбра с общими концами.

одному разу. (Начало и конец в цикле a_1–a_2–…–a_n мы считаем за один раз. Аналогично определяется и *гамильтонов путь* — путь, в котором каждая вершина встречается ровно один раз (тут уже начало и конец должны быть разными вершинами и считаются каждый за себя).

Хотя определение и похоже, но ситуация с поиском гамильтоновых путей (циклов) принципиально сложнее, чем с поисками эйлеровых: тут уже нет простого критерия или алгоритма проверки (есть ли гамильтонов путь в данном графе). Более того, есть некоторые причины считать, что эффективного (быстро работающего) алгоритма проверки нет. Как говорят, эта задача является NP-полной, и хотя пока и не доказано, что для таких задач нет быстрых алгоритмов, но мало кто на это надеется — «too good to be true», как говорится. Впрочем, их отсутствие, может быть, и к лучшему — если такие алгоритмы обнаружатся, то нынешним системам криптографии (используемым в банковском деле, в протоколах https, ssl и других), придёт конец.

3.5 Двудольные графы

3.5.1 Определение

Двудольным графом называется неориентированный граф, в котором вершины заранее разделены на две *доли* — *левые* и *правые*, и все рёбра соединяют вершины из разных долей (нет рёбер, соединяющих вершины одной доли). Другими словами, чтобы задать двудольный граф, надо указать два конечных множества L (левую долю) и R (правую долю) и указать, какие вершины левой доли соединены с какими вершинами правой доли. В программе такой граф представляется булевым массивом $edge[0..l-1, 0..r-1]$, где l, r — количества вершин в левой и правой доле; мы считаем, что левые вершины пронумерованы от 0 до $l-1$, а правые от 0 до $r-1$. При этом $edge[i][j]$ истинно, если i-я вершина левой доли соединена с j-й вершиной правой доли. Наглядно можно представлять себе прямоугольную таблицу $l \times r$, в которой в некоторых клетках стоят плюсы (означающие наличие ребра). Наконец, можно определить двудольный граф формально как тройку (L, R, E), где L и R — конечные множества, а $E \subset L \times R$ — некоторое множество упорядоченных пар (первые элементы в L, вторые в R).[13]

В жизни встречаются разные ситуации, которые можно изобразить двудольными графами. Например, в левой доле могут быть учёные, а в правой статьи, и наличие ребра $l - -r$ означает, что учёный l является одним из авторов статьи r. Или в левой доле могут быть студенты, а в правой — учебные курсы, и ребро означает, что студент прослушал курс. Или слева могут быть кавалеры, а справа дамы, и ребро означает, что они танцевали друг с другом. Наконец, можно вспомнить о теореме

[13] В этом определении озникает вопрос, могут ли множества L и R иметь общие элементы. В принципе это можно и разрешить, но тогда, говоря о вершине, надо всегда уточнять, рассматриваем ли мы её как вершину левой доли или как вершину правой. Поэтому обычно предполагают, что L и R не имеют общих элементов.

Холла и рассмотреть двудольный граф, в котором в левой доле стоят школьные кружки, в правой — школьники, и ребро означает участие школьника в кружке.

3.5.2 Двудольные графы и раскраска в два цвета

Мы предполагали, что при задании двудольного графа указано, какие вершины в левой доле, а какие в правой. Но можно действовать в другом порядке: взять произвольный неориентированный граф и спросить себя, можно ли так разделить его вершины на левую и правую долю, чтобы все рёбра соединяли вершины разных долей. Если обозначать это деление цветом: можно ли так раскрасить вершины графа в два цвета, чтобы концы любого ребра были разного цвета?

Если требуется доказать, что такое возможно, достаточно предъявить раскраску. А что надо предъявить, чтобы доказать, что такое невозможно? Наверно, вы уже догадались, но вот формальное утверждение.

Теорема. Раскраска описанного типа возможна когда и только тогда, когда в графе нет циклов нечётной длины.

Напомним, что циклом называется последовательность вершин $a_1, \ldots, a_n$, в которой вершины a_i и a_{i+1} соединены ребром, и $a_1 = a_n$. Длиной такого цикла является число рёбер, то есть $n-1$. (Не удивляйтесь, что рёбер в цикле меньше, чем вершин — это только кажется, потому что a_1 и a_n одна и та же вершина).

Доказательство. Если в графе есть цикл нечётной длины, то его нельзя раскрасить: соседние вершины должны быть противоположных цветов, и дойдя до конца, мы получим противоречие. (Попробуйте раскрасить вершины треугольника так, чтобы концы любого ребра были разного цвета!)

Чтобы доказать обратное, предположим, что циклов нечётной длины нет. Выберем некоторую вершину a и решим, что она белая. (Это не ограничивает общности — всегда можно поменять цвета местами.) Для любой другой вершины b посмотрим, сколько рёбер в пути от a к b.

Лемма. Если есть два пути из a в b, то либо в обоих чётное число рёбер, либо в обоих нечётное.

Доказательство. Если есть путь $a \to b$ с чётным числом рёбер, а также другой путь $a \to b$ с нечётным числом рёбер, то есть цикл с нечётным числом рёбер. А именно, пойдём из a в b по первому пути и вернёмся по второму. Это противоречит предположению.

Таким образом, мы поделили вершины графа на два типа: соединённые с a путями чётной длины и путями нечётной длины. Если какая-то вершина соединена с a путями чётной длины, то её соседи соединены путями нечётной длины (один такой путь — через соседа — заведомо есть, а тогда и все пути имеют нечётную длину). Требуемая раскраска построена.

Заметили пробел в рассуждении? На самом деле мы раскрасили не весь граф, а только связную компоненту вершины a. Но это легко исправить: каждую связную компоненту можно раскрашивать независимо, так как рёбер между ними нет.

Контрольный вопрос. Сколькими способами можно раскрасить в два цвета (с соблюдением требований) граф, в котором k связных компонент? (Ответ: 2^k.)

Замечание для программистов. Если попытаться реализовать это доказательство как алгоритм раскраски, получится сложно. Проще совместить поиск раскраски с поиском связных компонент и сразу же их раскрашивать: когда связная компонента растёт за счёт добавления соседа, цвет добавленной вершины определён однозначно.

Задача для самостоятельного решения. Можно ли в критерии раскрашиваемости графа запретить лишь *простые* циклы нечётной длины?

Задача для самостоятельного решения. *Булев куб* размерности n — это неориентированный граф, вершинами которого являются двоичные слова длины n, а рёбра соединяют слова, отличающиеся в одной позиции. Всегда ли такой граф можно раскрасить в два цвета?

Задача для самостоятельного решения. Всякое ли дерево можно раскрасить в два цвета?

3.5.3 Степени вершин

Для вершин двудольного графа можно определить *степени* обычным образом — как число выходящих рёбер (или, что то же самое, как число соседей в другой доле). Скажем, в примере с авторами и статьями степень автора — это число публикаций (важный параметр — им пытаются измерять научную значимость автора!). А степень статьи — это число её авторов.

Контрольный вопрос. Каков житейский смысл степеней в других приведённых нами примерах?

Как и раньше, есть простой факт о сумме степеней.

Теорема. Сумма степеней всех вершин левой доли равна сумме степеней всех вершин правой доли, и равна числу рёбер в графе.

Контрольный вопрос. Почему это верно?

Часто иллюстрируют эту теорему так: если на контрольной было 20 задач, каждую задачу решили три школьника, и каждый школьник решил две задачи, то сколько было школьников?

Контрольный вопрос. Где здесь граф, степени, и сколько-таки было школьников?

3.5.4 Паросочетания

Задачу, решённую в разделе 3.1.6, теперь можно переформулировать так: имеется двудольный граф, при этом степени всех вершин (и слева, и справа) равны 2. Тогда можно удалить часть рёбер так, чтобы степени всех вершин стали равны 1. (Такой граф задаёт схему разбора задач.)

Задачу из раздела про индукцию можно переформулировать так. Пусть есть двудольный граф, причём если мы возьмём любые k вершин из левой части, то у них всего не меньше k соседей (внимание: мы не складываем их степени, а смотрим, сколько вершин в правой части являются соседями одной из выбранных в левой части вершин). Тогда можно удалить часть рёбер графа так, чтобы у всех вершин слева степени стали равны 1.

Обе эти задачи можно сформулировать в терминах «паросочетаний».[14] Мы называем *паросочетанием* двудольный граф, у которого степени всех вершин не больше 1. Такой граф устаналивает некоторое взаимно однозначное соответствие между частью вершин левой доли и частью вершин правой доли. Размер паросочетания не превосходит степени любой из долей (очевидно), и часто задача состоит в том, чтобы найти в графе паросочетание максимального размера. Лучшее, на что мы можем надеяться — достигнуть размера меньшей из долей, и в упомянутых выше задачах мы доказывали, что это возможно. Сформулируем ещё раз результат второй задачи, используя введённую терминологию.

Теорема Холла. Пусть в двудольном графе любой набор из k левых вершин имеет не менее k соседей справа. Тогда в этом графе существует паросочетание, включающее все вершины левой доли.

Соседи набора — соседи хотя бы одной из его вершин; «в графе есть паросочетание» — можно получить это паросочетание, удалив часть вершин графа.

Контрольный вопрос. Почему верно обратное к теореме Холла утверждение (если есть паросочетание со всеми левыми вершинами, то любой набор с k вершинами имеет не менее k соседей)? [Соседи берутся из паросочетания.]

Мы привели доказательство теоремы Холла в разделе про индукцию. Если это рассуждение показалось сложным или искусственным, то, может быть, вам больше понравится другое доказательство, которое будет дано ниже в разделе про сети и потоки (с использованием критерия Форда–Фалкерсона).

Задача для самостоятельного решения. Рассмотрим граф с N вершинами слева и справа и будем искать в нём паросочетание размера N. Можно применить теорему Холла «слева-направо» и «справа-налево»: во втором случае требуется, чтобы любой набор из k вершин справа имел не менее k соседей слева. Как понять, не доказывая теорему Холла, что эти два условия эквивалентны?

В заключение этого раздела объясним, почему первая из рассмотренных задач (если все вершины графа имеют степень 2, то в нём есть паросочетание, включающее все вершины) следует из теоремы Холла. Мы уже знаем, что в таком графе поровну вершин слева и справа (половина числа рёбер), и надо доказать лишь, что любой набор из k вершин слева имеет не менее k соседей справа. Из вершин этого набора выходит $2k$ рёбер, и все эти $2k$ рёбер ведут в соседей. Поскольку в каждого соседа ведёт не более двух рёбер, то соседей не меньше k.

Преимущество этого рассуждения перед тем, которое у нас было раньше, в разделе 3.1.6: его легко обобщить на случай, когда каждый школьник решил 3 задачи, а каждую задачу решили 3 школьника. Тут уже граф не разбивается на циклы и как быть без теоремы Холла, непонятно. А с теоремой Холла всё так же: набор из k вершин имеет не менее k соседей, иначе выходящим из него $3k$ рёбрам некуда будет приткнуться (больше чем по три им собираться не разрешено). То же самое и для любого другого числа вместо 3.

Когда-то это утверждение предлагалось на школьных олимпиадах с таким «ожив-

[14] Это традиционный термин, хотя звучит довольно странно. По-английски используется слово *matching*.

ляжем»: каждый из 100 завов соединён телефонными проводами с 15 замами, при этом каждый из 100 замов соединён с 15 завами; надо доказать, что можно обрезать часть проводов так, чтобы каждый зав остался соединённым ровно с одним замом.

Задача для самостоятельного решения. Покажите, что двудольный граф с долями размера n, в котором все вершины имеют степень k, можно разбить на k паросочетаний размера n. (Другими словами, можно раскрасить все рёбра в k цветов, и рёбра каждого цвета будут паросочетанием.)

Задача для самостоятельного решения. В квадратной таблице $n \times n$ стоят неотрицательные числа, причём суммы чисел в каждой строке и в каждом столбце равны 1.[15] Докажите, что можно поставить n ладей, не бьющих друг друга, на клетки с положительными числами.

Задача для самостоятельного решения. В каждой клетке прямоугольной таблицы $m \times n$ стоит по гному. Они хотят произвести перестановку (каждый гном перемещается в выбранную им клетку, причём ни одну клетку не выброало несколько гномов). Докажите, что это всегда можно сделать в три дня, причём в первый день все гномы остаются с своих столбцах (перемещаются лишь по вертикали), второй день — в своих строках, а третий день — снова в столбцах.

3.6 Клики и независимые множества

Кликой называется такое подмножество вершин графа, каждая пара которых связана ребром.

Независимым множеством называется такое подмножество вершин графа, никакая пара которых не связана ребром.

Задача. Найдите максимальный размер клики и максимальный размер независимого множества в графе–пути P_n.

Решение. Будем считать, что вершины графа–пути занумерованы числами от 1 до n так, что рёбра соединяют пары вершин с номерами i, $i+1$. Если разность номеров двух вершин больше 1, ребра между ними нет. Поэтому клики размера 3 в графе–пути нет: из любых трёх целых чисел хотя бы одна пара отличается хотя бы на 2.

С другой стороны, любая пара вершин, соединённых ребром, образует клику размера 2. Поэтому максимальный размер клики в графе P_n равен 2 при $n \geqslant 2$. (Оставшийся случай графа–пути P_1 разберите самостоятельно.)

Вершинным покрытием называется такое множество вершин S, что для любого ребра один из концов лежит в S.

С независимыми множествами чуть сложнее. Во-первых, ответ зависит от чётности n. Для чётных n максимальный размер независимого множества равен $n/2$. Например, независимое множество образуют вершины с чётными номерами $2, 4, \ldots,$

[15] В теории вероятностей это называется *дважды стохастическими матрицами*; для любой такой таблицы можно определить распределение вероятностей на бесконечных в обе стороны последовательностях символов в n-буквенном алфавите: в клетке i, j стоит условная вероятность появления j за i, равная условной вероятности появления i перед j.

n. Для нечётных n максимальный размер независимого множества равен $(n+1)/2$. Независимое множество такого размера образуют вершиными с нечётными номерами 1, 3, ..., n.

Чтобы объединить эти два ответа в один, можно использовать функцию $\lceil x \rceil$, которая равна наименьшему целому числу, которое не меньше x. В обоих предыдущих примерах размер независимого множества равен $\lceil n/2 \rceil$. (Проверьте!)

Пока мы лишь привели примеры независимых множеств достаточно большого размера. Теперь нужно доказать, что бóльших независимых множеств в графе–пути нет. Пусть вершины с номерами $i_1 < i_2 < \cdots < i_k$ образуют независимое множество в графе–пути. По определению это означает, что между этими вершинами нет ребра, то есть $i_{j+1} - i_j > 1$. Значит, между каждой парой вершин независимого множества есть ещё хотя бы одна. А всего вершин в графе–пути n. Поэтому

$$2k - 1 = k + (k-1) \leqslant n, \quad \text{что равносильно } k \leqslant (n+1)/2.$$

Проверьте, что последнее неравенство означает, что $k \leqslant \lceil n/2 \rceil$ (не забудьте, что размер независимого множества — целое число).

Задача. Докажите, что S — вершинное покрытие тогда и только тогда, когда $V \setminus S$ — независимое множество.

С независимыми множествами и кликами связана одна из самых интересных теорем в комбинаторике — теорема Рамсея. В первом разделе этой главы мы уже её обсуждали, говоря о попарно знакомых и незнакомых людях.

Теорема Рамсея. Для любых k, n найдётся такое число $R(k,n)$, что в любом графе на $R(k,n)$ вершинах есть или клика размера k, или независимое множество размера n.

Ясно, что если утверждение теоремы справедливо для графа на R вершинах, то оно справедливо и для графов с $N > R$ вершинами. Поэтому обычно под $R(k,n)$ понимают *число Рамсея* —- минимальное количество вершин, для которого справедлива теорема.

Доказательство. Мы перескажем рассуждение, намеченное в первом разделе этой главы, для общего случая. Для этого применим индукцию.

Будем доказывать индукцией по s, что для любой пары чисел k, n такой, что $k + n = s$ справедливо утверждение теоремы.

База индукции $s = 2$ очевидна: $2 = 1 + 1$ — это единственный способ разложить число 2 в сумму целых положительных слагаемых, а одна вершина является одновременно и кликой, и независимым множеством.

Теперь докажем индуктивный переход. Предположим, что утверждение выполнено для всех пар (k,n) таких, что $k + n \leqslant s$.

Докажем его для пары (k,n) такой, что $k + n = s + 1$. По индуктивному предположению утверждение теоремы выполнено для пар $(k-1, n)$ и $(n, k-1)$.

Рассмотрим граф на $R(k-1,n) + R(k,n-1)$ вершине и возьмём какую-то вершину v этого графа. Пусть степень вершины v равна d.

Заметим, что если $d < R(k-1,n)$, то $R(k-1,n) + R(k,n-1) - 1 - d \geqslant R(k,n-1)$.

Вершин в графе за исключением вершины v ровно $R(k-1,n)+R(k,n-1)$ штук. Поэтому у вершины v есть или $R(k-1,n)$ соседей ($d \geqslant$, или $R(k,n-1)$ вершин, не связанных с v ребром.

Оба случая рассматриваются аналогично.

Первый случай. В индуцированном соседями вершины v подграфе по предположению найдётся или клика размера $k-1$, или независимое множество размера n. В первом варианте добавление вершины v даёт клику в исходном графе размера k, во втором варианте в исходном графе есть независимое множество размера n.

Второй случай. В индуцированном несоседями вершины v подграфе по предположению найдётся или клика размера k, или независимое множество размера $n-1$. В первом варианте в исходной графе есть клика размера k, а во втором добавление вершины v даёт независимое множество размера n в исходном графе.

Итак, мы доказали утверждение теоремы и для произвольной пары (k,n), для которой $k+n = s+1$. Индуктивный переход доказан, и теорема следует из принципа математической индукции.

Из доказательства теоремы получается также неравенство для чисел Рамсея

$$R(k,n) \leqslant R(k-1,n) + R(k,n-1),$$

которое напоминает рекуррентное соотношение для биномиальных коэффициентов. Поскольку $R(1,n) = R(k,1) = 1$, то легко убедиться, что

$$R(k,n) \leqslant \binom{k+n-2}{k-1}.$$

Лекция 4

Арифметика остатков

4.1 Чётные и нечётные числа

Все знают, что числа бывают чётные и нечётные. Чётные делятся на 2 без остатка, а нечётные дают остаток 1. Другими словами, чётные числа имеют вид $2k$ для целых k, а нечётные $2k+1$, тоже при целых k. Скажем, число 0 чётное ($0 = 2 \cdot 0$), а число -3 нечётное ($-3 = 2 \cdot (-2) + 1$).

Несложно проверить, что сумма двух чётных или двух нечётных чисел чётна, а сумма чётного и нечётного числа нечётна:

$+$	Ч	Н
Ч	Ч	Н
Н	Н	Ч

$\times$	Ч	Н
Ч	Ч	Ч
Н	Ч	Н

Например, если мы складываем чётное и нечётное число, то получаем $2k + (2l + 1) = 2(k + l) + 1$, то есть нечётное число. А если мы умножаем два нечётных числа, то получаем

$$(2k + 1)(2l + 1) = 4kl + 2k + 2l + 1 = 2(2kl + k + l) + 1,$$

то есть нечётное число.

Задача 4.1. Папа с сыном играют в такую игру: каждый пишет на бумажке число, не говоря его другому, потом они открывают эти числа, и если сумма чётная, выигрывает папа, а если нечётная, то сын. Честная ли это игра? Тот же вопрос, если вместо суммы чисел берётся их произведение.

Среди целых чисел чётные и нечётные встречаются одинаково часто: если мы возьмём большой интервал подряд идущих чисел, то количество чётных и нечётных в этом интервале будет почти одинаково (отличаться максимум на 1)

Задача 4.2. Почему?

Если заменить в таблицах буквы Ч и Н на 0 и 1, взяв нуль и единицу в качестве представителей классов чётных и нечётных чисел, то получатся почти что обычные таблицы сложения и умножения:

+	0	1
0	0	1
1	1	0

×	0	1
0	0	0
1	0	1

Разница с обычными только в одном месте: $1 + 1 = 0$.

4.2 Деление на 3 и остатки

Попробуем составить аналогичные таблицы сложения и умножения для чисел, делящихся и не делящихся на 3. Тут сразу же возникает проблема: мы не знаем, что сказать про сумму двух чисел, не делящихся на 3. Она может делиться на 3 (например, $1 + 5 = 6$), а может и не делиться (например, $2 + 5 = 7$). Дело в том, что не делящиеся на 3 числа могут быть двух видов: одни дают остаток 1 (имеют вид $3k + 1$ при целом k), а другие дают остаток 2 (имеют вид $3k + 2$).

Задача 4.3. К какому типу относится число 1000? А число -1? Найдите соответствующие значения k.

Задача 4.4. Покажите, что все три типа чисел (делящиеся на 3, дающие остаток 1 и дающие остаток 2) встречаются одинаково часто: в большом отрезке подряд идущих чисел их количества отличаются максимум на 1.

Теперь можно составить аналогичную таблицу сложения и умножения остатков по модулю 3:

+	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

×	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

Каждую клеточку в этой таблице несложно проверить. Например,

$$(3k + 1) + (3l + 2) = 3(k + l) + 3 = 3(k + l + 1) + 0$$

и

$$(3k + 2)(3l + 2) = 9kl + 6k + 6l + 4 = 3(3kl + 2k + 2l + 1) + 1.$$

Задача 4.5. Папа с сыном играют в такую игру: каждый пишет на бумажке число, не говоря его другому, потом они открывают эти числа, и если сумма не делится на 3, выигрывает папа, а если делится, то сын. Честная ли это игра?

Аналогичные таблицы можно составить не только для деления на 2 и 3, но и для любого числа. Но сначала мы дадим более формальные определения.

4.3 Деление с остатком

Говорят, что целое число a делится на целое число b, если $a = bk$ для некоторого целого числа k. В этом случае говорят также «a кратно b», и «b является делителем числа a».

В этом определении можно было бы сказать: «если частное a/b целое», но этим бы исключался случай $b = 0$, который формально допустим по нашему определению. Правда, особого смысла в нём всё равно нет: единственное число, которое делится на 0, это число 0. Определение допускает также отрицательные a и b: скажем, число -6 делится на -2 (а также и на 2), всего у него 8 делителей, если считать и положительные, и отрицательные.

Задача 4.6. Что это за делители?

Впрочем, обычно, говоря о количестве делителей у положительного целого числа, имеют в виду только положительные делители (считая единицу и само число).

Задача 4.7. Сколько (положительных) делителей у числа $30 = 2 \cdot 3 \cdot 5$? у числа $210 = 2 \cdot 3 \cdot 5 \cdot 7$?

Задача 4.8. Придумайте (положительное целое) число, у которого было бы ровно 6 (положительных) делителей. Тот же вопрос для 7 делителей.

Задача 4.9. Говорить о кратных можно не только для целых чисел, но и для отрезков: один отрезок кратен другому, если второй укладывается в первом целое число раз, то есть если отношение (длина первого)/(длина второго) целое. Докажите, что два отрезка имеют *общую меру* (отрезок, которому они оба кратны) тогда и только тогда, когда они имеют общее кратное.

Задача 4.10. Найдите наименьшее общее кратное отрезков длиной $15/6$ и $21/10$.

Задача 4.11. Могут ли два отрезка не иметь общего кратного?

Если два числа a и b делятся на третье число c, то и их сумма $a + b$ и разность $a - b$ делятся на это c. В самом деле, если $a = kc$ и $b = lc$, то $a + b = (k + l)c$ и $a - b = (k - l)c$.

Для делимости произведения достаточно делимости одного из сомножителей: если a делится на c, то и ab делится на c, каково бы ни было (целое) b. В самом деле, если $a = kc$, то $ab = k(bc) = (kb)c$.

Множество целых чисел называют *идеалом*, если вместе с любыми двумя числами оно содержит их сумму и разность, и вместе с любым числом оно содержит все его кратные. Используя эту терминологию, можно сказать, что для любого c множество всех кратных числа c является идеалом.

Задача 4.12. Докажите, что произведение любых трёх последовательных целых чисел делится на 3.

Задача 4.13. Докажите, что число $a^3 - a$ делится на 3 при любом целом a.

Задача 4.14. Докажите, что сумма $84+85+86+87+88+89+90$ делится на 7 и на 87.

Задача 4.15. Найдите трёхзначный и семизначный делители числа 103103103.

Задача 4.16. Какие из следующих утверждений верны: (1) если a делится на c, а b не делится на c, то $a+b$ не делится на c; (2) если a не делится на c и b не делится на c, то $a+b$ не делится на c; (3) если a не делится на c и b не делится на c, то ab не делится на c? Докажите верные и приведите контрпримеры к неверным.

Задача 4.17. Известно, что a, b, c, d — положительные целые числа, и $ab = cd$. Докажите, что если a делится на c, то d делится на b.

Задача 4.18. Числа a и b целые, причём $2a+3b$ делится на 7. Докажите, что $a+5b$ также делится на 7.

Задача 4.19. Положительное целое число a чётно, но не делится на 4. Покажите, что количество (положительных) чётных делителей a равно количеству (положительных) нечётных делителей a.

Задача 4.20. Докажите, что произведение любых k подряд идущих целых чисел делится на $k! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot k$.

Теперь определим деление с остатком. Пусть b — целое положительное число. Деля на b с остатком, мы связываем предметы в пачки по b в каждой, пока это возможно: количество полных пачек называется *частным* (говорят ещё «неполное частное», чтобы отличать от частного как дроби), и сколько-то предметов останется, их количество и называют *остатком*

Формально: разделить целое a на целое положительное b означает найти такое целое q (*частное*) и такое целое r (остаток), что

$$a = b \cdot q + r; \quad 0 \leq r < b.$$

Ограничения на r понятны: остаётся сколько-то предметов (возможно, ни одного), но меньше b, иначе возникла бы ещё одна целая пачка.

Теперь можно сформулировать теорему: *деление с остатком всегда возможно, притом единственным образом.*

Не очень понятно, что тут доказывать (неужели это не очевидно, если подумать о связывании предметов в пачки?). Тем не менее для педантов можно привести такое доказательство.

Единственность. Если $a = bq + r = bq' + r'$, то $r - r' = b(q' - q)$ и потому $r - r'$ делится на b. Но оба числа r, r' находятся в интервале $0, 1, \ldots, b-1$, так что их разность (если из большего вычесть меньше) не больше $b-1$, и может делиться на b, лишь если равна нулю. Поэтому $r = r'$, откуда и $b = b'$.

Существование можно доказать индукцией по a. Для $a = 0$ частное и остаток равны нулю: $0 = 0 \cdot b + 0$. Если $a = bq + r$, то $a + 1 = bq + (r+1)$. При этом $r + 1 \leq b$, так как $r < b$. Если $r + 1 < b$, то для $a + 1$ получаем частное q и остаток $r + 1$. Если же $r + 1 = b$, то $a + 1 = bq + b = b(q+1) + 0$, получаем частное $q + 1$ и остаток 0.

Программистам будет ближе другое доказательство существования:

```
q:=0; r:=a;
{a=bq+r; r>=0}
пока (r>=b):
   q:=q+1; r:=r-b;
{a=bq+r; r>=0; r<b}
```

Здесь равенство $a = bq + r$ и неравенство $r \geq 0$ являются, как говорят, *инвариантом цикла*, они выполняются после любого числа итераций. В самом деле, начальные значения $q = 0$ и $r = a$ удовлетворяют условию $a = qa + r$; итерация цикла происходит при $r \geq b$ и потому r после уменьшения на b остаётся неотрицательным, а условие $a = bq + r$ не нарушается, если увеличить q на единицу и одновременно уменьшить r на b. Выход из цикла происходит, когда условие нарушается, то есть $r < b$ (в дополнение к инварианту) — что и требуется определением остатка.

Внимательные читатели, наверно, уже заметили ошибку в рассуждении: оно годится лишь при $a \geq 0$, поскольку иначе инвариант $r \geq 0$ будет нарушен. Собственно, и индуктивное рассуждение годилось тоже только при $a \geq 0$. Так что для отрицательных a возможность деления с остатком надо доказывать отдельно. Пусть $a = -a'$, разделим a' на b с остатком: $a' = bq' + r'$. Если $r' = 0$, то $a = -a' = b(-q') + 0$, так что можно взять $q = -q'$ и $r = 0$. Если $r' > 0$, то можно записать

$$a = -a' = -bq' - r' = b(-q' - 1) + (b - r'),$$

так что можно взять $q = -q' - q$ и $r = b - r'$: поскольку мы предположили, что $0 < r' < b$, то $0 < b - r' < b$, и r попадает в нужный диапазон.

Аналогичная проблема возникает во многих языках программирования: целочисленное деление a на b даёт неполное частное в нашем смысле лишь при $a \geq 0$, а, скажем, -7 `div` 3 оказывается равным -2, а не -3, как требует наше определение. Тем не менее математически наше определение удобнее, так как при этом сохраняется общая формула для чисел с данным остатком: скажем, числа $3k + 1$ при всех целых k при делении на 3 дают остаток 1 (и частное k).

Задача 4.21. Можно ли разрезать шахматную доску 8×8 на прямоугольники 3×1?

Задача 4.22. Найти число вида $100 * **$ (звёздочки обозначают некоторые цифры), которое делится на 547.

Задача 4.23. Книги на столе пытались связывать в пачки по 2, по 3, по 4 и по 5 книг, и каждый раз оставалась одна лишняя. Сколько книг было на столе? (Известно, что их было больше одной и не больше 100.)

Задача 4.24. Число a даёт остаток 6 при делении на 12. Может ли оно давать остаток 12 при делении на 20?

|

4.4 Сравнения по модулю

Если два числа a и b дают одинаковые остатки при делении на положительное число N, то говорят, что они *сравнимы* по модулю N, и пишут $a \equiv b \pmod N$.

Эквивалентное определение: a и b сравнимы по модулю N, если разность $a - b$ делится на N. (В самом деле, если они дают одинаковый остаток r, то $a = kN + r$, $b = lN + r$, и $a - b = kN - lN = (k - l)N$. Наоборот, если $a - b = mN$, и b даёт остаток r, то $b = lN + r$ и $a = (a - b) + b = mN + lN + r = (m + l)N + r$, то есть a даёт тот же остаток r.)

Можно сказать, что при данном N все целые числа разбиваются на N классов в зависимости от остатков по модулю N: два числа в одном классе сравнимы, а числа в разных классах — нет.

Задача 4.25. Докажите, что числа a^2 и b^2 дают одинаковые остатки при делении на $a - b$, если a и b — положительные целые числа, и $a > b$.

Важное свойство сравнений: чтобы узнать, в какой класс попадет сумма или произведение двух чисел, достаточно знать, в каком классе лежат слагаемые или сомножители: если одно из слагаемых (один из сомножителей) изменить на кратное N, то сумма (произведение) тоже изменится на кратное N.

В самом деле, если к одному из слагаемых прибавить kN, то к сумме тоже прибавится kN, аналогично для разности. С произведением: $(a+kN)b = ab+kbN \equiv ab \pmod N$.

Благодаря этому свойству в выражении, содержащем операции сложения и умножения (или возведение в целую степень, которое сводится к многократному умножению), можно заменять слагаемые или сомножители на сравнимые по модулю N — если результат нам важен лишь по модулю N. Например, можно найти $2^{100} \bmod 7$ (остаток от деления 2^{100} на 7): поскольку $2^3 = 8 \equiv 1 \pmod 7$, то $2^{99} \equiv (2^3)^{33} \equiv 1^{33} = 1 \pmod 7$, так что $2^{100} \equiv 2^{99} \cdot 2 \equiv 1 \cdot 2 = 2 \pmod 7$.

Задача 4.26. Какой остаток даёт число 100^{100} при делении на 99?

Задача 4.27. Найдите две последние цифры числа 99^{1000}.

Сравнения по модулю (пусть не под таким названием) часто встречаются в быту. Скажем, циферблат показывает количество прошедших часов по модулю 12, а также количество минут по модулю 60. Измеряя углы в градусах, мы фактически измеряем число градусов по модулю 360.

Задача 4.28. Как построить угол в 1°, если задан угол в 19°?

Последняя цифра (для положительного целого числа) сравнима в этим числом по модулю 10, а две последние цифры — по модулю 100. В музыке двенадцать полутонов составляют целую октаву, и потом названия нот (до, до диез и так далее) повторяются.

Известный признак делимости на 9 (число делится на 9 тогда и только тогда, когда его сумма цифр делится на 9) можно обобщить и сказать, что число и его сумма

цифр сравнимы по модулю 9. Это легко следует из наших рассуждений. Скажем, для четырёхзначного числа:

$$\overline{abcd} = 1000a + 100b + 10c + d \equiv 1a + 1b + 1c + 1d = a + b + c + d \pmod 9,$$

поскольку $10 \equiv 1 \pmod 9$ и, следовательно, 10^2, 10^3,... все сравнимы с 1 по модулю 9.

Задача 4.29. Придумайте признак делимости на 11, использующий знакопеременную сумму цифр.

Если мы хотим представить себе наглядно числа по модулю N, можно вообразить кольцевое шоссе длиной в N километров: на нём километровые столбы будут $0, 1, 2, \ldots, N-1$, далее идёт столб N, который на том же месте, где 0, затем $N+1$ на том же месте, где 1, и так далее. Можно пойти в другую сторону и поставить столб -1 на том же месте, где $N-1$: это соответствует тому, что $(N-1) \equiv (-1) \pmod N$.

Для действительных (не целых) чисел равенство дробных частей можно назвать сравнением по модулю 1. Точнее говоря, целой частью числа x называют наибольшее целое число, не превосходящее x; целую часть обозначают обычно $\lfloor x \rfloor$. Разницу $x - \lfloor x \rfloor$ называют дробной частью и иногда обозначают $\{x\}$. Числа с одинаковой дробной частью отличаются на целое число единиц; можно сказать, что они «сравнимы по модулю 1». Обратите внимание, что с отрицательными числами та же ситуация: $\lfloor -2.3 \rfloor = -3$ и $\{-2.3\} = 0.7$.

Задача 4.30. Нарисуйте на плоскости пары (x, y), для которых $\lfloor x \rfloor = \lfloor y \rfloor$. Тот же вопрос для условия $\{x\} = \{y\}$.

4.5 Таблицы сложения и умножения по модулю N

Мы уже видели таблицы сложения и умножения по модулю 2 и 3. Теперь мы знаем, что аналогичную таблицу можно составить по любому модулю. Например, по модулю 10 получатся таблицы сложения и умножения, которые получаются из обычных, если оставить только последнюю цифру:

+	0	1	2	3	4	5	6	7	8	9
0	0	1	2	3	4	5	6	7	8	9
1	1	2	3	4	5	6	7	8	9	0
2	2	3	4	5	6	7	8	9	0	1
3	3	4	5	6	7	8	9	0	1	2
4	4	5	6	7	8	9	0	1	2	3
5	5	6	7	8	9	0	1	2	3	4
6	6	7	8	9	0	1	2	3	4	5
7	7	8	9	0	1	2	3	4	5	6
8	8	9	0	1	2	3	4	5	6	7
9	9	0	1	2	3	4	5	6	7	8

×	0	1	2	3	4	5	6	7	8	9
0	0	0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7	8	9
2	0	2	4	6	8	0	2	4	6	8
3	0	3	6	9	2	5	8	1	4	7
4	0	4	8	2	6	0	4	8	2	6
5	0	5	0	5	0	5	0	5	0	5
6	0	6	2	8	4	0	6	2	8	4
7	0	7	4	1	8	5	2	9	6	3
8	0	8	6	4	2	0	8	6	4	2
9	0	9	8	7	6	5	4	3	2	1

Скажем, $7 \times 8 = 56$, поэтому в восьмой клетке седьмого ряда мы пишем $56 \bmod 10 = 6$.

Задача 4.31. Найдите последнюю ненулевую цифру числа $10! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot 10$? Тот же вопрос для числа 100!

Задача 4.32. На какие цифры могут оканчиваться точные квадраты (=квадраты целых чисел)?

Задача 4.33. В обычной алгебре уравнение $x^2 = x$ имеет только решения $x = 0$ и $x = 1$. Покажите, что по модулям 10, 100 и 1000 оно имеет ещё два решения. (Скажем, существуют два трёхзначных числа x, для которых $x^2 \equiv x \pmod{1000}$, то есть x^2 оканчивается на x.)

Имея такое «сложение» и «умножение», стоит задуматься о том, верны ли для них знакомые из школьного курса алгебры правила. Скажем, будет ли сложение коммутативно ($a + b = b + a$)? Это можно проверить по таблице: она симметрична относительно главной диагонали. Но можно и так сообразить: чтобы найти, скажем, число в седьмой клетке третьего ряда, надо сложить числа с остатками 3 и 7 и взять остаток по модулю 10. А в третьей клетке седьмого ряда мы складываем числа с остатками 7 и 3. Поскольку обычное сложение коммутативно, то получится одно и то же.

Аналогичное рассуждение показывает, что и другие обычные свойства сложения и умножения сохраняются:

- $a + (b + c) = (a + b) + c$ (ассоциативность сложения);
- $ab = ba$ (коммутативность умножения)
- $a(bc) = (ab)c$ (ассоциативность умножения)
- $a(b + c) = ab + ac$ (дистрибутивность)

Числа 0 и 1 (точнее, классы чисел, дающие остаток 0 и 1 при делении на N) обладают обычными свойствами:

- $0 + a = 0$;
- $1 \cdot a = a$;
- $0 \cdot a = 0$.

Для каждого остатка a есть противоположный, который в сумме с a равен нулю (а именно, $N - a$ при $a \neq 0$, и нуль при $a = 0$). Его естественно обозначить $-a$. Как обычно, противоположные остатки можно использовать при вычитании. В самом деле, по определению вычитания, $b - a$ это такое x, что $a + x = b$. В качестве такого x годится $b + (-a)$, потому что $(b + (-a)) + a = b + ((-a) + a) = b + 0 = b$ (пользуемся ассоциативностью, свойствами противоположного, свойствами нуля). Решение уравнения $a + x = b$ не только существует, но и единственно: прибавляя $(-a)$ к обеим частям, получаем $(-a) + (a + x) = (-a) + b$, но левая часть равна $((-a) + a) + x = 0 + x = x$.

Так что с вычитанием проблем нет и всё как обычно. А вот с делением, то есть с решением уравнений $ax = b$, всё сложнее. Решая такое уравнение, мы должны в a-ой строке найти b, и посмотреть, в каком столбце это b окажется (то есть на что надо умножить a, чтобы получить b).

Задача 4.34. Разделить 3 на 7 по модулю 10, пользуясь таблицей.

При делении 3 на 7 проблем не возникло, потому что число 3 встречается в строке 7 ровно один раз. Вот если бы мы делили 3 на 6, то ничего бы не вышло: в строке 6 таблицы числа 3 нет (там стоят только чётные числа). А при делении 4 на 6 мы не знали бы, что выбрать: и $6 \cdot 4$, и $6 \cdot 9$ кончаются на 4, так что уравнение $6x = 4$ в остатках (или, как говорят ещё, «вычетах») по модулю 10 имеет два решения.

Чтобы в этом разобраться, нам понадобится понятие обратимого элемента.

4.6 Обратимые элементы по модулю N

Остаток (вычет) по модулю N называется *обратимым*, если в произведении с каким-то другим остатком он даёт 1. Другими словами, a обратим, если уравнение $ax = 1$ имеет решение, то есть если в строке a таблицы умножения встречается единица.

Задача 4.35. По таблице умножения по модулю 10 найдите все обратимые элементы.

На обратимые элементы можно делить: *если a обратим, то уравнение $ax = b \pmod N$ имеет единственное решение при любом b.*

В самом деле, обозначим через a^{-1} элемент, который в произведении с a даёт единицу. (Пока мы не знаем, что такой только один, так что обозначим какой-то.) Положим $x = a^{-1}b$. Тогда $ax = a(a^{-1}b) = (aa^{-1})b = 1 \cdot b = b$, так что одно решение уравнения $ax = b$ мы нашли. Оно будет единственным: если $ax = b$, то $a^{-1}(ax) = a^{-1}b$, но $a^{-1}(ax) = (a^{-1}a)x = 1 \cdot x = x$, так что для x есть только одна возможность.

Остаётся выяснить, какие элементы (остатки, вычеты) обратимы по модулю N. Решив задачу в начале этого раздела, мы знаем, что по модулю 10 это будут $1, 3, 7, 9$.

Задача 4.36. При любом N легко указать как минимум два обратимых элемента. Что это за элементы?

Ответ на этот вопрос составляет первый существенный результат этого раздела.

Определение 4.1. *Взаимно простыми* называются числа, которые не имеют общего делителя, не считая 1.

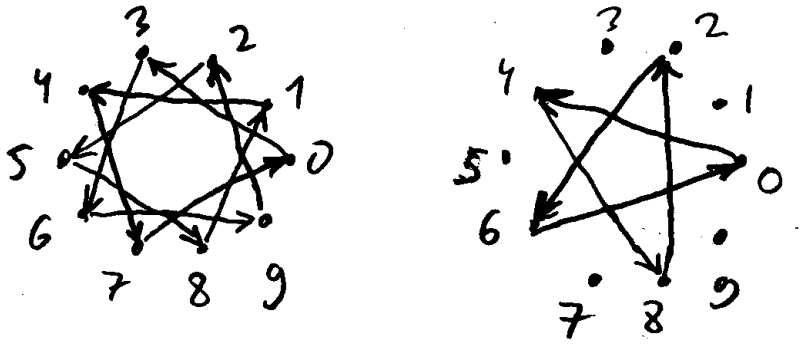
Теорема 4.1. *Обратимыми по модулю N являются те и только те остатки, которые взаимно просты с N.*

Например, по модулю 10 отпадают все чётные числа (у которых общий делитель 2), а также число 5 (общий делитель 5), остаются как раз $1, 3, 7, 9$.

В частности, если N простое (не разлагается в произведение меньших чисел), то общим делителем могут быть только 1 и N, так что все остатки, кроме нуля, взаимно просты с N и обратимы. Для простого модуля всё как в обычной алгебре: делить можно на всё кроме нуля. (Математики выражают это словами: вычеты по простому модулю p образуют *поле*.)

Прежде чем доказывать эту теорему, представим себе наглядно, что означает обратимость остатка a. Для этого вспомним, что остатки по модулю N можно расположить на круговом шоссе длины N, как автобусные остановки. Запустим маршрут автобуса, которые делает остановки каждые a километров: у него первая остановка будет в a, вторая в $a + a = 2a$, третья в $3a$ (всё по модулю N, естественно). Обратимость означает, что таким странным способом все остановки будут обслужены (в строке таблицы умножения встретятся все остатки).

Вот две картинки, показывающие, что по модулю 10 остаток 3 будет обратимым, а остаток 4 нет:



в первом случае мы проходим все остановки, а во втором не все (только чётные).

Кстати, из этой картинки хорошо видно, что если элемент обратим (мы попадаем в соседнюю остановку), то деление всегда возможно (мы попадём во все остановки). В самом деле, если через k шагов мы попали в соседнюю, то ещё через k мы попадём в следующую и так далее, пройдя через все остановки.

Теперь легко доказать простую часть утверждения: если N и шаг a имеют общий делитель d, то элемент a необратим (мы не попадём в соседнюю остановку). В самом деле, будем отмечать остановки через $d, 2d, 3d$ и так далее от начальной. Поскольку N делится на d, то мы дойдём до N-й (то есть начальной) остановки, пройдя весь круг. Если a кратно d, то a-автобус будет останавливаться только в выделенных остановках, и в соседнюю никогда не попадёт.

Осталось доказать сложную часть утверждения: если N и шаг a не имеют общих делителей, то все остановки будут обслужены. Тут полезно вспомнить о перестановках и отвечающих им ориентированных графах, см. раздел **??**.

В нашем случае вершины графа расположены по кругу, как вершины правильного N-угольника. Стрелки (рёбра графа) соответствуют движению автобуса, проезжающего a перегонов до следующей остановки. Другими словами, из вершины x (остатка по модулю N) стрелка ведёт в вершину $x + a$. По построению из каждой вершины выходит только одна стрелка, и входит тоже только одна, из вершины $x-a$ (однозначность вычитания). Нам надо доказать, что если a взаимно просто с N, то есть только один цикл, включающий все вершины.

Пусть цикл, начатый из вершины 0, включает только часть вершин. Посмотрим, в какие вершины он попадает. Пусть ближайшая из них (по кругу) имеет номер d. Ясно, что построение цикла можно начать с любой вершины (круг везде одинаков), поэтому если мы начнём с d, то следующая обслуженная вершина будет $2d$. Значит, в цикл входят вершины 0, потом (по кругу) d, потом $2d$ и так далее, пока мы не вернёмся обратно в начальную вершину 0. В результате мы пройдём полный круг в N вершин, двигаясь шагами по d, поэтому N кратно d. С другой стороны, вершина a обслужена (на первой же остановке автобуса), поэтому и a тоже кратно d. Получается, что d — общий делитель N и a. А мы предположили, что они взаимно просты, то есть $d = 1$, что означает, что все вершины (все остатки по модулю N)

попадают в один цикл.

Задача 4.37. Покажите, что построенное в этом рассуждении число d делится на любой общий делитель чисел a и N и является их наибольшим общим делителем.

Последняя задача показывает, что уравнение $ax = b \pmod N$ разрешимо тогда и только тогда, когда b делится на НОД(a, N) (наибольший общий делитель чисел a и N).

Задача 4.38. Покажите, что в общем случае граф отображения $x \mapsto x + a$ состоит из НОД(a, N) циклов одинаковой длины $N/\,$НОД(a, N).

Если наше доказательство с циклами и многоугольниками, вписанными в окружность, кажется непонятным, ничего страшного: другое доказательство будет приведено в следующем разделе.

4.7 Обратимые элементы и диофантовы уравнения

Мы только что доказали, что всякий остаток a, взаимно простой с N, имеет обратный. Но как этот обратный найти? Можно пробовать все возможности по очереди. При небольших N это реально, но что делать, если, скажем $N = 5704689200685129054721$ (хотите верьте, хотите проверьте, но это простое число, один из делителей числа $2^{128} + 1$). Проверить столько вариантов даже и с помощью какого-нибудь суперкомпьютера не удастся. Нет ли какого-то более быстрого способа? (Научно говоря, нет ли полиномиального алгоритма — время работы которого было бы ограничено многочленом от числа цифр в N?)

Оказывается, что такой алгоритм есть, и по существу он был известен ещё Евклиду в древней Греции. Для начала переформулируем нашу задачу более симметричным образом. Мы хотим решить сравнение $ax \equiv b \pmod N$. Это сравнение означает, что разность $b-ax$ должна быть целым числом, кратным N, то есть должна равняться Ny при каком-то y. Таким образом, нам надо решить в целых числах уравнение $ax + Ny = b$. Здесь a, N, y — известные целые числа, и мы ищем целые x и y, при которых левая часть равна правой. Найдя их, мы можем про y забыть, а x будет решением.

Уравнения, в которых нам нужны целые решения, называются *диофантовыми*, в честь другого древнего грека — Диофанта. (Он, правда, не такой древний, как Евклид, и жил уже после Р.Х.)

Таким образом, теорема об обратных элементах сводится к такому утверждению: *уравнение $ax + Ny = b$ имеет решение в целых числах тогда и только тогда, когда b кратно НОД(a, N)*. Из него следует, что при взаимно простых a и N это уравнение имеет решение при любом b.

Чтобы сделать запись ещё более симметричной и забыть о неравноправии a и N в исходной постановке, заменим букву N на b, а b на c. Таким образом, нам надо доказать такое утверждение:

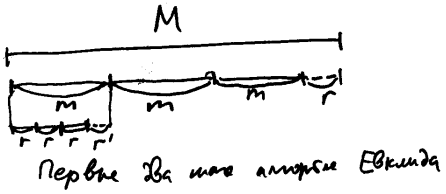
Теорема 4.2. *Пусть a, b, c — произвольные целые числа. Уравнение $ax+by = c$ имеет целочисленнное решение тогда и только тогда, когда число c кратно НОД(a, b).*

Опять в одну сторону это очевидно: если d есть (наибольший) общий делитель a и b, то $ax + by$ всегда делится на d, так что уравнение может иметь решение только при c, кратном d. Интересно обратное утверждение, и мы его докажем с помощью алгоритма Евклида, заодно показав, как можно искать это самое решение на практике.

4.8 Алгоритм Евклида

Давайте сначала посмотрим на этот алгоритм в том виде, как это было у Евклида в его учебнике геометрии («Начала»). Пусть нам надо найти общую меру двух отрезков, то есть третий отрезок, который укладывается целое число раз в первом и во втором. (Другими словами, мы хотим найти меру длины, при которой длины обоих данных нам отрезков будут целыми числами.)

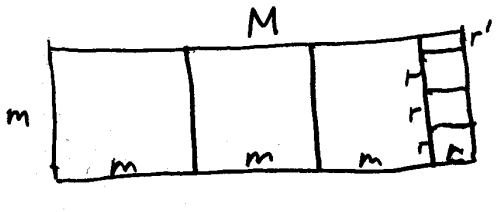
Евклид предлагает делать это так: будем откладывать меньший отрезок m внутри большего M. Если нам повезёт и он уложится целое число раз, то меньший отрезок m и будет общей мерой. Если нет, то он уложится сколько-то раз и что-то (уже меньшее m) останется. Обозначим этот остаток за r. Теперь повторим эту процедуру с отрезками m и r, укладывая меньший из них (то есть r) в большем. Снова либо он уложится без остатка, либо получится остаток r', меньший r, и мы применяем алгоритм к r и r', и так далее.



Первые два шага алгоритма Евклида

Алгоритм заканчивает свою работу, когда и если меньший отрезок укладывается в большем без остатка.

Объясняя эту процедуру школьникам, часто считают отрезки сторонами прямоугольников: сначала есть прямоугольник $M \times m$, от которого отрезают квадраты $m \times m$, пока это возможно. Когда останется прямоугольник $r \times m$, в котором $r < m$, от него отрезают квадраты $r \times r$, остаётся прямоугольник $r \times r'$ с $r' < r$. Можно считать, что у нас есть автомат, который отрезает от прямоугольника квадрат со стороной, равной меньшей стороне прямоугольника (он сам разбирается, какая сторона меньше). Мы кромсаем прямоугольник на квадратные части, засовывая остаток снова и снова в этот автомат.



По существу это, конечно, тот же самый процесс, может быть, в немного более наглядной форме.

Основное свойство алгоритма Евклида теперь можно сформулировать так:

Теорема 4.3. *Если исходные отрезки имеют общую меру, то алгоритм заканчивает работу и последний отрезок (тот, что уложится целое число раз) будет наибольшей их общей мерой. Если же исходные отрезки не имеют общей меры, то алгоритм никогда не остановится.*

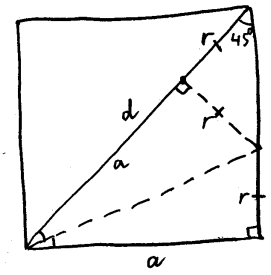
Доказательство. (1) Если исходные отрезки имеют общую меру d, примем её за единицу измерения. Тогда оба наших отрезка имеют целую длину, и мы делим одно целое число (M) на другое (m) с остатком r, потом делим m на r с остатком r' и так дальше, при этом $m > r > r' > \ldots$ (остаток меньше делителя) и все они целые, так что бесконечно это продолжаться не может. В терминах прямоугольников: нарисуем всё по клеточкам на клетчатой бумаге с шагом d, тогда и все разрезы пройдут по клеточкам, и возможных мест разреза конечное число.

(2) Теперь покажем, что если алгоритм заканчивает работу, то последний отрезок будет общей мерой. В самом деле, если принять его за единицу измерения, то предыдущий отрезок будет целым числом (ведь последний укладывается целое число раз), перед ним тоже будет целый отрезок (равный сумме нескольких целых чисел) и пр. В терминах квадратов: если принять самый маленький квадрат при разрезании за клетку на клетчатой бумаге, то все большие квадраты и составленные их них прямоугольники пойдут по линиям сетки, и их стороны будет целыми, то есть сторона маленького квадрата будет общей мерой. (Отсюда уже следует последнее утверждение: если общей меры нет, то алгоритм не заканчивает работу.)

(3) Осталось показать, что последний отрезок будет *наибольшей* общей мерой. Более того, он даже будет кратен любой другой общей мере d. Почему? Приняв d за единицу измерения, мы видим, что оба наших отрезка имеют целую длину, и все последующие отрезки будут целыми. Значит, ответ алгоритма Евклида тоже целый (кратен d), как мы и утверждали.

Алгоритм Евклида можно использовать, чтобы показать, что сторона квадрата и его диагональ не имеют общей меры. На рисунке сторона бумажного квадрата загнута по биссектрисе так, чтобы она пошла по диагонали. Видно, что сторона a укладывается в диагонали d один раз и остаётся отрезок r. Этот же отрезок можно найти ещё в двух местах, и если его один раз (уже на следующем шаге алгоритма Евклида) отложить на стороне, то останется как раз диагональ квадрата со стороной r. Таким образом, через два шага мы приходим к тому же соотношению между

диагональю квадрата и его стороной, и всё повторяется, так что алгоритм никогда не закончит работу.



Задача 4.39. Посмотрите внимательно на картинку и восстановите пропущенные геометрические детали.

Задача 4.40. Каково должно быть отношение сторон прямоугольника, чтобы после отрезания от него одного квадрата получался прямоугольник, подобный исходному? Покажите, что стороны такого прямоугольника несоизмеримы (=не имеют общей меры).

Это отношение называют *золотым сечением*; говорят, что оно встречается в красивых постройках и картинах.

Задача 4.41. Найдите целые положительные числа x, y, z, t, при которых

$$\frac{17}{10} = x + \cfrac{1}{y + \cfrac{1}{z + \cfrac{1}{t}}}$$

При чём тут алгоритм Евклида?

4.9 Алгоритм Евклида и диофантовы уравнения

Весь этот экскурс в историю нужен нам не сам по себе, а чтобы научиться решать уравнения $ax + by = c$ при c, кратном НОД(a, b). Ясно, что достаточно уметь это делать при $c = \text{НОД}(a, b)$, потом можно умножить на отношение $c/\,\text{НОД}(a, b)$, раз оно целое.

Мы можем найти $d = \text{НОД}(a, b)$, разрезая на квадраты прямоугольник $a \times b$, это будет сторона наименьшего из квадратов. (Про квадраты мы говорим только для наглядности, можно было говорить про отрезки.) Ключевое наблюдение: *все отрезки, появляющиеся в ходе алгоритма, представляются в виде* $ax + by$ *с некоторыми целыми* a *и* b. Как говорят, они являются «целочисленными линейными комбинациями» a и b. Это же относится и к последнему отрезку, то есть d, и мы получаем искомое решение.

Почему они будут целочисленными линейными комбинациями? Пусть мы сначала делим a на b с остатком r, тогда $a = bq + r$, и $r = a - bq$ представлен такой комбинацией. Теперь мы делим b на r, получаем остаток r', то есть $b = q'r + r'$, и $r' = b - q'r$ есть целочисленная комбинация b и r. Вспомним, что само r есть комбинация a и b, получится

$$r' = b - q'r = b - q'(a - bq) = b - q'a + q'bq = (qq' + 1)b - q'a,$$

то есть r' тоже есть целочисленная комбинация a и b, и так далее.

Задача 4.42. Какая целочисленная комбинация получается для r' на приведённой выше картинке?

Для программистов всё сказанное можно заменить алгоритмом:

```
{a,b - целые положительные числа}
делим a на b, получаем частное q и остаток r
M:=b; m:=r;
Ma:=0; Mb:=1; ma:=1; mb:=-q;
{M>m>=0; НОД(M,m)=НОД(a,b); M=Ma*a+Mb*b; m=ma*a+mb*b}
пока m>0:
    делим M на m, получаем частное q и остаток r
    M, Ma, Mb, m, ma, mb <- m, ma, mb, r, Ma-q*ma,Mb-q*mb
{m=0; НОД(a,b)=НОД(M,m)=НОД(M,0)=M=Ma*a+Mb*b}
ответ: M=НОД(a,b); (x,y)=(Ma,Mb) - решение ax+by=НОД(a,b)
```

Здесь стрелка `<-` означает одновременное присваивание: шесть переменных получают новые значения, указанные справа, при этом при вычислении выражений в правой части используются старые значения. Если, как это принято, выполнять присваивания последовательно, то надо написать

```
M_new:=m; Ma_new:=ma; Mb_new:=mb
m_new:=r; ma_new:=Ma-q*ma; mb_new:=Mb-q*mb
M:=M_new; Ma:=Ma_new; Mb:=Mb_new
m:=m_new; ma:=ma_new; mb:=mb_new
```

Этот алгоритм иногда называют «расширенным алгоритмом Евклида» (по-английски “extended Euclidean algorithm”). Что можно сказать о количестве операций в этом алгоритме? Можно заметить, что за два шага алгоритма Евклида больший отрезок (большее число для алгоритма с целыми числами) уменьшается по крайней мере вдвое. В самом деле, если меньший отрезок не превосходит половины большего, то это случится уже за один шаг; если же нет, то на первом шаге частное равно 1, а остаток не больше половины большего отрезка, так что за два шага это всё равно случится. Поэтому число шагов для n-битовых чисел равно $O(n)$. Каждый шаг требует деления с остатком — если это делать «уголком», как в школе, то число действий будет $O(n^2)$, так что всего получается $O(n^3)$. Конечно, если следовать

определению буквально и делить с помощью последовательного вычитания (как мы делали, доказывая существование частного и остатка), то будет плохо (число действий будет экспоненциальным).

Задача 4.43. Напишите алгоритм решения уравнения $ax + by = \text{НОД}(a, b)$, использующий три соотношения:

- $\text{НОД}(2a, 2b) = 2\,\text{НОД}(a, b)$;
- $\text{НОД}(2a, b) = \text{НОД}(a, b)$ при нечётном b;
- $\text{НОД}(a, b) = \text{НОД}(a - b, b)$, которое верно всегда, но нужно при нечётных a и b и $a \geq b$.

В каждом случае задачу решения уравнения можно рекурсивно свести к той же задаче для новой пары. Сколько действий требует этот вариант алгоритма Евклида?

Задача 4.44. Известно, что алгоритм Евклида для пары целых положительных чисел (a, b) требует n шагов. Каково наименьшее возможное значение a? (Считаем, что $a \geq b$.)

Если не интересоваться алгоритмической стороной дела, то можно доказать теорему о разрешимости уравнений $ax + by = c$ следующим образом. Напомним, что мы называли *идеалом* множество целых чисел, которое вместе с любыми двумя элементами содержит их сумму и разность, и вместе с любым элементом содержит все его кратные. В качестве примера идеала мы приводили множество I_c всех кратных некоторого числа c. Такие идеалы алгебраисты называют *главными*. Оказывается, что (для целых чисел) других и нет. В самом деле, пусть I — некоторый идеал. Если он состоит только из нуля, то $c = 0$. Если в нём есть ненулевые числа, возьмём минимальное по модулю число c. Будем считать, что оно положительно (перейдя к $-c$, если надо). Теперь ясно, что все числа в I кратны c: если какое-то $x \in I$ даёт ненулевой остаток r при делении на c, то $x = qc + r$ и $r = x - qc$ есть разность двух чисел из I и потому тоже принадлежит I, что противоречит минимальности c. Такое число c, при котором $I = I_c$, называется *образующей* идеала I.

Пусть теперь a, b — произвольные целые числа, а I состоит из всех целых чисел, которые можно представить в виде $ax + by$. Это будет идеал, и по доказанному он совпадает с некоторым I_c. Поскольку $a, b \in I$, то они кратны c, так что c будет общим делителем. С другой стороны, любой общий делитель d чисел a, b делит все элементы I, в том числе и c, так что c будет наибольшим общим делителем (и кратен любому делителю — это мы тоже доказали заодно). Наконец, c представим в виде $ax + by$ по построению I.

Задача 4.45. Покажите, что общие кратные двух чисел a и b образуют идеал, и выведите отсюда, что наименьшее общее кратное $\text{НОК}(a, b)$ является делителем любого общего кратного.

Задача 4.46. Покажите, что $\text{НОД}(a, b) \cdot \text{НОК}(a, b) = ab$.

Задача 4.47. Покажите, что для данных целых a, b числа x, при которых ax кратно b, образуют идеал. Как найти его образующую, зная a, b и НОД(a, b)?

Задача 4.48. При каких целых a, b, c, d уравнение $ax+by+cz = d$ разрешимо в целых числах?

4.10 Однозначность разложения на множители

Теперь мы можем доказать теорему об однозначности разложения целых чисел на простые множители, которую иногда торжественно называют «основной теоремой арифметики». Напомним, что целое число $p > 1$ называется *простым*, если оно не разлагается в произведение меньших чисел (то есть не имеет делителей, кроме 1 и p).

Теорема 4.4. *Всякое целое положительное число, большее* 1*, разлагается на простые множители, причём единственным образом: любые два разложения отличаются только перестановкой сомножителей.*

(Можно сказать, что единица тоже разлагается в произведение нуля простых множителей — если принять, что произведение нуля сомножителей равно 1. Как вы думаете, кстати, что следует считать суммой нуля слагаемых?)

Трудность с этой теоремой в том, что она (видимо, со школы) кажется само собой разумеющейся, и непонятно, что тут доказывать. Тем не менее доказательство требует некоторых усилий (которые мы по большей части уже проделали).

Существование разложения совсем просто. Если данное число N простое, то получилось разложение из одного сомножителя. Если нет, то $N = ab$ для каких-то меньших a, b. Если a и b простые, то хорошо, если нет, то разложим их в произведение меньших и так далее до тех пор, пока дальше уже ничего не раскладывается, поскольку числа простые. (Формально говоря, мы рассуждаем по индукции и считаем, что для меньших чисел a и b существование разложения уже известно.)

Единственность разложения. Пусть некоторое число N имеет два разложения

$$N = p_1 \cdot p_2 \cdot \ldots \cdot p_n = q_1 \cdot q_2 \cdot \ldots \cdot q_m$$

(они могут отличаться и числом сомножителей, m не обязано равняться n). Мы хотим получить противоречие. Сократим на общие сомножители (если они есть). Если сократится не всё, то получим два разложения одного числа, не имеющих общих сомножителей

$$p_1 \cdot p_2 \cdot \ldots \cdot p_n = q_1 \cdot q_2 \cdot \ldots \cdot q_m.$$

Как говорят, «без ограничения общности» можно предположить, что общих сомножителей нет (сократив на них, если есть).

В чём тут противоречие? С одной стороны, левая часть делится на p_1 (можно было бы взять любой другой p_i, если там несколько сомножителей). А правая часть равна произведению чисел, ни одно из которых не делится на p_1: они ведь простые и p_1 среди них по предположению нет. Осталось доказать, что такого не бывает, то есть доказать следующую лемму.

Лемма 4.5. *Если p — простое число, то произведение чисел, не делящихся на p, не может делиться на p.*

По существу мы уже это доказали: в терминах вычетов по модулю p нужно доказать, что произведение ненулевых вычетов не равно нулю. А мы знаем, что если $a \neq 0 \pmod p$, то a взаимно просто с p, поэтому a обратим и уравнение $ax = 0$ имеет единственное решение (нулевое).

Можно повторить эти рассуждения более подробно. Во-первых, достаточно доказать лемму для двух сомножителей. Если есть, скажем, три числа a, b, c, не делящиеся на p, то мы применяем лемму для двух сомножителей a и b и заключаем, что ab не делится на p. После этого уже можно применить лемму к двум сомножителями ab и c и заключить, что $(ab)c$ не делится на p. (Аналогично для любого числа сомножителей — формально говоря, мы используем индукцию по числу сомножителей.)

Для двух сомножителей мы можем доказать более общий факт:

Лемма 4.6. *Если ab делится на n, и при этом a взаимно просто с n, то b делится на n.*

(Более общий он потому, что если a не делится на простое p, то a взаимно просто с p — других общих делителей быть не может.)

Доказательство. Если $\text{НОД}(a, n) = 1$, можно найти x, y, для которых $ax+ny = 1$. Умножим это равенство на b, получим, что

$$b = abx + ny.$$

Осталось заметить, что оба слагаемых в правой части делятся на n: по предположению ab делится на n, и очевидным образом ny делится на n. Значит, и сумма (то есть b) делится на n, что и требовалось доказать.

Задача 4.49. Как вычислить $\text{НОД}(a, b)$ и $\text{НОК}(a, b)$, зная разложения a и b на простые множители? Докажите, что $\text{НОД}(a, b) \cdot \text{НОК}(a, b) = ab$, используя (очевидное) равенство $\min(x, y) + \max(x, y) = x + y$. Можно ли использовать эту задачу, чтобы найти более эффективный алгоритм поиска наибольшего общего делителя?

Однозначность разложения на множители делает очевидными многие утверждения, например, такое: *если число a делится на b и на c, и при этом b и c взаимно просты, то a делится на bc.* В самом деле, если a делится на b, то разложение a на множители содержит все множители из разложения b (и ещё что-то), поскольку можно разложить по отдельности b и a/b. Аналогичным образом разложение a содержит все множители из разложения c. Но общих множителей у b и c нет, так что это разные множители, и в разложении a можно выделить разложение для bc.

Можно, конечно, обойтись и без теоремы об единственности разложения: поскольку b и c взаимно просты, то $bx + cy = 1$ при некоторых x и y, так что $a = a(bx + cy) = abx + acy$, и оба слагаемых abx и acy делятся на bc (почему?) — но это всё же выглядит более искусственно...

Задача 4.50. Докажите, что если a взаимно просто с b и с c по отдельности, то оно взаимно просто с bc. (Можно воспользоваться разложением на множители, а можно и перемножить равенства $ax + by = 1$ и $au + cv = 1$.)

4.11 Китайская теорема об остатках

Как найти число, которое даёт остаток 1 при делении на 23 и остаток 1 при делении на 37? Понятно как: надо взять число, делящееся и на 23, и на 37, например, $23 \cdot 37$, и прибавить единицу.

Немного сложнее найти число, которое даёт остаток 22 при делении на 23 и 36 при делении на 37. Надо заметить, что если прибавить 1, то получится число, делящееся и на 23, и на 37, так что можно попробовать $23 \cdot 37 - 1$, и всё получится.

А что делать, если надо найти число, дающее (скажем) остаток 17 при делении на 23, и одновременно остаток 24 при делении на 37? И вообще, возможно ли это? Оказывается, что возможно, причём для любой комбинации остатков, поскольку 23 и 37 взаимно просты. Это гарантирует следующая теорема, традиционно называемая «китайской теоремой об остатках» (Chinese remainder theorem).

Теорема 4.7. *Пусть числа m и n взаимно просты, и пусть u и v — любые целые числа. Тогда можно найти число x, для которого $x \equiv u \pmod m$ и одновременно $x \equiv v \pmod n$.*

Прежде чем доказывать это, посмотрим на какой-нибудь пример. Числа 23 и 37 слишком большие, возьмём, скажем, 3 и 4 и составим таблицу:

x	0	1	2	3	4	5	6	7	8	9	10	11	
$x \bmod 3$	0	1	2	0	1	2	0	1	2	0	1	2	
$x \bmod 4$	0	1	2	3	0	1	2	3	0	1	2	3	

(Дальше можно уже не продолжать, поскольку 12 делится и на 3, и на 4, и всё повторится с начала.)

Так вот, китайская теорема учит, что в двух нижних строках встретятся все возможные комбинации остатков: любой из остатков $0, 1, 2$ комбинируется с любым из остатков $0, 1, 2, 3$ (всего как раз 12 вариантов: три варианта mod 3 комбинируются с четырьмя вариантами mod 4, так что каждая комбинация встречается ровно по одному разу.)

Эта таблица подсказывает доказательство китайской теоремы об остатках. Построим аналогичную таблицу, записав в первой строке числа $0, 1, 2, \ldots, mn - 1$, во второй строке их остатки при делении на m (получится $0, 1, 2, \ldots, m-1$, повторённое n раз), а в третьей строке их остатки при делении на n (получится $0, 1, 2, \ldots, n - 1$, повторённое m раз). Покажем, что все комбинации остатков (их будет mn) встретятся ровно по одному разу. Для этого заметим, что никакая комбинация не может повториться дважды: если числа u и v дают одинаковые остатки и при делении на m, и при делении на n, то их разность $u - v$ делится и на m, и на n. Как мы уже

видели, это значит, что $u-v$ кратно mn (поскольку m и n взаимно просты), а у нас все остатки при делении на mn представлены по одному разу.

Осталось заметить, что если ни одна из mn комбинаций не повторяется в имеющихся mn столбцах, то придётся использовать все mn комбинаций. (Если N голубей разместить в N норах, причём в каждой норе не больше одного голубя, то все норы будут заполнены. Это называется по-английски pigeon-hole principle, а по-русски принципом Дирихле, поскольку Дирихле использовал это соображение, изучая приближения действительных чисел рациональными.)

Задача 4.51. Докажите, что если числа m, n, k попарно взаимно просты, то для любых u, v, w найдётся такое x, что

$$x \equiv u \pmod m; \quad x \equiv v \pmod n; \quad x \equiv w \pmod k.$$

Приведённое нами доказательство китайской теоремы об остатках не даёт алгоритма нахождения искомого x. Но легко понять, что эта задача сводится к уже известным. В самом деле, нам нужно найти x, для которого $x \equiv u \pmod m$ и $x \equiv v \pmod n$. Первое условие означает, что $x = u + tm$, и надо искать t. Второе условие, переписанное в терминах t, означает, что $u + tm \equiv v \pmod n$, то есть надо решить сравнение $tm \equiv -u \pmod n$, что мы уже умеем делать с помощью алгоритма Евклида (надо найти обратный для m в вычетах по модулю m, используя взаимную простоту m и n).

Заодно мы получили другое доказательство китайской теоремы об остатках.

4.12 Малая теорема Ферма

Этот раздел называется «малая теорема Ферма», хотя исторически это не вполне справедливо: как и в случае «большой», или «последней» теоремы Ферма (о том, что уравнение $x^n + y^n = z^n$ при $n > 2$ не имеет решений в целых положительных числах), сам Ферма не предложил доказательства — то ему полей не хватало (с большой теоремой Ферма), то он боялся надоесть адресату (je vous envoierois la démonstration, si je n'appréhendois d'être trop long). Но в этом случае не понадобилось ждать до конца XX века, доказательство вроде бы знал Лейбниц в том же XVII веке, а опубликовано оно было Эйлером в 1736 году, и вполне можно допустить, что это доказательство было известно Ферма.

Вот о каком утверждении идёт речь:

Теорема 4.8. *Если p — простое число, то*

$$a^{p-1} \equiv 1 \pmod p$$

при любом a, не делящемся на p.

Вот одно из доказательств (пожалуй, наиболее естественное). Рассмотрим все ненулевые остатки по модулю p. Их всего $p-1$: от единицы до $p-1$. Сделаем их

вершинами графа, проведя стрелки из x в ax (умножение по модулю p). Разница с тем графом, что у нас уже был, двоякая: во-первых, мы рассматриваем не все p остатков, а только $p-1$ ненулевых остатков. Но главное, мы не прибавляем a, а умножаем на a: стрелка из x ведёт не в $x+a$, а в ax. Так что красивых картинок с правильными звёздчатыми многоугольниками, как раньше, не получится.

Но тем не менее из каждой вершины выходит одна стрелка (по построению) и в каждую вершину входит одна стрелка (поскольку a не делится на p, оно обратимо и уравнение $ax=b$ имеет единственное решение при любом b, то есть в b входит ровно одна стрелка). Значит, граф, как и раньше, разбивается, на циклы.

Что это за циклы? Начнём с какой-то вершины, например, с остатка 1, и будем идти по стрелкам, пока цикл не замкнётся:

$$1 \mapsto a \mapsto a^2 \mapsto a^{k-1} \mapsto a^k = 1$$

Здесь k — минимальная степень a, равная 1 (как мы знаем, цикл может замкнуться, только придя в начальную вершину). Если начать построение с какой-то вершины b вместо 1, то получится цикл того же размера:

$$b \mapsto ab \mapsto a^2b \mapsto a^{k-1}b \mapsto a^kb = 1b = b.$$

Почему? Ясно, что если $a^k=1$, то и $a^kb=b$, так что после k шагов цикл замкнётся. Но надо понять, почему этот цикл не замкнётся раньше. В самом деле, если $a^lb=b$, то можно домножить справа на b^{-1} (ведь остаток b тоже обратим), и получится $a^l=1$, а мы предположили, что k минимальное.

Таким образом, $p-1$ ненулевых остатков разбиваются на циклы одинакового размера k, так что $p-1$ делится на k, то есть $p-1=km$ при каком-то m. Тогда

$$a^{p-1} = a^{km} = (a^k)^m = 1^m = 1 \pmod p,$$

что и требовалось доказать.

Есть и другие доказательства, приведём два любопытных.

Достаточно доказать, что $a^p \equiv a \pmod p$, потом можно сократить на обратимый элемент a. Представим себе, что имеется a цветов, и ими нужно раскрасить круг из p равных секторов (каждый сектор в какой-то цвет), причём раскраски, отличающиеся лишь поворотом круга, считаются за одну. Сколькими способами это можно сделать? Можно все сектора красить в один цвет, это можно сделать a способами (по числу цветов). А можно использовать более одного цвета, на это остаётся a^p-a способов, но они делятся на группы по p, отличающихся поворотами (каждую неоднородную раскраску можно повернуть p различными способами). Получается ответ $a+\frac{a^p-a}{p}$. Но число способов должно быть целым, так что a^p-a делится на p.

Второе доказательство использует отображение остатков по модулю p в себя, заданное формулой $f(x)=x^p$. Нам надо доказать, что оно тождественное, то есть что $f(x)=x$ при всех x. Это вытекает из трёх его свойств:

- $f(0)=0$;

- $f(1) = 1$;
- $f(x + y) = f(x) + f(y)$.

В самом деле, третье свойство можно по индукции распространить на любое число слагаемых, скажем,

$$f(x + y + z) = f((x + y) + z) = f(x + y) + f(z) = f(x) + f(y) + f(z)$$

и так далее, поэтому

$$f(x) = f(1 + 1 + \ldots + 1) = f(1) + f(1) + \ldots + f(1) = 1 + 1 + \ldots + 1 = x$$

(многоточия обозначают суммы из x слагаемых). Так что достаточно доказать эти три свойства. Первые два очевидны, а третье следует из бинома Ньютона: в разложении

$$(x + y)^p = x^p + \frac{p!}{1!(p-1)!}x^{p-1}y^p + \frac{p!}{2!(p-2)!}x^{p-2}y^2 + \ldots + \frac{p!}{(p-1)!1!}x^1y^{p-1} + y^p$$

все биномиальные коэффициенты, кроме двух крайних, делятся на p (у них есть p в числителе, и это простое число, которое не может сократиться с меньшими числами в знаменателе), так что по модулю p остаются только два крайних члена, что и требовалось.

Алгебраисты сказали бы, что эти три свойства верны в любом поле «характеристики p», где сумма p единиц равна нулю — их можно ещё дополнить четвёртым свойством $f(xy) = f(x)f(y)$. Но отображение f уже не будет (вообще говоря) тождественным; его называют «автоморфизмом Фробениуса» (слово «автоморфизм» означает, что выполнены эти четыре свойства, а Фробениус — немецкий математик, в честь которого он назван).

С точки зрения практиков, теорема Ферма может использоваться как способ убедиться, что число не простое. Скажем, можно вычислить $2^8 = 256 \equiv 4 \pmod 9$ и заключить отсюда, что число 9 не простое (если бы мы этого и так не знали). Это выглядит глупо, но на самом деле имеет вычислительный смысл: возвести какое-то a в степень $p - 1$ по модулю p не так сложно (надо вычислять $a, a^2, a^4, a^8, \ldots \bmod p$ последовательными возведениями в квадрат, а потом перемножить нужные из них, следуя двоичному разложению для $p - 1$), это требует полиномиального числа действий (от n для n-битового числа p), и если нам повезёт и получится не 1 по модулю p, то мы сможем убедиться, что число p составное. Можно назвать число a, для которого $a^{p-1} \not\equiv 1 \pmod p$, «свидетелем» того, что p составное; как мы обсудили, если такой свидетель известен, «проверить его показания» можно быстро.

Собственно говоря, проверка разложения на множители (перемножение сомножителей) тоже полиномиальна, но преимущество нового способа в том, что свидетелей обычно бывает много, и есть шанс наткнуться на них, взяв наугад число от 1 до $p - 1$. (А на разложение на множители так просто не наткнуться; полиномиального алгоритма разложения на множители не известно, и для чисел из нескольких тысяч

цифр искать разложение на множители никто не умеет.) Так что если вам нужно быстро найти какое-нибудь большое простое число, можно выбрать случайную последовательность цифр и проверить теорему Ферма для (скажем) $a = 2, 3, 5$; если она не выполнена, число не простое и надо взять другую случайную последовательность цифр, пока этот тест не будет пройден. С большой вероятностью этот способ даёт простые числа.

Задача 4.52. Мы предлагали проверять теорему Ферма для $2, 3, 5$, но почему-то пропустили 4. Как вы думаете, почему?

4.13 Функция Эйлера и теорема Эйлера

На самом деле приведённое доказательство малой теоремы Ферма имеет смысл не только для простых модулей. Пусть N — какое-то число, не обязательно простое. Мы можем по-прежнему рассмотреть остатки по модулю N, но взять только взаимно простые с N. (В алгебре это называют «мультипликативной группой кольца вычетов по модулю N».) Произведение двух таких остатков тоже взаимно просто с N, и все они обратимы, так что взяв какой-то остаток a среди них, мы можем построить ориентированный граф $x \mapsto xa$, и в нём из каждой вершины будет исходить одна стрелка и в каждую вершину будет входить одна стрелка. Этот граф разбивается на циклы, и по тем же причинам все циклы будут одинаковой длины. Эта длина видна на примере цикла, начинающегося с 1:

$$1 \mapsto a \mapsto a^2 \mapsto \ldots \mapsto a^k = 1,$$

где k — наименьшая степень a, равная 1 по модулю N. Значит, k является делителем числа остатков, взаимно простых с N, и мы получаем такой результат, называемый «теоремой Эйлера».

Теорема 4.9. *Пусть $N > 1$ — произвольное целое число, а $\varphi(N)$ — количество остатков среди $0, 1, \ldots, N-1$, взаимно простых с N. Пусть a — один из этих остатков. Тогда*

$$a^{\varphi(N)} \equiv 1 \pmod{N}.$$

Определённую таким образом функцию φ называют *функцией Эйлера* и традиционно обозначают буквой φ. Если число N простое, то $\varphi(N) = N - 1$, и теорема Эйлера превращается в малую теорему Ферма.

Как вычислить функцию Эйлера $\varphi(N)$, зная N? Это несложно, если мы дополнительно знаем разложение числа $N - 1$ на простые множители (как мы уже говорили, разложение на множители вычислительно сложно), и делается с помощью таких свойств:

- $\varphi(p^n) = p^n(1 - 1/p)$ для простого p;
- $\varphi(uv) = \varphi(u)\varphi(v)$, если u и v взаимно просты.

Первое свойство позволяет найти φ для степеней простых чисел, а второе свойство позволяет собрать из них любое число. Свойства эти нам по существу уже известны. Первое говорит, что доля чисел, не взаимно простых с p^n (то есть делящихся на p) составляет $1/p$ среди всех чисел, и их надо вычеркнуть. Второе легко доказать, сославшить на китайскую теорему об остатках и на то, что взаимная простота с uv означает одновременную взаимную простоту с u и v. В терминах теории вероятностей можно сказать, что определённые на вычетах по модулю uv функции «остаток по модулю u» и «остаток по модулю v» независимы и имеют равномерное распределение, и вероятность события «быть взаимно простым с uv» можно вычислять как произведение вероятностей событий «быть взаимно простым с u» и «быть взаимно простым с v».

Второе свойство называют *мультипликативностью* функции Эйлера. В теории чисел встречаются разные мультипликативные функции.

Задача 4.53. Покажите, что функция $d(n)$, равная числу положительных делителей числа n, мультипликативна.

Задача 4.54. Покажите, что функция $\sigma(n)$, равная сумме всех положительных делителей числа n, мультипликативна.

4.14 Что дальше?

Мы познакомились с базовыми свойствами сложения и умножения по модулю N, или, как говорят, *кольца вычетов по модулю* N, стараясь не использовать алгебраического языка. Есть и много других интересных (и не таких сложных) результатов. Скажем, есть такая теорема: при простом N найдётся остаток a по модулю N, при котором все элементы $1, a, a^2, \ldots, a^{N-2}$ различны (следующий элемент по теореме Ферма равен 1). Или другой результат: ровно половина всех ненулевых вычетов по простому модулю являются точными квадратами (квадратами других вычетов). В принципе и их можно было бы изложить, не используя языка алгебры (группы, подгруппы, кольца, гомоморфизмы, многочлены и пр.), но в этом нет смысла: пора с ним познакомиться. Это можно сделать по любому учебнику алгебры — есть старинный «Курс высшей алгебры» Куроша (когда-то бывший стандартным мехматским учебником), или несколько более новый учебник Кострикина (тоже по материалам мехматского курса). Стандартные учебники, покрывающие существенно больший материал — классические книги ван дер Вардена и Ленга. В программе ВШЭ курс алгебры предусмотрен в 4 модуле первого курса.

Где все эти алгебраические премудрости используются «на практике»? Вот игрушечный пример: пусть вы хотите послать кому-то номер своей кредитной карточки, но боитесь, что ваше сообщение будет перехвачено. Тогда есть такой способ: каждую цифру a_i этого номера можно представить в виде суммы по модулю 10 случайно выбранной цифры r_i и недостающей цифры $s_i = a_i - r_i$, и послать последовательности цифр r_i и s_i (той же длины, что и номер карточки) в двух разных электронных письмах. Получив оба этих письма, адресат легко восстановил $a_i - r_i + s_i$. Но если

враг перехватит только одно из них, а второе нет — то у него не будет никакой информации о вашей карточке. В самом деле, сообщение r не содержит никакой информации (это просто случайная последовательность цифр), и сообщение s в отдельности тоже (потому что при данном a это сообщение может быть любым, и все они равновероятны).

Конечно, защита тут слабая — лишь на случай перехвата одного из писем. Но, скажем, если одно посылается по электронной почте, а другое — как сообщение в Skype, то враг должен получить доступ к обоим сервисам (они разные, и каналы передачи сообщений разные). Существуют и более надёжные методы передачи сообщений по открытой сети, прежде всего так называемая система RSA (она используется в банках, в протоколах https, ssh, цифровой подписи PGP и вообще на каждом шагу). И она тоже основана на арифметике остатков. Наших знаний достаточно, чтобы понять, почему она не искажает сообщения — но вот доказать, что её нельзя взломать, мы не можем, и никто пока не может — если кто-то вдруг научится быстро разлагать большие числа на множители, эта система, а вместе с ней и все банки, рухнет. Но для этого (быстрого разложения) нужны либо новые алгоритмы, либо квантовые компьютеры (которые, если вдруг их удастся построить, могут разлагать числа на множители — это обнаружил P. Shore). Другие системы шифрования основаны на более сложных алгебраических методах (эллиптические кривые) — так что знания по алгебре лишними не будут.

Вообще алгебраические методы незаменимы там, где нужно построить «регулярно устроенное» семейство объектов с хорошими свойствами. Попробуйте, например, найти в 7-элементном множестве как можно больше 3-элементных подмножеств, любые два из которых имеют общий элемент. (Сколько дней подряд можно назначать по три дежурных в походе из 7 человек, чтобы наряды на любые два дня имели ровно одного общего?) Не так просто — если не знать, что двумерные подпространства в трёхмерном пространстве над полем вычетов по модулю 2 содержат (не считая нуля) по три элемента и любые два их них пересекаются по одномерной прямой, содержащей ровно один ненулевой элемент.

Часть II

Основные конструкции

Лекция 5

Множества

Мы уже использовали понятие множества и в дальнейшем будем его использовать постоянно. Сейчас мы обсудим это понятие и связанные с ним более систематически.

Стоит также предупредить читателя, что мы не собираемся излагать теорию множеств формально. Это довольно трудное занятие. Наша цель более скромная — объяснить, как понятия множества, операций с множествами и свойств множеств используются для формулировки математических утверждений и описания математических рассуждений. Далее мы будем при необходимости свободно использовать теоретико-множественный язык. Эту главу можно рассматривать как краткий «русско – теоретико-множественный разговорник».

5.1 Основные свойства множеств и операции с множествами

Неформально, множество — это совокупность каких-то элементов. Природа элементов неважна, как и возможные взаимоотношения между ними. Единственное, что существенно для определения множества — это какие элементы в него входят, а какие — нет. При этом множество не может содержать «половину» элемента[1]. Элемент либо входит в множество, либо нет.

Замечание 5.1. В жизни мы постоянно имеем дело с более сложными совокупностями, которые в бытовом языке тоже называются «множествами». Например, «множество лысых людей». Кто именно входит в это «множество», не вполне понятно и определяется разными людьми по-разному, зачастую из очень субъективных соображений. И даже «множество людей» — это совокупность, не являющаяся множеством в математическом смысле. Из-за различного понимания содержания этой совокупности происходят серьёзные политические дебаты и даже убийства (имеются в виду дискуссии об абортах и убийства абортмахеров).

Математики любят иллюстрировать абстрактные идеи на конкретных примерах. Поэтому вы вполне можете столкнуться в математических книгах с множеством лю-

[1] Однако дробные части элементов — это не бессмысленная забава. Им можно придать разумный смысл при изучении теории вероятностей, см. лекцию 11.

дей или, скажем, множеством пароходов. Нужно помнить, что в этих иллюстрациях всегда неявно предполагается, что мы заранее как-то договорились о том, какие элементы входят в множество людей, а какие — в множество пароходов.

Из сделанных уточнений ясно, что множество полностью определяется своими элементами. Два множества A и B равны тогда и только тогда, когда любой элемент множества A является элементом множества B, а любой элемент множества B является элементом множества A. Это определение равенства множеств является по сути самым важным из свойств множеств. В дальнейшем это определение нужно постоянно иметь в виду, когда речь заходит о равенстве множеств. А такое будет происходить очень часто, так как мы собираемся выражать на языке множеств другие математические конструкции.

Среди множеств есть уникальное множество — пустое, — которое не содержит никаких элементов. Пустое множество обозначается $\varnothing$.

Контрольный вопрос 5.1. Используя определение равенства множеств, аккуратно докажите единственность пустого множества.

Если элементов в множестве мало, его можно задать, указав все эти элементы. При этом принято заключать список элементов в фигурные скобки. Хотя на письме мы вынуждены записывать множества в каком-то порядке, этот порядок не играет роли. Поэтому

$$\{0, 2, 4, 6, 8\} = \{4, 2, 0, 8, 6\}$$

и это просто разные обозначения множества чётных цифр.

Способ задания множества списком его элементов очень удобен для компьютеров, но не всегда удобен для людей.

Пример 5.1. Множество простых десятизначных чисел содержит достаточно мало элементов, чтобы задать его списком элементов. Вы без труда можете написать программу, которая составит такой список и распечатает его красивым шрифтом.[2] Но попробуйте без компьютерных вычислений указать хотя бы один элемент этого множества! Например, входит ли в это множество число 1000001009? а 1000001011?

Как видно, мы тремя словами описали множество, об элементах которого имеем лишь весьма смутное представление.

Бывают однако такие множества, которые уж никак не задать списком элементов. Просто потому, что элементов в этих множествах много. Точнее, бесконечно много. Скажем, множество целых чисел $\mathbb{Z}$. Для каждого целого числа n найдётся целое число, которое больше n. Так что со списком элементов ничего не выйдет.

Как же мы справляемся с такой ситуацией? Во-первых, для некоторых множеств мы используем специальные обозначения (как для множества целых чисел). Во-вторых, разобранный выше пример показывает основную идею гибкого описания множеств. Множества можно задавать как части других множеств, выделяя в них

[2] Предупреждение: в рюкзаке эту распечатку не унести.

элементы, удовлетворяющие некоторому свойству. Скажем, из множества целых чисел таким способом выделяется множество простых чисел. Это те целые числа, у которых нет целого обратного и делителей, отличных от ± 1 и самого числа.

Выделение множества из другого множества указанием некоторого свойства элементов встречается очень часто в математике. Поэтому полезно ознакомиться с принятыми формальностями при записи таких определений.

Говорят, что элемент x *принадлежит* множеству A, если он является его элементом. Стандартное обозначение этого утверждения[3] такое: $x \in A$. Соответственно, отрицание принадлежности, то есть «x не принадлежит множеству A», обозначается $x \notin A$.

Части множества называются *подмножествами*. По определению множество A является подмножеством множества B, если каждый элемент A принадлежит множеству B. Обозначается это так: $A \subseteq B$ (словами: «A подмножество B»).

Равенство множеств и включение напоминают равенство чисел (и достижимость в графах) и сравнение чисел по величине (а также достижимость в орграфах) и обладают похожими свойствами. Для равенства это рефлексивность «$A = A$», симметричность «если $A = B$, то $B = A$» и транзитивность «если $A = B$, $B = C$, то $A = C$». Для включения множеств это рефлексивность «$A \subseteq A$», и транзитивность «если $A \subseteq B$, $B \subseteq C$, то $A \subseteq C$» и ещё одно свойство антисимметричности: «если $A \subseteq B$ и $B \subseteq A$, то $A = B$» (это сразу следует из определений — проверьте!).

Как уже говорилось, подмножества выделяют указанием некоторых свойств элементов множества. Например, чётные числа — это те целые числа, которые делятся на 2. Часто используется формальная запись такого определения множества. Она напоминает перечисление множества списком элементов, но теперь список «неявный». На примере чётных чисел это выглядит так:

$$\{x \in \mathbb{Z} \mid x = 2y \text{ и } y \in \mathbb{Z}\}.$$

Если ясно, о каком объемлющем множестве идёт речь, его обозначение опускают. Скажем, если ясно, что мы определяем подмножество множества целых чисел, допустимо определить натуральные числа[4] так:

$$\mathbb{N} = \{x \mid x \geqslant 0\}.$$

Тут, как и обычно в математике, гибкость и понятность обозначений обратно пропорциональна их точности.

Замечание 5.2. Последнее замечание приложимо и к самой идее выделения множества свойствами элементов некоторого другого множества. Рассмотрим такой пример. Определим множество сложных чисел как множество, состоящее из тех натуральных чисел, которые нельзя определить, используя не больше четырнадцати слов

[3] Это утверждение: элемент либо принадлежит множеству, либо нет; поэтому $x \in A$ либо истинно, либо ложно.

[4] Мы предпочитаем именно такое определение натуральных чисел, включающее 0.

русского языка. Казалось бы, мы задали подмножество множества натуральных чисел аналогично предыдущим простым примерам. Однако в этом множестве должно быть наименьшее число. И с этим числом возникает проблема: фраза из тринадцати слов «наименьшее натуральное число, которое нельзя определить, используя не больше четырнадцати слов русского языка» определяет это число в противоречии с самим определением!

В этой книге нет нужды сталкиваться со столь замысловатыми случаями. Однако полезно, в том числе и по другим причинам, продумывать каждое определение и проверять, что оно не содержит в себе таких несуразностей.

Есть ещё один удобный способ определять новые множества из уже имеющихся. Для этого к множествам можно применять *операции*.

На множествах определено большое количество операций, самые важные из которых мы сейчас перечислим.

Объединение множеств. Обозначение $A \cup B$. Это множество, состоящее в точности из тех элементов, которые принадлежат хотя бы одному из множеств A и B.

Пересечение множеств. Обозначение $A \cap B$. Это множество, состоящее в точности из тех элементов, которые принадлежат обоим множествам A и B.

Разность множеств. Обозначение $A \setminus B$. Это множество, состоящее в точности из тех элементов, которые принадлежат множеству A, но не принадлежат множеству B.

Симметрическая разность множеств. Обозначение $A \triangle B$. Это множество, состоящее в точности из тех элементов, которые принадлежат ровно одному из множеств: либо A, либо B.

Помимо словесных определений, приведённых выше, есть наглядный графический способ иллюстрировать операции с множествами: круги Венна (или Эйлера–Венна, как пишут в некоторых книгах). При этом способе множество изображается условным кругом (или другой геометрической фигурой) и предполагается, что внутренность круга изображает элементы множества.

На паре кругов легко изобразить объединение, пересечение, разность и симметрическую разность множеств, как это сделано на рисунке 5.1 — выделяя результат применения операции штриховкой или цветом.

Контрольный вопрос 5.2. Определите, на каких картинках рис. 5.1 какие операции изображены.

Элементами множества могут быть другие множества.

Пример 5.2. Зададим множество и его элементы–множества списками элементов:

$$\{\varnothing, 1, \{1, \varnothing\}, \{1, 2\}\}$$

Изучив эту запись, легко убедиться, что всего в множестве 4 элемента, один из которых — число 1, ещё один — пустое множество и есть также два элемента, кото-

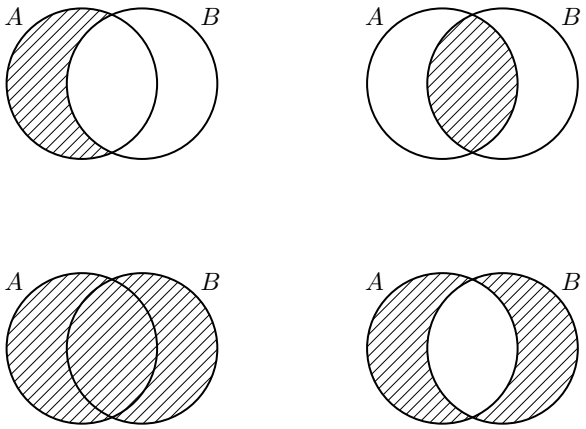


Рис. 5.1: Теоретико-множественные операции на кругах Венна

рые сами по себе являются 2-элементными множествами. Причём в одном из этих элементов-множеств один из элементов также не число, а пустое множество.[5]

Пример 5.3. Вспомним, как мы определяли неориентированные графы. Для каждой пары вершин мы указывали, соединены ли они ребром. Пары вершин, соединённых ребром образуют множество, элементами которого являются 2-элементные подмножества вершин (попросту пары вершин).

Приходим к другому определению неориентированного графа (без петель и кратных рёбер) как пары множеств (V, E) (множества вершин и множества рёбер), в которой E является подмножеством множества, состоящего из всех пар вершин. Это определение общеупотребительно в книгах по теории графов.

Пример 5.4. Ещё один пример конечного множества, состоящего из бесконечных множеств, встретился нам в арифметике остатков. Вычет по модулю n — это множество целых чисел, имеющих одинаковый остаток при делении на n. Всего разных вычетов ровно n штук. Используя обозначения для задания множества списком элементов с некоторой вольностью, можно записать множество вычетов по модулю 3 как

$$\{\{\dots,-3,0,3,\dots\},\ \{\dots,-2,1,4,\dots\},\ \{\dots,-1,2,5,\dots\}\}.$$

Замечание 5.3. На всякий случай укажем ещё одну тонкость в построении множеств. Давайте «определим» множество A списком его элементов: $A = \{A\}$. То есть единственным элементом множества является оно само. Что же это за множество, которое является своим собственным элементом и притом единственным? Понять это затруднительно.

В «настоящих» множествах такая конструкция запрещена и никакое множество не является своим собственным элементом.

[5]Если заняться строгим построением математики из множеств, то числа тоже станут множествами, причём довольно хитроумно устроенными. Мы в эти вопросы углубляться не будем.

Множество подмножеств. По любому множеству A определено множество его подмножеств, которое обозначается 2^A. Элементами этого множества являются в точности подмножества множества A.

Есть также способ определить возведение в степень A^B для любых множеств, но мы его не будем обсуждать.

Замечание 5.4. Внимательный читатель обратил, наверное, внимание на то, что среди теоретико-множественных операций мы не назвали *дополнение*. Тому есть очень существенная причина.

Дополнение к множеству (обозначение $\bar{A}$) определяют обычно как те элементы, которые не входят в множество A. Если использовать такое определение буквально, то дополнение — не очень осмысленное понятие. Скажем, в дополнение к множеству $\{1\}$ нужно включить и число 2, и функцию синус, и полный граф на 140 вершинах.

Более того, если строить теорию множеств строго, такое определение дополнения становится совсем неудачным. Мы уже говорили о том, что не всякая совокупность является множеством. Так вот, совокупность элементов, не принадлежащих множеству множеством быть не может.

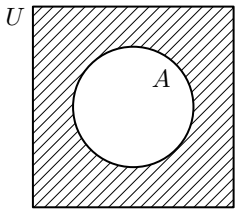


Рис. 5.2: Дополнение $\bar{A} = U \setminus A$ на рисунке заштриховано

Определение дополнения имеет смысл в более узкой ситуации, которая, однако, встречается очень часто. Предположим, мы рассуждаем о подмножествах некоторого множества U. В этом случае элементы множества U, которые не принадлежат некоторому подмножеству A, образуют множество, которое и называется дополнением $\bar{A}$ к множеству A. Определённое таким образом дополнение выражается через операцию разности множеств: $\bar{A} = U \setminus A$ (см. рис. 5.2). В этом случае дополнение является по сути сокращённым обозначением разности некоторого фиксированного в рассуждении множества U и подмножества этого множества. Именно так оно будет возникать в этой книге в дальнейшем.

5.2 Теоретико-множественные тождества

Может оказаться так, что применение теоретико-множественных операций в разном порядке даёт одно и то же множество, к каким бы множествам не применять данные последовательности операций. В таких случаях говорят о теоретико-множественных тождествах.

Простейшие примеры тождеств

$$A\cup B = B\cup A;\ A\cap B = B\cap A;\ (A\cup B)\cup C = A\cup(B\cup C);\ (A\cap B)\cap C = A\cap(B\cap C).$$

Они очевидны из определения: достаточно пересказать определения словами и станет ясно, что левые и правые части равенств задают одно и то же.

И вообще, определения объединения и пересечения имеют смысл не только для двух (как в исходном определении) или трёх (как в этом примере) множеств, но и для произвольного *семейства* множеств.[6] Объединением семейства множеств является множество, состоящее в точности из элементов, которые входят хотя бы в одно множество семейства. Пересечением семейства множеств является множество, состоящее в точности из элементов, которые входят во все множества семейства.

Вот менее очевидный пример тождества:

$$(A\cap B)\setminus C = (A\setminus C)\cap B. \tag{5.1}$$

Нарисуем левую и правую часть на картинке кругов Венна, рис. 5.3. Видим, что множество получается одно и то же в обоих случаях.

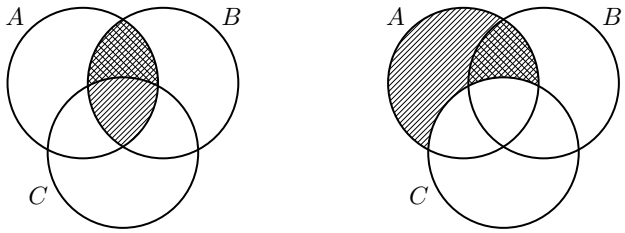


Рис. 5.3: На рисунках штриховкой вправо выделен результат первой операции, двойной штриховкой выделен результат второй операции

Почему рассуждение с картинкой корректно? Оно на самом деле скрывает под собой разбор возможных случаев. Каждый элемент может принадлежать или не принадлежать одному из трёх множеств. Поэтому всего есть 8 вариантов, которые отмечены на рисунке 5.4. Варианты обозначены двоичными словами.

Цифра 0 на первом месте означает, что данная область обозначает элементы, которые не входят в множество A, цифра 1 означает вхождение в множество A. Вторая цифра аналогично означает принадлежность или непринадлежность множеству B; третья — множеству C.

Если известно, в какие множества из данных трёх входит элемент, то можно определить, входит ли он в пересечение, разность и т.д. каких-то из этих трёх мно-

[6] Слово «семейство» здесь и в аналогичных ситуациях является синонимом слова «множество» и используется для гладкости речи.

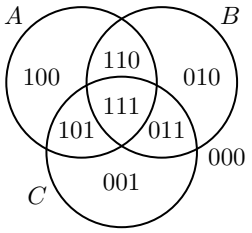


Рис. 5.4: 8 вариантов принадлежности или не принадлежности каждому из трёх множеств

жеств. Составим таблицу всех возможных вариантов:

$x \in A$	$x \in B$	$x \in C$	$x \in A \cap B$	$x \in (A \cap B) \setminus C$	$x \in A \setminus C$	$x \in (A \setminus C) \cap B$
0	0	0	0	0	0	0
0	0	1	0	0	0	0
0	1	0	0	0	0	0
0	1	1	0	0	0	0
1	0	0	0	0	1	0
1	0	1	0	0	0	0
1	1	0	1	1	1	1
1	1	1	1	0	0	0

Легко видеть, что пятый и седьмой столбцы таблицы совпадают. Это означает, что каждый элемент входит в множество, задавамое левой частью (5.1), тогда и только тогда, когда он входит в множество, задаваемое правой частью. По определению равенства множеств эти множества совпадают.

Аналогично можно доказывать и другие тождества.

Задача 5.5. Докажите следующие тождества, используя круги Венна и таблицы значений:

(a) $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$ (дистрибутивность объединения и пересечения);

(b) $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$ (дистрибутивность пересечения и объединения);

(c) $(A \triangle B) \triangle C = A \triangle (B \triangle C)$ (ассоциативность симметрической разности).

При большом количестве переменных-множеств круги Венна становятся не слишком наглядными, а таблицы значений становятся слишком длинными. Поэтому удобнее использовать свойства операций с множествами и выполнять равносильные преобразования аналогично тому, как мы это делаем в элементарной алгебре. Тождеств с множествами много, они не исчерпываются примерами, которые выше разобраны.