КАЗАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Ю.А. Альпин, С.Н. Ильин

ДИСКРЕТНАЯ МАТЕМАТИКА: ГРАФЫ И АВТОМАТЫ

УЧЕБНОЕ ПОСОБИЕ

Казань
2006

УДК 519

Печатается по решению учебно-методической комиссии
механико-математического факультета КГУ

Научный редактор
кандидат физико-математических наук, доцент Кугураков В.С.

Альпин Ю.А., Ильин С.Н.
Дискретная математика: графы и автоматы. Учебное пособие. — Казань: Казанский государственный университет им. В.И. Ульянова-Ленина, 2006. — 78 с.

Учебное пособие предназначено для студентов 2-го курса механико-математического факультета КГУ и содержит разделы, традиционно излагаемые в общем курсе дискретной математики. Оно также может быть использовано в качестве основы для специальных курсов по теории графов и теории автоматов.

Оглавление

Глава 1
Неориентированные графы

§ 1. Первые понятия теории графов

Графом называется пара (V, E), где V — непустое множество, E — некоторое множество двухэлементных подмножеств V. Элементы V называют *вершинами*, а элементы E — *рёбрами* графа.[1] Мы будем рассматривать только *конечные* графы, то есть, графы с конечными множествами вершин. Графы удобно изображать в виде рисунков, состоящих из точек и линий, соединяющих некоторые пары точек. Точки соответствуют вершинам графа, а линии — рёбрам. Например, рисунок

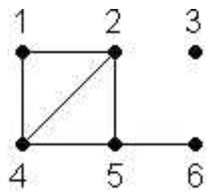


Рис. 1

изображает граф с множеством вершин $V = \{1, 2, 3, 4, 5, 6\}$ и множеством рёбер $E = \{\{1, 2\}, \{1, 4\}, \{2, 4\}, \{2, 5\}, \{4, 5\}, \{5, 6\}\}$.

Если $e = \{u, v\}$ — ребро графа, то вершины u и v называются *концами* ребра e. Говорят, что ребро $e = \{u, v\}$ *соединяет* вершины u и v. Вершины, соединенные ребром, называются *смежными*. Ещё говорят, что вершины u и v *инцидентны* ребру $e = \{u, v\}$. Отметим крайние случаи: граф называется *полным*, если любые его две вершины смежны, и *пустым*, если множество рёбер пусто.

Маршрутом в графе называется всякая чередующаяся последовательность вершин и рёбер

$$v_1, e_1, v_2, , \ldots, e_k, v_{k+1}, \tag{1}$$

в которой ребро e_i соединяет вершины v_i и v_{i+1} при $i = 1, 2, \ldots, k$. Обычно маршрут записывается короче, как последовательность

$$v_1, v_2, \ldots, v_{k+1},$$

[1] Точнее говоря, приведённое определение относится к *неориентированным* графам, но в первой главе это прилагательное опускается.

в которой соседние вершины v_i и v_{i+1} смежны (иногда запятые заменяются чёрточками). Про маршрут (1) говорят, что он связывает вершины v_1 и v_{k+1}, или является (v_1, v_{k+1})-маршрутом. Число k рёбер в маршруте (1) называется его *длиной*. Маршрут называется *замкнутым*, если $v_1 = v_{k+1}$.

Маршрут называется *цепью*, если все его рёбра различны. Цепь называется *простой*, если её вершины различны, кроме, может быть, первой и последней. Замкнутая цепь называется *циклом*, замкнутая простая цепь — *простым циклом*. Заметим, что длина незамкнутой простой цепи в графе с n вершинами не больше, чем $n-1$, а длина простого цикла — не больше, чем n.

Лемма 1. *При $u \neq v$ всякий (u, v)-маршрут содержит простую (u, v)-цепь.*

Доказательство. Если маршрут не является простой цепью, то он имеет вид

$$u - \ldots - s - t - \ldots - s - w - \ldots - v.$$

Но тогда имеется более короткий (u, v)-маршрут

$$u - \ldots - s - w - \ldots - v.$$

Если и он — не простая цепь, то снова возможно сокращение. Ясно, что после нескольких сокращений получится простая (u, v)-цепь. $\square$

Следующая лемма доказывается аналогично и рекомендуется в виде упражнения.

Лемма 2. *Всякий цикл содержит простой цикл, причём каждая вершина и ребро цикла принадлежат некоторому простому циклу.*

Лемма 3. *Если в графе есть разные простые (u, v)-цепи, то граф содержит простой цикл, составленный из вершин и рёбер этих цепей.*

Доказательство. Рассмотрим неравные простые цепи

$$u = l_1 - l_2 - \ldots - l_k = v,$$
$$u = p_1 - p_2 - \ldots - p_m = v.$$

Пусть i — наименьший номер, для которого $l_i = p_i$, но $l_{i+1} \neq p_{i+1}$. Пусть $j > i$ — наименьший номер, такой, что $l_j = p_t$ при некотором t. Тогда $l_i - l_{i+1} - \ldots - l_j - p_{t-1} - \ldots - p_i$ — простой цикл. $\square$

Говорят, что вершины u и v *связаны*, если $u = v$ или существует (u, v)-маршрут. Граф называется *связным*, если любые две его вершины связаны. Если граф не связен, то он представляет собой объединение нескольких связных подграфов. Чтобы выразиться точнее, введем понятие подграфа.

Граф (V_1, E_1) называется *подграфом* графа (V, E), если $V_1 \subseteq V$, $E_1 \subseteq E$. Например, цепь в графе можно рассматривать как подграф. Говорят, что подграф *порождён подмножеством вершин* V_1, если E_1 состоит из рёбер, соединяющих вершины из V_1. Говорят, что подграф *порождён подмножеством рёбер* E_1, если V_1 состоит из концов рёбер из E_1.

Легко убедиться, что бинарное отношение связанности на множестве вершин графа рефлексивно, симметрично и транзитивно, то есть, связанность вершин является отношением эквивалентности.

Как известно, всякое отношение эквивалентности разбивает множество, на котором оно задано, на классы. Классы связанности вершин назовём *компонентами* графа и тем же термином обозначим подграфы, порождённые этими классами. Ясно, что компоненты — это связные графы и что вершины из разных компонент несмежны. Сколько компонент у графа, изображенного на рис. 1?

Задача 1. Докажите, что связный подграф является компонентой тогда и только тогда, когда он не содержится в другом связном подграфе.

Определение графа не предполагает нумерации его вершин. Но часто это бывает удобно, поскольку тогда мы можем сопоставить графу матрицу и использовать для изучения графов матричную алгебру. Если на множестве V вершин графа с n вершинами задана нумерация, то есть, некоторая биекция $\sigma : V \to \{1, 2, \ldots, n\}$, то граф называется *нумерованным*. Вершины нумерованного графа будем обозначать числами $1, 2, \ldots, n$. *Матрицей смежности* нумерованного графа с n вершинами называется матрица $A = (a_{ij})$ порядка n, в которой

$$a_{ij} = \begin{cases} 1, & \text{если вершины } i \text{ и } j \text{ смежны}, \\ 0 & \text{в противном случае.} \end{cases}$$

Очевидно, матрица смежности симметрична и ее главная диагональ состоит из нулей. Наоборот, любая симметричная (0,1)-матрица с нулевой диагональю определяет нумерованный граф, в котором вершины i и j смежны, если (i, j)-элемент матрицы равен 1. В дальнейшем, когда речь идёт о матрице смежности графа, то обычно молчаливо предполагается, что граф нумерован.

Упражнение 1. Пусть $V_1, \ldots, V_s$ — компоненты графа. Перенумеруем сначала вершины из V_1, затем вершины из V_2 и так далее. Докажите, что при такой нумерации матрица смежности графа имеет блочно-диагональный вид:

$$\begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_s \end{pmatrix},$$

где A_i, $i = 1, \ldots, s$, — матрицы смежности компонент.

Элементы степеней матрицы смежности имеют прозрачный графовый смысл.

Предложение 1. *Если A — матрица смежности графа, то (ij)-элемент $a_{ij}^{(k)}$ матрицы A^k равен количеству (i,j)-маршрутов длины k в этом графе.*

Доказательство. Для элементов k-й степени матрицы A имеет место равенство

$$a_{ij}^{(k)} = \sum a_{il_1} a_{l_1 l_2} \ldots a_{l_{k-1},j}, \tag{2}$$

где суммирование производится по всевозможным последовательностям индексов $l_1, l_2, \ldots, l_{k-1}$. (Формула (2) выводится индукцией по параметру k.) Очевидно, произведение $a_{il_1} a_{l_1 l_2} \ldots a_{l_{k-1}j}$ равно единице, если существует маршрут $i, l_1, l_2, \ldots, l_{k-1}, j$, и равно нулю в противном случае. Отсюда и следует утверждение. □

Если нам требуется знать не количество (i,j)-маршрутов, а лишь выяснить, существует ли хотя бы один маршрут, то удобнее считать символы 0 и 1 не натуральными числами, а элементами двухэлементной *булевой алгебры*, в которой сложение и умножение задаются следующими таблицами:

$$\begin{array}{c|c|c} \oplus & 0 & 1 \\ \hline 0 & 0 & 1 \\ \hline 1 & 1 & 1 \end{array}\,, \quad \begin{array}{c|c|c} \odot & 0 & 1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 0 & 1 \end{array}\,.$$

Можно непосредственно проверить, что операции в этой двухэлементной алгебре обладают привычными свойствами ассоциативности, коммутативности и дистрибутивности умножения относительно сложения, но относительно сложения булева алгебра не является группой (вычитание не определено).

Для упрощения записи вместо $a \oplus b$ и $a \odot b$ мы будем писать $a + b$ и ab, так как из контекста видно, когда действия происходят в булевой алгебре.

Матрицы над булевой алгеброй складывают и умножают по обычным правилам, при этом сохраняются известные свойства операций с матрицами (с теми же доказательствами). Но нужно забыть о вычитании матриц. Впрочем, оно и не потребуется.

Условимся называть (0,1)-матрицу *булевой*, если с её элементами мы намерены обращаться по правилам булевой алгебры. Приведём "булевский" вариант предложения 1.

Предложение 2. *Если A — булева матрица смежности графа, то $a_{ij}^{(k)} = 1$ тогда и только тогда, когда в этом графе существует (i, j)-маршрут длины k.*

Через I будем обозначать прямоугольную булеву матрицу, все элементы которой равны 1, а размеры таковы, что формулы, в которых она участвует, имеют смысл. Буква E, как обычно, обозначает квадратную матрицу, диагональные элементы которой равны 1, а прочие равны 0.

Задача 2. Пусть граф с n вершинами задан булевой матрицей смежности A. Докажите, что

a) $(E + A)^m = E + A + \ldots + A^m, \quad m = 1, 2, \ldots;$

b) вершины i, j связаны $\Leftrightarrow$ (ij)-элемент матрицы $(E + A)^{n-1}$ равен 1;

c) граф связен $\Leftrightarrow (E + A)^{n-1} = I$.

Задача 3. Пусть A — матрица смежности графа. Чему – в графовых терминах – равно число $\operatorname{tr} A^2$?

Графы (V, E) и (V', E') называются *изоморфными*, если существует такая биекция $\sigma : V \to V'$, что $(u, v) \in E \;\Leftrightarrow\; (\sigma(u), \sigma(v)) \in E'$. С точки зрения теории изоморфные графы представляют, по существу, один граф, но с различными обозначениями вершин.

Изоморфизм графов можно определить в матричных терминах. Предварительно введём понятие перестановочного подобия матриц. Оно формулируется одинаково для булевых матриц и матриц над полем.

Квадратная $(0, 1)$-матрица P называется *перестановочной*, если она имеет в каждой строке и каждом столбце ровно одну единицу. Легко проверить, что тогда $PP^T = P^T P = E$, то есть $P^T = P^{-1}$. Матрицы A и B называются *перестановочно подобными*, если $A = PBP^T$ для некоторой перестановочной матрицы P. Содержательный смысл этого определения заключается в том, что A получается из B одинаковыми перестановками строк и столбцов.

Пусть графы с n вершинами, заданные матрицами смежности A и B, изоморфны, то есть существует такая биекция (перестановка) σ

на множестве $\{1, 2, \ldots, n\}$, что для любых i, j

$$a_{ij} = b_{\sigma(i)\sigma(j)}.$$

Сопоставим перестановке σ перестановочную матрицу $P = (p_{ij})$ порядка n, где

$$p_{ij} = \begin{cases} 1, & \text{если } j = \sigma(i), \\ 0 & \text{в противном случае.} \end{cases} \tag{3}$$

Прямыми вычислениями проверяется, что

$$A = PBP^T.$$

Итак, если графы изоморфны, то их матрицы смежности перестановочно подобны. Наоборот, если матрицы смежности перестановочно подобны, то графы изоморфны, причем изоморфизм σ определяется по матрице подобия P из равенств (3). Суммируем наши рассуждения.

Предложение 3. *Графы изоморфны тогда и только тогда, когда их матрицы смежности перестановочно подобны.*

Замечание 1. После нумерации σ вершин графа (или перенумерации, если вершины уже нумерованы) получается, формально говоря, другой граф, изоморфный данному, с некоторой матрицей смежности A. Но поскольку по существу это тот же граф, мы будем в этой ситуации говорить так: после нумерации σ матрица смежности графа равна A. По той же причине перестановочно подобные (0,1)-матрицы естественно рассматривать как матрицы смежности одного графа при различных нумерациях вершин.

В заключение параграфа отметим, что иногда используются графы, в которых, как на рис. 2, две вершины соединены несколькими рёбрами (из называют *кратными* рёбрами) или имеются рёбра с совпадающими концами (их называют *петлями*).

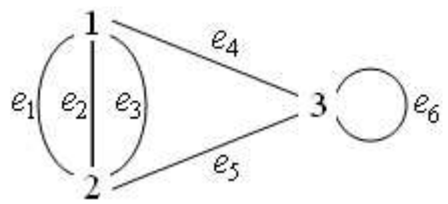


Рис. 2

В нашем курсе, без существенного ущерба для содержания теории, рассматриваются графы без кратных рёбер и петель. Исключением является первая теорема следующего параграфа — теорема

Эйлера, где из уважения к исторически первой задаче теории графов допускаются кратные рёбра.

§ 2. Эйлеровы и гамильтоновы графы

Теория графов началась со следующей *задачи о кёнигсбергских мостах*, поставленной и решённой Эйлером. В этом городе семь мостов переброшены через реку Прегель, как показано на карте ниже.

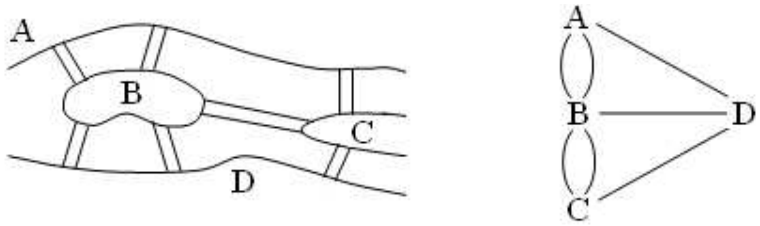


Рис. 3

Спрашивается, можно ли, выйдя из дома, вернуться обратно, пройдя по каждому мосту ровно один раз? Граф, отображающий математическое существо задачи, изображен на рис. 3 рядом с картой.

Общая задача будущей теории графов, поставленная Эйлером, формулируется так: при каких условиях связный граф содержит цикл, проходящий через каждое его ребро?

Цикл, проходящий через все рёбра графа, называют *эйлеровым*, а граф, содержащий эйлеров цикл, — *эйлеровым графом.*

Для решения задачи потребуется новое понятие. Назовём *степенью вершины* количество инцидентных ей рёбер.

Теорема 1 (Эйлер, 1736 г). *Связный граф содержит эйлеров цикл тогда и только тогда, когда степень каждой его вершины есть чётное число.*

Доказательство. Необходимость. Пусть эйлеров цикл существует. Будем проходить по нему, удаляя из графа пройденные рёбра. Ясно, что при входе в вершину её степень уменьшается на единицу, а при выходе — ещё на единицу, то есть, при проходе через вершину из её степени вычитается двойка. Когда все рёбра цикла пройдены, остаётся пустой граф: степень каждой вершины равна нулю. Таким образом, степень каждой вершины исходного графа есть сумма двоек, то есть, чётное число.

Достаточность. Начиная с произвольно выбранной вершины v, строим маршрут, соблюдая единственное правило: не проходить дважды по одному ребру. Поскольку число рёбер конечно, построение закончится попаданием в такую вершину v', что все инцидентные ей рёбра уже пройдены. Докажем, что $v' = v$, следовательно, построен цикл. Предположим, что $v' \neq v$, тогда выходит, что наш маршрут несколько раз прошёл через вершину v', используя по два инцидентных ей ребра, и на последнем шаге вошел в неё по последнему свободному ребру. Получается, что степень вершины v' — нечётное число, но это противоречит условию теоремы. Итак, построен некоторый цикл C_1. Заметим, что при этом связность графа не потребовалась. Если C_1 содержит все рёбра графа, то этот цикл эйлеров. Если не все, то существует некоторая вершина u цикла C_1, инцидентная ребру, не входящему в C_1. В противном случае подграф, порождённый рёбрами, не входящими в C_1, не имел бы с C_1 общих вершин. Но это противоречит связности графа.

Согласно рассуждению в части необходимости, каждая вершина цикла инцидентна чётному числу рёбер этого цикла. Отсюда и из условия теоремы следует, что рёбер, инцидентных любой из вершин графа и притом не принадлежащих C_1, тоже чётное число. Поэтому можно построить новый цикл C_2, начиная с вершины u по сказанному выше правилу, но не пользуясь рёбрами C_1. Соединяя C_1 и C_2 в вершине u, как показано на рисунке 4, получим цикл C'.

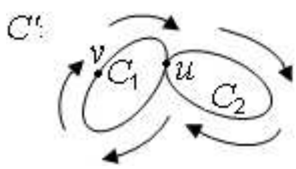


Рис. 4

Если он содержит все рёбра графа, то построение закончено. Если нет, то найдётся вершина $s \in C'$, инцидентная ребру не из C'. Сделаем её началом нового цикла C_3 и так далее. Ясно, что на некотором шаге будет получен цикл, содержащий все рёбра графа, то есть, эйлеров цикл. □

Упражнение 1. Решите задачу о кёнигсбергских мостах.

Незамкнутую цепь назовём эйлеровой, если она содержит все рёбра графа. Вопрос о существовании такой цепи легко решается с помощью теоремы 1.

Следствие 1. *Незамкнутая эйлерова цепь в связном графе существует тогда и только тогда, когда граф содержит ровно две вершины нечётной степени.*

Доказательство. Пусть эйлерова цепь существует. Соединив её концы u и v новым ребром, получим эйлеров граф, в котором все вершины имеют чётную степень. Учитывая, что при добавлении ребра степени вершин u, v увеличились на 1, а степени других вершин не изменились, получаем требуемое утверждение.

Обратно, пусть связный граф содержит две вершины нечётной степени. Соединив их новым ребром, получим граф, все вершины которого имеют чётную степень. С учётом теоремы 1 завершение доказательства вполне очевидно. $\square$

Теперь от эйлеровых графов перейдем к понятию, внешне близкому, но имевшему иную историю. Известный ирландский математик У. Гамильтон в 1859 году опубликовал игру "Кругосветное путешествие". Её суть в следующем. Каждой из двадцати вершин додекаэдра приписано название одного из крупных городов мира. Требуется, переходя по ребрам додекаэдра, посетить каждый город ровно один раз и вернуться в исходный город. Разумеется, решение можно искать (попытайтесь найти его) на плоском изображении додекаэдра, а именно на рис 5.:

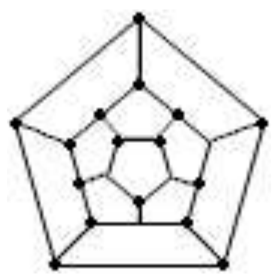


Рис. 5

Общая постановка задачи такова: в связном графе требуется найти простой цикл, содержащий все вершины графа. Такой цикл называется *гамильтоновым*, и сами графы, содержащие гамильтонов цикл, называются *гамильтоновыми*.

Несмотря на внешнее сходство, вопрос о том, как устроены гамильтоновы графы, в отличие от вопроса об эйлеровости, до сих пор не имеет удовлетворительного решения. Однако известны достаточные условия гамильтоновости графа. Рассмотрим два из них.

Теорема 2 (Дирак, 1952). *Пусть дан граф с n вершинами ($n \geqslant 3$). Для существования гамильтонова цикла достаточно, чтобы степень каждой вершины была не меньше, чем $n/2$.*

Доказательство. Предположим, что граф удовлетворяет условиям теоремы, но не является гамильтоновым. Добавим к нему новые вершины, соединяя каждую из них с каждой “старой” вершиной. Ясно, что если добавить n новых вершин $1'$, $2', \ldots, n'$, то в таком расширенном графе есть гамильтонов цикл $1 - 1' - 2 - 2' - \ldots - n - n' - 1$. Пусть $k > 0$ — наименьшее количество новых вершин, после добавления которых граф становится гамильтоновым. Рассмотрим гамильтонов цикл $p - q - r - \ldots - p$ в новом графе. Можно считать, что p, r — несмежные старые вершины, а q — новая (почему?). Докажем вспомогательное утверждение: *после каждой вершины, смежной с p, следует вершина, не смежная с r.* Предположим, случилось противное и цикл имеет вид

$$p - q - r - \ldots - s - t - u - \ldots - p,$$

где s смежна с p, а t смежна с r. Но тогда существует цикл

$$p - q - r \leftarrow \ldots \leftarrow s - t \rightarrow u \rightarrow \ldots \rightarrow p,$$

обозначенный стрелками. Как видим, новая вершина q оказывается лишней, но это противоречит минимальности числа k новых вершин. Итак, наше вспомогательное утверждение верно. Но тогда число вершин, не смежных с r, не меньше числа вершин, смежных с p, то есть, не меньше, чем $n/2 + k$. С другой стороны, число вершин, смежных с r, тоже не меньше, чем $n/2 + k$. Следовательно, общее количество вершин в расширенном графе не меньше, чем $n + 2k$, то есть, получено неравенство $n + k \geqslant n + 2k$, верное лишь при при $k = 0$, что противоречит предположению о негамильтоновости графа. $\square$

Интересно, что изменив лишь окончание доказательства предыдущей теоремы, можно получить более точный результат, известный как теорема Оре (1960).

Теорема 3. *Для существования гамильтонова цикла в графе с $n \geqslant 3$ вершинами достаточно, чтобы сумма степеней любых двух несмежных вершин была не меньше n.*

Д о к а з а т е л ь с т в о. Вначале повторяем доказательсто теоремы Дирака, включая вспомогательное утверждение и его следствие: число вершин, несмежных с r в расширенном графе, не меньше суммы степени вершины p в исходном графе и числа k добавленных вершин. А дальше будем рассуждать так. Поскольку степень вершины r в новом графе равна её степени в старом графе и числа k, то, учитывая условие теоремы, количество вершин в расширенном графе должно удовлетворять неравенству:

$$n + k \geqslant \text{степень } p + \text{степень } r + 2k \geqslant n + 2k.$$

Но это неравенство возможно лишь при $k = 0$. $\square$

§ 3. Деревья. Минимальные остовы графов

1. Основные свойства деревьев. Перейдём к изучению графов, в некотором смысле противоположных эйлеровым и гамильтоновым графам.

Деревом называется связный граф без циклов. Граф без циклов называют *лесом*.

На рис. 6 изображены примеры деревьев с 4-мя вершинами.

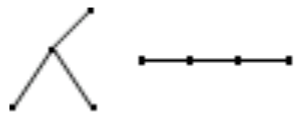


Рис. 6

Рассматриваемые вместе, эти деревья определяют лес.

Прежде, чем доказать теорему о свойствах деревьев, введём новое понятие и установим два полезных факта.

Ребро связного графа называется *мостом*, если удаление его приводит к несвязному графу (очевидно, с двумя компонентами). Вообще, ребро графа называется мостом, если оно является мостом компоненты, которой оно принадлежит.

Предложение 1. *Ребро является мостом тогда и только тогда, когда оно не принадлежит ни одному циклу.*

Доказательство достаточно провести для связного графа. Если ребро $\{u, v\}$ принадлежит циклу, то вершины u и v связаны маршрутом, не содержащим этого ребра. Этим маршрутом можно заменить все вхождения ребра $\{u, v\}$ в маршруты. Следовательно, удаление ребра $\{u, v\}$ не нарушает связанности вершин.

Обратно, если после удаления ребра $\{u, v\}$ получается связный граф, то в нём существуют (u, v)-маршрут и, следовательно, простая (u, v)-цепь, не содержащие ребра $\{u, v\}$. Но это значит, что в исходном графе это ребро принадлежит простому циклу. $\square$

Предложение 2. *Связный граф с n вершинами имеет не меньше, чем $n - 1$ ребер.*

Доказательство. Для графов с малым числом вершин утверждение очевидно. Пусть оно верно для графов с числом вершин, строго меньшим n. Докажем его для графов с n вершинами. Будем последовательно удалять из графа рёбра до тех пор, пока не получим несвязный граф с двумя компонентами. Пусть k — число удалённых рёбер. Ясно, что $k \geqslant 1$. Обозначим через n_1 и n_2 количества вершин в компонентах полученного графа. По предположению индукции первая компонента содержит не менее $n_1 - 1$ рёбер, вторая — не менее $n_2 - 1$ рёбер. Тогда исходный граф содержал рёбер не менее, чем $(n_1 - 1) + (n_2 - 1) + k = n - 2 + k \geqslant n - 1$. □

Теорема 1. *Для графа с n вершинами следующие свойства эквивалентны:*

1) граф связный и без циклов (то есть, дерево);

2) граф связный и у него $n - 1$ рёбер;

3) граф не содержит циклов и имеет $n - 1$ рёбер;

4) для любых различных вершин u, v существует единственная простая (u, v)-цепь;

5) в графе нет циклов, но добавление любого ребра дает граф с единственным циклом.

Доказательство. 1) ⇒ 2). Для малых n утверждение очевидно. Предположим, что оно верно для графов с числом вершин, меньшим n, и рассмотрим связный граф с n вершинами, не содержащий циклов. Удалив из него произвольное ребро (u, v) (в силу предложения 1 оно является мостом), получим граф с двумя компонентами, в каждой из которых вершин меньше, чем n. По предположению индукции эти компоненты содержат, соответственно, $n_1 - 1$ и $n_2 - 1$ рёбер, где n_1 и n_2 — количества вершин в компонентах. Тогда общее количество рёбер в исходном графе равно $(n_1 - 1) + (n_2 - 1) + 1 = n - 1$.

2) ⇒ 3). Допустим, что граф содержит цикл. Тогда, удалив какое-либо ребро этого цикла, получим связный граф с $n - 2$ рёбрами, что противоречит предложению 2.

3) ⇒ 4). Докажем сначала, что граф связный. Пусть у него k компонент, содержащих $n_1, \ldots, n_k$ вершин. Каждая компонента является деревом и по импликации 1) ⇒ 2), доказанной выше, i-я компонента содержит $n_i - 1$ рёбер, значит, всего рёбер в графе

$$(n_1 - 1) + \ldots + (n_k - 1) = n - k = n - 1.$$

Следовательно, $k = 1$ и граф связный. Согласно лемме 1 §1 для любых неравных вершин u, v существует простая (u, v)-цепь, а её единственность сразу следует из леммы 3 §1.

4) ⇒ 5). В графе нет циклов, поскольку наличие цикла по лемме 2 влечёт и наличие простого цикла, а в простом цикле любые две вершины, очевидно, соединены двумя простыми цепями. Легко также видеть, что добавление любого нового ребра дает простой цикл.

Наконец, единственность получающегося простого цикла следует из того, что в противном случае вершины, инцидентные добавленному ребру, соединены в исходном графе двумя простыми цепями, что противоречит условию 4).

5) $\Rightarrow$ 1). Если добавление нового ребра дает простой цикл, значит, вершины, соединенные этим ребром, были соединены в исходном графе простой цепью. Следовательно, граф связный. $\square$

2. Остовы наименьшего веса. Алгоритм Краскала. Одна из причин популярности деревьев заключается в том, что любой связный граф содержит в качестве подграфа дерево с тем же множеством вершин, что и сам граф. Это дерево называется *остовом* графа. В существовании остова можно убедиться, последовательно удаляя рёбра, принадлежащие циклам. Ясно, что граф, не являющийся деревом, имеет несколько остовов.

Рассмотрим следующую задачу. Имеется несколько городов, которые нужно соединить между собой сетью дорог. Для каждой пары городов известна стоимость строительства соединяющей их дороги. Требуется построить сеть дорог, имеющую минимальную суммарную стоимость. Эта и подобные ей практические задачи допускают следующую математическую формулировку в терминах графов.

Дан граф, каждому ребру которого приписано положительное число — вес. *Весом остова* назовём сумму весов его рёбер. Требуется найти остов графа, имеющий наименьший вес. Наиболее известен следующий алгоритм решения этой задачи, обычно именуемый алгоритмом Краскала.

Построим графы $T_0, T_1, \ldots, T_{n-1}$ следующим образом. Возьмём в качестве T_0 пустой граф с множеством вершин V. Добавив ребро наименьшего веса, получим граф T_1. Ясно, что у него нет циклов. Пусть уже построен лес T_i с i рёбрами ($i < n-1$). Он несвязный по предложению 2.1. Выберем ребро наименьшего веса из тех, которые не образуют циклов с рёбрами T_i (таковым является любое ребро, соединяющее различные компоненты графа T_i). Добавив выбранное ребро к T_i, получим T_{i+1}. Продолжаем, пока не получим граф T_{n-1}. У этого графа по построению $n-1$ ребер и нет циклов. По теореме 1 граф T_{n-1} — дерево (остов исходного графа).

Теорема 2. *Остов T_{n-1} имеет минимально возможный вес.*

Доказательство. Пусть D — любой другой остов графа. Докажем, что вес T_{n-1} не больше веса D. Запишем последовательность рёбер $e_1, e_2, \ldots, e_{n-1}$ остова T_{n-1}, выбранных алгоритмом Краскала. Пусть e_k — ребро с наименьшим номером, не принадлежащее D. Добавив его к D, получим граф с единственным простым циклом (см. теорему 1). В этом цикле найдется ребро e', не принадлежащее T_{n-1}. Докажем, что вес e' не меньше, чем вес e_k. Действительно, допустим, что вес e' строго меньше, чем вес e_k. Но тогда на k-м шаге алгоритм Краскала не должен был выбрать e_k, поскольку есть более лёгкое ребро e', не образующее циклов с рёбрами $e_1, e_2, \ldots, e_{k-1}$. Итак, вес e' не меньше, чем вес e_k. Поэтому, удалив ребро e', мы получим новый остов D', который имеет с остовом T_{n-1} больше общих рёбер, чем D, а его вес не больше, чем вес D. Если $D' \neq T_{n-1}$, то повторим вышеописанную процедуру замены ребра и получим дерево D'', которое имеет ещё больше общих рёбер с D и вес его не больше, чем вес D'. Продолжая таким образом, получим последовательность остовов

$$D, D', D'', \ldots, D^{(m)} = T_{n-1},$$

веса которых образуют невозрастающую последовательность. Следовательно, вес T_{n-1} не больше, чем вес D, что и завершает доказательство. □

Пример 1. Требуется найти остов наименьшего веса в графе, изображенном на рис. 7:

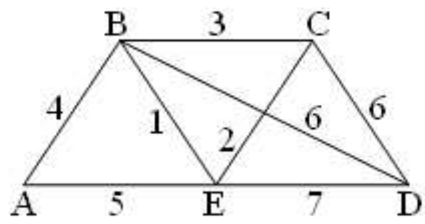


Рис. 7

Следуя алгоритму Краскала, выбираем ребро BE наименьшего веса 1. Среди оставшихся рёбер наименьший вес 2 имеет ребро CE. Следующее по весу ребро BC образует цикл с ранее выбранными рёбрами, поэтому пропускаем его и берём ребро AB веса 4. Ребро AE веса 5 образует цикл с уже выбранными рёбрами AB и BE, следовательно, его брать нельзя. Легко видеть, что любое из рёбер BD и CD веса 6 подходит для последнего шага. Получившиеся остовы наименьшего веса $1 + 2 + 4 + 6 = 13$ изображены на рисунке 8.

§ 4. Листы и деревья

Как следует из предложения 1 §3, дерево — это связный граф, каждое ребро которого является мостом. Противоположным свойством обладает *лист*. Связный граф называется листом, если у него

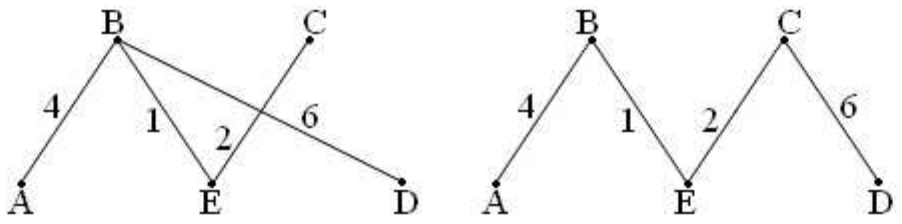


Рис. 8

нет мостов, то есть, каждое ребро принадлежит циклу. Эйлеров граф — это, конечно, лист. Нетрудно показать, что и гамильтонов граф — лист. Но некоторые листы не эйлеровы и не гамильтоновы.

Мы покажем, что всякий связный граф есть, образно говоря, дерево с листами (а то, что мы назвали деревом, это, скорее, дерево зимой, без листьев).

Назовём вершины u, v графа *циклически-рёберно связанными*, если $u = v$, или существует (u, v)-маршрут, не содержащий мостов. Циклически-рёберная связанность есть, очевидно, отношение эквивалентности. Классы этой эквивалентности порождают подграфы, являющиеся листами. Будем называть эти подграфы, как и сами классы, листами графа. Докажите следующее несложное

Предложение 1. *Каждый мост графа соединяет вершины из разных листов, причём, если два листа соединены, то лишь одним мостом.*

В параграфе 1 введено понятие подграфа, и, в частности, подграфа, порождённого подмножеством вершин. Введём понятие, в некотором смысле двойственное понятию подграфа.

Пусть задано разбиение множества вершин графа G. Определим *факторграф* графа G по данному разбиению следующим образом. Вершины факторграфа — это классы разбиения, два класса считаются смежными вершинами, если какие-нибудь вершины из этих классов смежны в G.

Упражнение 1. Что представляет собой факторграф по разбиению на компоненты?

Некоторые свойства графа наследуются факторграфом. Например, нетрудно проверить, что справедливо

Предложение 2. *Факторграф связного графа по любому разбиению — связный граф.*

Основной результат этого параграфа:

Теорема 1. *Факторграф связного графа по разбиению на листы есть дерево.*

Д о к а з а т е л ь с т в о. Вследствие предложения 2 нужно лишь убедиться, что факторграф не содержит циклов. Предположим противное: существует цикл листов $L_1, \ldots, L_k, L_1$ и пусть $e_1, \ldots, e_k$ — мосты исходного графа, последовательно соединяющие эти листы. Тогда, ввиду связности листов, в исходном графе должен быть цикл, содержащий эти мосты. Но это противоречит предложению 2. □

Пример 1. На рис. 9 изображены граф с семью листами и его факторграф по разбиению на листы.

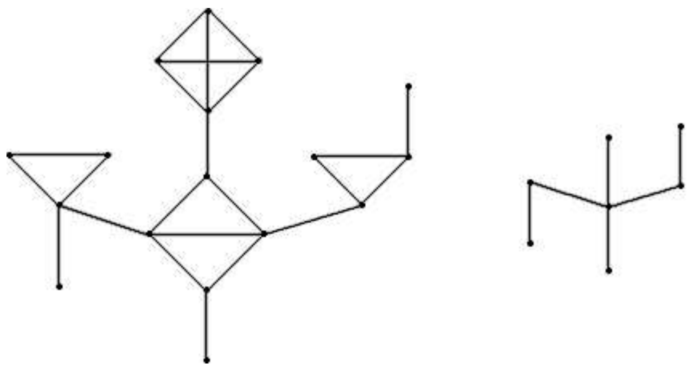


Рис. 9

Замечание 1. Говоря неформально, факторграф является упрощённой моделью графа. В зависимости от выбора разбиения эта модель может содержать бо̀льшую или меньшую информацию об исходном графе. Полезно сравнить в этом отношении факторграфы из упражнения 1 и из теоремы 1.

§ 5. Теорема о свадьбах, двудольные графы и (0,1)-матрицы

1. Теорема о свадьбах. Одна из наиболее известных задач дискретной математики часто формулируется в шутливой форме: имеется множество юношей и множество девушек, и для любых юноши и девушки известно, знакомы ли они. Задача: можно ли женить каждого юношу на знакомой ему девушке? Пусть количество юношей равно m, девушек — n.

Теорема 1 (Ф. Холл, 1935). *Задача о свадьбах разрешима тогда и только тогда, когда любые k юношей ($k = 1, \ldots, m$) знакомы в совокупности не менее, чем с k девушками.*

Доказательство. Необходимость почти очевидна: если некоторые k юношей знакомы в совокупности меньше, чем с k девушками, то им не хватит невест и задача в целом неразрешима. Достаточность доказывается индукцией по числу юношей. При $m = 1$ утверждение очевидно. Предположим, что оно верно для числа юношей, меньшего m, и докажем его для m юношей. Возможны два случая.

1. Условие теоремы выполнено "с избытком": любые k юношей ($k \leqslant m-1$) знакомы не меньше, чем с $k+1$ девушками. Тогда женим какого-нибудь юношу на знакомой ему девушке. Для остальных $m-1$ юношей и $n-1$ девушек условие теоремы выполнено и по предположению индукции их можно женить на знакомых девушках.

2. Имеется k юношей ($k \leqslant m-1$), знакомых в совокупности ровно с k девушками. По предположению индукции этих юношей можно женить. Если мы теперь докажем, что любые p из остальных $m-k$ юношей знакомы не меньше, чем с p из остальных $n-k$ девушек, то и для них по предположению индукции задача будет разрешимой. Допустим противное: среди оставшихся есть p юношей, знакомых в совокупности лишь с $q < p$ девушками из оставшихся. Но тогда во всём множестве юношей имеется $k+p$ юношей, знакомых лишь с $k+q$ девушками, причем $k+q < k+p$. Это противоречие завершает доказательство теоремы. $\square$

В следующих параграфах показано, что задача о свадьбах и теорема Холла допускают формулировку и в других, на первый взгляд весьма отличных, терминах. При этом математическое их содержание сохраняется. Заметим, что каждая из этих формулировок имеет хождение в своей области дискретной математики не может быть сочтена "лишней".

2. Паросочетания в двудольных графах. Граф называется *двудольным*, если его множество вершин V можно так разбить на два подмножества V_1 и V_2, что любое ребро графа соединяет вершины из разных подмножеств. *Совершенным паросочетанием* из V_1 в V_2 в двудольном графе называется такая инъекция $\varphi : V_1 \to V_2$ что вершины v и $\varphi(v)$ смежны для всех $v \in V_1$. Вопрос: при каких условиях существует совершенное паросочетание? Выведите из теоремы Холла, что оно существует тогда и только тогда, когда любые k вершин из V_1 имеют в совокупности не меньше, чем k смежных вершин.

3. Системы различных представителей. Рассмотрим семейство $S_1, \ldots, S_m$ подмножеств конечного множества E. Говорят, что множество $\{s_1, \ldots, s_m\} \subseteq E$ является *системой различных представителей* (с.р.п.) для данного семейства, если $s_i \in S_i$, $i = 1, \ldots, m$.

Вопрос: существует ли с.р.п.? Ответ: существует тогда и только тогда, когда объединение любых k подмножеств ($k = 1, \ldots, m$) содержит не менее k элементов. Как это следует из теоремы Холла?

4. Теорема Фробениуса–Кёнига и перманенты. В этом пункте $A = (a_{ij})$ — $m \times n$-матрица с элементами 0 или 1. Любой набор элементов $a_{1j_1}, a_{2j_2}, \ldots, a_{mj_m}$, где вторые индексы попарно различны, называется *диагональю* матрицы. Если все элементы диагонали равны 1, то говорят, что диагональ *положительна.* Попросту говоря, положительная диагональ — это набор m единичных элементов матрицы, любые два из которых стоят в разных строках и столбцах.

Лемма 1. *Матрица A содержит положительную диагональ тогда и только тогда, когда любые k строк ($k = 1, 2, \ldots, m$) матрицы A образуют матрицу, имеющую не менее k ненулевых столбцов.*

Доказательство. Любая (0,1)-матрица A задает задачу о свадьбах, если считать, что юноши перенумерованы числами $1, \ldots, m$, девушки — числами $1, \ldots, n$, и равенство $a_{ij} = 1$ означает, что i-й юноша знаком с j-й девушкой. Положительная диагональ $a_{1j_1}, a_{2j_2}, \ldots, a_{mj_m}$ определяет решение задачи: 1-го юношу женим на j_1-й девушке и так далее. И наоборот, всякое решение задачи о свадьбах определяет положительную диагональ. Заметим, что j-я девушка имеет знакомого в множестве $\{i_1, \ldots, i_k\}$ юношей тогда и только тогда, когда j-й столбец матрицы, образованной строками $i_1, \ldots, i_k$ матрицы A, содержит единицу. Так что у этих k юношей в совокупности ровно столько знакомых девушек, сколько ненулевых столбцов в соответствующей матрице. Теперь уже ясно, что наша лемма — переформулировка теоремы Холла. $\square$

Лемма 2. *Следующие свойства матрицы A эквивалентны:*

1) любая подматрица, составленная из k строк ($k = 1, \ldots, m$), имеет не меньше, чем k ненулевых столбцов;

2) сумма размеров любой нулевой подматрицы не превышает n.

Доказательство. Если верно 1), то любая нулевая подматрица, расположенная в k строках, имеет не больше, чем $n - k$ столбцов. Следовательно, сумма её размеров не больше, чем $k + (n - k) = n$. Если же 1) не выполнено и подматрица, составленная из некоторых k строк, имеет $l < k$ ненулевых столбцов, то в этих же строках расположена нулевая подматрица с суммой размеров $k + (n - l) > n$. $\square$

Из лемм 1 и 2 сразу следует

Теорема 2 (Фробениус–Кёниг). *Пусть A — $(0,1)$-матрица размеров $m \times n$ $(m \leqslant n)$. Она содержит положительную диагональ тогда и только тогда, когда сумма размеров любой её нулевой подматрицы не больше n.*

Перманентом матрицы $A = (a_{ij})$ размеров $m \times n$, $m \leqslant n$, называется число

$$\operatorname{per} A = \sum a_{1j_1} a_{2j_2} \dots a_{mj_m},$$

где суммирование ведется по всевозможным последовательностям $\{j_1, \dots, j_m\}$ различных элементов из множества $\{1, 2, \dots, n\}$. Докажите, что перманент (0,1)-матрицы равен количеству положительных диагоналей или, в "свадебных" терминах, — количеству решений задачи о свадьбах.

5. Граничный ранг (0,1)-матрицы размеров $m \times n$ определяется как максимальное число $\rho(A)$ единиц, любые две из которых стоят в разных строках и столбцах. Если считать, что матрица A определяет задачу о свадьбах, то её граничный ранг равен максимально возможному числу бракосочетаний, заключенных между знакомыми. Ясно, что $\rho(A) \leqslant \min(m, n)$. Определим ещё одну характеристику (0,1)-матриц.

Ранг покрытия $\mu(A)$ равен минимальному числу линий (строк и столбцов) матрицы, содержащих (покрывающих) все единицы матрицы A.

Мы докажем, что граничный ранг и ранг покрытия совпадают, но прежде приведем лемму, которую можно интерпретировать как условие разрешимости задачи о свадьбах для девушек. Она не требует отдельного доказательства.

Лемма 3. *Пусть A — $(0,1)$-матрица размеров $m \times n$, где $m \geqslant n$. В ней можно найти n единиц в попарно разных строках и столбцах тогда и только тогда, когда любые k столбцов A $(k = 1, \dots, n)$ образуют подматрицу, в которой ненулевых строк не меньше k.*

Теорема 3 (Кениг–Эгервари, 1931). *Для любой $(0,1)$-матрицы A граничный ранг и ранг покрытия совпадают: $\rho(A) = \mu(A)$.*

Д о к а з а т е л ь с т в о. Если матрица содержит ρ единиц, любые две из которых находятся на разных линиях (то есть, в разных строках и разных столбцах), то ясно, что для покрытия матрицы требуется не менее ρ линий. Поэтому $\rho(A) \leqslant \mu(A)$.

Докажем, что $\rho(A) \geqslant \mu(A)$. Пусть в минимальный список покрывающих линий входят r строк и s столбцов, причем можно считать (подумайте, почему), что строки занимают верхние r позиций, а столбцы — последние s позиций, так что матрица имеет вид

$$A = \begin{array}{r@{}c} \begin{array}{r} {\scriptstyle r} \\ {\scriptstyle m-r} \end{array} & \begin{pmatrix} B & * \\ 0 & C \end{pmatrix} \\ & \begin{array}{cc} {\scriptstyle n-s} & {\scriptstyle s} \end{array} \end{array}$$

Докажем, что матрица B содержит r единиц, любые две из которых стоят на разных линиях. Допустим противное. Тогда по лемме 1 существуют k строк матрицы B, все единицы которых стоят в l столбцах, где $l < k$. Удалив номера этих k строк из списка линий и добавив номера этих l столбцов, снова получим набор покрывающих линий в количестве $(r - k) + (s + l) = r + s - (k - l) < r + s = \mu$, что противоречит минимальности μ.

Теперь докажем, что подматрица C содержит s единиц в попарно разных линиях. Вследствие леммы 3 противное означало бы, что имеются k столбцов матрицы C, все единицы которых находятся в p строках, где $p < k$. Удалив номера этих k столбцов из списка покрывающих линий и добавив номера этих p строк, получим набор покрывающих линий числом, меньшим μ, — противоречие. Таким образом, доказано существование $r + s = \mu(A)$ единиц, любые две из которых стоят на разных линиях, значит, $\rho(A) \geqslant \mu(A)$. Учитывая полученное ранее противоположное неравенство, заключаем, что $\rho(A) = \mu(A)$. $\square$

Глава 2
Ориентированные графы

§ 1. Взаимодостижимость, компоненты и конденсация

Во многих теоретических и прикладных задачах полезно использовать графы, рёбра которых имеют направление. В связи с этим введём новый тип графов.

Ориентированным графом (орграфом) называется пара (V, E), где V — непустое множество вершин, $E \subseteq V \times V$ — множество *ориентированных рёбер* или *дуг*. Таким образом, дуга — это упорядоченная пара вершин. Если $e = (u, v)$, то вершина u — *начало* дуги e, вершина v — её *конец*. Говорят, что дуга e *выходит* из u и *входит* в v. Говорят также, что e *инцидентна* вершинам u и v. Дуга (v, v) называется *петлёй*. На рисунке дуги графа изображаются стрелками. Иногда в тексте вместо выражения "дуга ведёт из вершины i в вершину j" будем писать $i \to j$. Вершины орграфа *смежны*, если они соединены дугой.

Путём длины k в орграфе называют любую последовательность вершин
$$v_1, v_2, \ldots, v_{k+1},$$
такую, что (v_i, v_{i+1}) — дуга, $i = 1, \ldots, k$. Говорят, что v_1 — начало, v_{k+1} — конец пути. Когда хотят указать начало и конец пути, пишут: (v_1, v_{k+1})-путь.

Простой путь — это путь, в котором все вершины различны, кроме, возможно, первой и последней. Если $v_1 = v_{k+1}$, то путь (простой путь) называется *контуром* (*простым контуром*).

Приведём леммы, аналогичные леммам 1 и 2 из §1 гл.1. Их доказательства в ориентированном случае даже упрощаются. Рекомендуем доказать их самостоятельно.

Лемма 1. *Если существует (u, v)-путь, то существует и простой (u, v)-путь.*

Лемма 2. *Всякий контур содержит простой контур, причём каждая вершина и дуга контура принадлежат некоторому простому контуру.*

На орграфы естественным образом переносятся такие понятия, как подграф, в частности, подграф, порождённый множеством вершин, изоморфизм графов, факторграф. Например, факторграф графа G по разбиению множества вершин определяется так: его вершины — это классы разбиения, причём из класса S в класс T ведёт дуга, если $S \neq T$ и в исходном графе существует дуга с началом в S и концом в T.

Очевидным образом распространяется на орграфы понятие матрицы смежности и булевой матрицы смежности. После соответствующей замены терминов остаются верными предложения 1, 2 и 3 из §1 гл.1 и их доказательства. Советуем читателю убедиться в этом. Разумеется, матрица смежности орграфа уже не обязана быть симметричной и иметь нулевую диагональ, а может быть какой-угодно $(0, 1)$-матрицей.

Говорят, что вершина v *достижима* из вершины u, если $u = v$ или существует (u, v)-путь. Достижимые друг из друга вершины u и v называются *взаимодостижимыми.* Если любые две вершины орграфа взаимодостижимы, то он называется *сильно связным.* В противном случае граф называется *приводимым.* Взаимодостижимость — рефлексивное, симметричное и транзитивное бинарное отношение, то есть, отношение эквивалентности. Оно разбивает множество вершин графа на классы взаимодостижимых вершин. Класс взаимодостижимости, как и сильно связный подграф, им порождённый, будем называть *компонентой орграфа.*

Задача 1. Сформулируйте для орграфов утверждения задачи 2 §1 гл. 1 и убедитесь, что они остаются верными.

Конденсацией орграфа называется его факторграф по разбиению на компоненты.

Теорема 1. *Конденсация не содержит контуров.*

Д о к а з а т е л ь с т в о. Если предположить существование контура

$$S \to T \to \ldots \to U \to S,$$

то, очевидно, вершины различных компонент, входящих в этот контур, взаимодостижимы и, следовательно, лежат в одной компоненте. Противоречие. □

Замечание 1. Аналогом сильно связного орграфа является в неориентированном случае скорее лист, а не связный граф. Действительно, любая дуга сильно связного орграфа принадлежит контуру (докажите это), и любое ребро листа принадлежит циклу. Очевидно также сходство теоремы 1 и теоремы 1 из §4 гл.1.

Пример 1. Орграф отображения $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 5 & 2 & 5 & 1 \end{pmatrix}$ выглядит так:

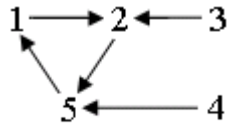


Рис. 1

На этом примере виден и общий принцип построения графа отображения множества в себя. Для нашего примера компоненты и конденсация таковы:

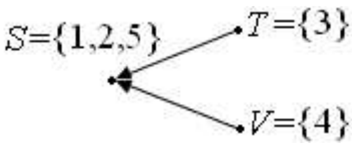


Рис. 2

Если отображение — биекция (перестановка), то граф распадается на простые контуры, его конденсация — пустой граф.

Пример 2.

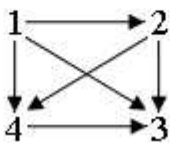


Рис. 3

На рис. 3 изображён пример *турнира*. Так называются орграфы, в которых любые две вершины соединены единственной дугой. В данном случае турнир не содержит контуров, и его конденсация совпадает с ним самим.

Пример 3.

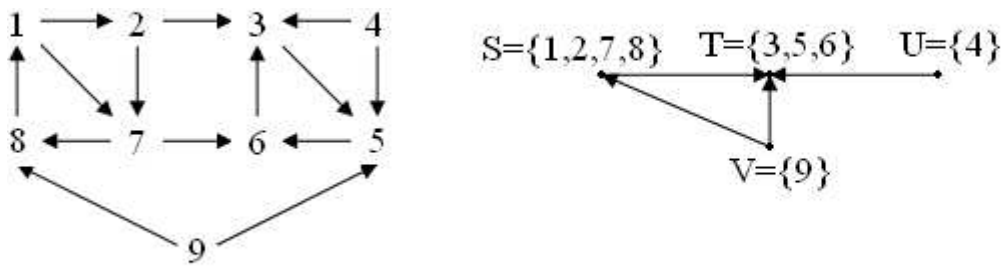


Рис. 4

При подходящей нумерации вершин матрица смежности отражает строение графа. Это верно и в неориентированном случае (см.

упр. 1 §1 гл. 1), но для орграфов нумерация имеет большее значение. Пример такой нумерации содержит следующее

Предложение 1. *Пусть орграф не содержит контуров. Тогда существует нумерация вершин, при которой дуга ведёт из вершины i в вершину j, только если $i > j$, то есть, матрица смежности — нижняя нильтреугольная.*

Доказательство. В орграфе без контуров существует *минимальная* вершина, то есть вершина, из которой не выходят дуги (подумайте, почему?). Присвоим минимальной вершине u номер 1. Рассмотрим множество вершин $V\setminus\{u\}$. Оно содержит вершину v, из которой не выходят дуги в другие вершины этого множества. Присвоим этой вершине номер 2. Затем по тому же принципу выделим в множестве $V\setminus\{u, v\}$ вершину w и так далее. В итоге получим нумерацию $u \mapsto 1, v \mapsto 2, \ldots, z \mapsto n$, которая, как нетрудно убедиться, обладает нужным свойством. □

Поскольку на компоненты орграфа мы можем смотреть как на вершины конденсации, то из теоремы 1 и предложения 1 вытекает

Теорема 2. *Существует такая нумерация компонент*

$$V_1, V_2, \ldots, V_s, \tag{1}$$

что если начало дуги орграфа лежит в V_i, а конец — в V_j, то $i > j$.

Компонента графа называется *минимальной*, если из неё не выходят дуги в другие компоненты. В списке (1) компонента V_1 — заведомо минимальная, но необязательно единственная минимальная компонента.

Задача 2. Назовём множество вершин орграфа *замкнутым*, если из него не выходят дуги в вершины, не принадлежащие этому множеству. Докажите, что непустое множество вершин орграфа образует минимальную компоненту тогда и только тогда, когда оно замкнуто и не содержит собственных замкнутых подмножеств.

Нумерацию компонент, описанную в теореме 2, назовём *нормальной*, если, дополнительно, минимальные компоненты перечислены в ней в первую очередь. Нумерацию вершин графа будем считать *нормальной*, если она согласована с некоторой нормальной нумерацией (1) компонент в том смысле, что при этой нумерации первые номера получают вершины из V_1, затем нумеруются вершины из V_2 и так далее.

Вершина графа называется *невозвратной*, если из неё есть путь в такую вершину, из которой она недостижима. Докажите самостоятельно

Предложение 2. *Невозвратные вершины — это в точности те вершины, которые принадлежат неминимальным компонентам.*

Назовём простой путь *максимальным*, если он не имеет простого продолжения. Докажите самостоятельно

Предложение 3. *Конец простого максимального пути всегда принадлежит минимальной компоненте.*

§ 2. Нормальная форма матрицы смежности приводимого графа

Квадратную матрицу A (булеву матрицу или матрицу над полем) называют *приводимой*, если она перестановочно подобна матрице вида

$$\begin{pmatrix} B & 0 \\ C & D \end{pmatrix}, \tag{1}$$

где B, D — квадратные матрицы. В противном случае говорят, что A — *неприводимая* матрица. Приводимая матрица A имеет *нормальную* форму, если она блочно-треугольная:

$$A = \begin{pmatrix} A_1 & & & & & \\ & \ddots & & & & \\ & & A_l & & & \\ A_{l+1,1} & \dots & A_{l+1,l} & A_{l+1} & & \\ \vdots & & \vdots & \vdots & \ddots & \\ A_{s,1} & \dots & A_{s,l} & A_{s,l+1} & \dots & A_s \end{pmatrix} \tag{2}$$

причём $A_1, A_2, \dots, A_s$ $(1 \le l \le s)$ — неприводимые матрицы, а в каждой строке, начиная с $(l+1)$-го (при $l < s$), имеются ненулевые блоки, расположенные левее диагонали.

Предложение 1. *Орграф сильно связен тогда и только тогда, когда его матрица смежности неприводима. Если граф приводим, то существует нумерация вершин, при которой его матрица смежности имеет нормальную форму.*

Доказательство. Допустим, что орграф сильно связен, но его матрица смежности приводима. Это значит, что при некоторой нумерации вершин графу соответствует матрица смежности вида (1). Но тогда выходит, что из вершин с номерами $1, \dots, k$ нельзя перейти в вершины с номерами $k+1, \dots, n$ (где k — порядок B), — противоречие с сильной связностью.

Теперь пусть граф приводим и $V_1, V_2, \ldots, V_s$ — нормальная нумерация компонент. Перенумеруем вершины графа, придав первые номера вершинам из V_1, а остальные вершины — произвольно. При такой нумерации, очевидно, матрица смежности примет форму (1) и тем самым доказана её приводимость. Если же продолжить нумерацию, придав очередные номера вершинам из V_2, и так далее (то есть, произвести нормальную нумерацию), то получим желаемую форму (2). $\square$

Как видно из предложения 1, при "правильной" нумерации вершин матрица графа наглядно отражает его строение. А именно, её диагональные блоки являются по существу матрицами смежности компонент, а блоки, лежащие левее диагонали, содержат информацию о дугах, идущих из компоненты с бóльшим номером в компоненты с меньшим номером.

Любая квадратная булева матрица A может рассматриваться как матрица смежности орграфа. Учитывая, что при новой нумерации вершин графа его матрица смежности A заменяется перестановочно подобной матрицей PAP^T, замечаем, что второе утверждение предложения 1 принимает следующий вид:

Следствие 1. *Любая приводимая булева матрица некоторым перестановочным подобием приводится к нормальной форме.*

Полученные выше сведения об орграфах и их матрицах смежности можно применять для анализа матриц над полем. Пусть $C = (c_{ij})$ — комплексная матрица порядка n. *Графом матрицы* C называют нумерованный орграф с n вершинами, в котором $i \to j \Leftrightarrow a_{ij} \neq 0$. *Портретом комплексной матрицы* C в вычислительной линейной алгебре называют (0,1)-матрицу $A = (a_{ij})$ такую, что $a_{ij} = 1 \ \Leftrightarrow c_{ij} \neq 0$. Ясно, что булевский портрет A матрицы можно понимать как матрицу смежности графа матрицы C.

Поскольку перестановочное подобие P лишь переставляет строки и столбцы, то нули в матрицах PAP^{-1} и PCP^{-1} стоят на одних и тех же местах.[1)] Следовательно, комплексная матрица неприводима в точности тогда, когда неприводим её портрет. А если она приводима, то перестановочное подобие P, преобразующее её портрет к нормальному виду, преобразует и саму комплексную матрицу к нормальному виду. Поэтому следствие 1 распространяется и на комплексные матрицы.

[1)] Разумеется, в первом случае перестановочную матрицу P нужно считать булевой, а во втором — комплексной матрицей, но мы не применяем разных обозначений, поскольку в обоих случаях производится одинаковая перестановка рядов.

Следствие 2. *Любая приводимая комплексная матрица некоторым перестановочным подобием приводится к нормальной форме.*

Полезность нормальной формы приводимой комплексной матрицы, в частности, состоит в том, что вычисление её собственных значений сводится к вычислению собственных значений диагональных блоков. Но наибольшее применение эта форма имеет в теории неотрицательных матриц (то есть, матриц с вещественными неотрицательными элементами) и теории цепей Маркова.

Это объясняется, например, следующим легко доказываемым фактом. Обозначим булевский портрет матрицы C символом $Sg(C)$.

Предложение 2. *Если P, Q — неотрицательные матрицы одного порядка, то*

$$Sg(P+Q) = Sg(P) + Sg(Q),$$

$$Sg(PQ) = Sg(P)Sg(Q), \text{ *следовательно,* } Sg(P^k) = (Sg(P))^k.$$

Таким образом, если мы желаем знать, как расположены ненулевые элементы в сумме (произведении) неотрицательных матриц, то для этого достаточно вычислить сумму (произведение) их булевских портретов.

Задача 1. Докажите, что в сильно связном турнире с $n \geq 3$ вершинами каждая вершина принадлежит контуру длины 3. Является ли конденсация турнира турниром?

Задача 2. Опишите орграфы, для которых матрица смежности A — нильпотентная матрица. Какова нормальная форма матрицы смежности в этом случае? В частности, для каких графов $A^2 = 0$?

Задача 3. Опишите орграфы, для которых булева матрица смежности A — идемпотентная матрица, то есть, $A^2 = A$. Какова нормальная форма матрицы смежности в этом случае?

§ 3. Арифметические свойства сильно связного графа. Циклические классы.

Пусть дан сильно связный орграф с вершинами $1, 2, \ldots, n$. Обозначим через N_{ij} множество длин всевозможных (i, j)-путей. В частности, N_{ii} — множество длин контуров, проходящих через вершину i. Пусть d_i — наибольший общий делитель чисел из N_{ii}.

Предложение 1. *$d_i = d_j$ для всех i, j.*

Д о к а з а т е л ь с т в о. Если $k \in N_{ii}$, $l \in N_{ij}$, $m \in N_{ji}$, то существуют контуры длины $k + l + m$ и $l + m$, проходящие через вершину j. Следовательно,

$$k + l + m \equiv 0 \ (mod\ d_j),\ l + m \equiv 0 \ (mod\ d_j),$$

и потому $k \equiv 0 \ (mod\ d_j)$. Мы доказали, что d_j делит произвольное число k из N_{ii}, значит, $d_j \leq d_i$. Меняя в этом рассуждении i и j местами, получаем $d_i \leq d_j$. Значит, $d_i = d_j$. $\square$

Итак, согласно доказанному, число $d = d_i$ равно НОД длин *всех* контуров графа и, одновременно, для любой вершины i равно НОД длин контуров, проходящих через i.

Замечание 1. Согласно лемме 1 §2 каждый контур является простым или содержит простые контуры, поэтому его длина равна сумме длин простых контуров. Отсюда следует, что *число d равно НОД длин простых контуров графа.*

Поскольку множество простых контуров конечно, это замечание существенно упрощает вычисление параметра d.

Предложение 2. *Для любых вершин i, j длины различных (i, j)-путей сравнимы по модулю d.*

Д о к а з а т е л ь с т в о. Пусть $k, l \in N_{ij}$, $m \in N_{ji}$. Тогда $k + m, l + m \in N_{ii}$, значит,

$$k + m \equiv 0 \ (mod\ d),\ l + m \equiv 0 \ (mod\ d) \Rightarrow k \equiv l \ (mod\ d). \ \square$$

Обозначим через t_{ij} $(0 \leq t_{ij} \leq d - 1)$ остаток от деления на d длины любого из (i, j)-путей.

Из элементарных свойств делимости целых чисел следует полезное сравнение

$$t_{ik} + t_{kj} \equiv t_{ij} \ (mod\ d). \tag{1}$$

Введём бинарное отношение $\sim$ на множестве вершин:

$$i \sim j \ \Leftrightarrow \ t_{ij} = 0, \tag{2}$$

то есть, $i \sim j$, если длины всех (i, j)-путей кратны d.

Предложение 3. *Отношение $\sim$ является отношением эквивалентности.*

Доказательство рекомендуем как упражнение.

Предложение 4. *Пусть $i \sim u$. Тогда* $t_{ij} = t_{uv} \Leftrightarrow j \sim v$.

Д о к а з а т е л ь с т в о. Из (1) выводится сравнение

$$t_{ij} + t_{jv} + t_{vu} \equiv t_{iu} \equiv 0 \ (mod\ d). \tag{3}$$

Если теперь $t_{ij} = t_{uv}$, то

$$t_{ij} + t_{vu} \equiv t_{uv} + t_{vu} \equiv t_{uu} \equiv 0 \ (mod\ d),$$

следовательно, $t_{jv} = 0$, то есть, $j \sim v$.

Обратно, подставляя $t_{jv} = 0$ в (3), получаем

$$t_{ij} + t_{vu} \equiv 0 \ (mod\ d),$$

а поскольку

$$t_{uv} + t_{vu} \equiv 0 \ (mod\ d),$$

то $t_{ij} = t_{uv}$. $\square$

На языке классов отношения $\sim$ предложение 4 означает, что класс, в который приводит путь, однозначно определяется классом, в котором лежит начало пути, а также остатком от деления длины пути на число d. Если эти остатки для двух путей различны, то пути приводят в различные классы. Отсюда следует, что имеется ровно d классов отношения $\sim$ и что переходы из класса в класс совершаются циклически. Произведём *нормальную* нумерацию классов: выберем произвольно некоторый класс и обозначим его C_0. Через C_1 обозначим тот класс, в который ведут дуги из C_0, через C_2 — класс, в который ведут дуги из C_1 и так далее. Из класса C_{d-1} все дуги ведут в C_0. Таким образом, доказана

Теорема 1. *Если НОД длин (простых) контуров сильно связного графа равен d, то множество вершин можно так разбить на d классов*

$$C_0, C_1, ..., C_{d-1},$$

что дуги с началом в C_0 ведут в C_1, дуги с началом в C_1 ведут в C_2 и т.д., наконец, дуги с началом в C_{d-1} ведут в C_0.

Классы эквивалентности отношения $\sim$ называются *циклическими* классами.

Упражнение 1. Докажите, что факторграф сильно связного графа по разбиению на циклические классы есть простой контур.

Разумеется, картина, описанная в теореме 1, содержательна лишь при $d > 1$. Число d циклических классов называется *индексом импримитивности* сильно связного графа. При $d > 1$ граф называется *импримитивным*, а при $d = 1$ — *примитивным.*

Предположим, что вершины графа разбиты на циклические классы, как описано в теореме 1. Перенумеруем вершины, присвоив первые номера вершинам из C_0, последующие — вершинам из C_1, и так далее. При такой нумерации матрица смежности получает вид, ясно отображающий циклический характер переходов из класса в класс:

$$A = \begin{pmatrix} 0 & A_{12} & 0 & \dots & 0 \\ 0 & 0 & A_{23} & \dots & 0 \\ .. & .. & .. & .. & .. \\ 0 & 0 & 0 & 0 & A_{d-1,d} \\ A_{d1} & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{4}$$

Блок A_{12} содержит информацию о переходах из C_0 в C_1, блок A_{23} — о переходах из C_1 в C_2 и так далее, наконец, блок A_{d1} сообщает о дугах, ведущих из C_{d-1} в C_0. Таким образом, справедливо

Предложение 5. *Для любого сильно связного графа существует такая нумерация вершин, при которой его матрица смежности имеет блочную форму (4), где блочный порядок d равен индексу импримитивности графа.*

Описанную выше нумерацию вершин сильно связного графа, отражающую их разбиение на циклические классы и характер переходов из класса в класс, будем называть *нормальной*, а о соответствующей матрице (4) смежности графа будем говорить, что она имеет *нормальный* вид (*нормальную* форму). Если граф примитивен, то будем считать, что его матрица смежности всегда находится в нормальной форме.

Если матрица смежности находится в нормальной форме, то и её степени имеют прозрачную блочную структуру. А именно, матрица A^k $(k = 1, 2, \dots, d-1)$ содержит ровно один ненулевой блок в каждой блочной строке и каждом блочном столбце. Ненулевые блоки занимают две блочные диагонали, которые сдвигаются вправо и вверх при увеличении степени, а при $k = d$ сливаются в одну диагональ — главную. Точнее, матрица A^d имеет блочно-диагональный вид:

$$A^d = \begin{pmatrix} A_{11}^{(d)} & & & \\ & A_{22}^{(d)} & & \\ & & \ddots & \\ & & & A_{dd}^{(d)} \end{pmatrix}, \text{где} \quad \begin{matrix} A_{11}^{(d)} = & A_{12}A_{23} & \dots & A_{d1}, \\ A_{22}^{(d)} = & A_{23} & \dots & A_{d1}A_{12}, \\ \dots & \dots & \dots & \dots \\ A_{dd}^{(d)} = & A_{d1}A_{12} & \dots & A_{d-1,d}. \end{matrix}$$

Нетрудно понять, что матрицы A^k и A^{d+k} имеют одну и ту же блочную структуру для любого показателя k. Блочные структуры степеней A меняются периодически с периодом d, так что блочная структура матрицы A^k, $k = ld + r$ $(0 \le r \le d-1)$, совпадает с блочной структурой A^r. Но что происходит внутри блоков? Чтобы выяснить это, продолжим изучение арифметических свойств сильно связного графа.

Два контура длины l и m, проходящие через вершину i, можно соединить в один контур длины $l+m$, проходящий через i. Следовательно, если $l, m \in N_{ii}$, то и $l+m \in N_{ii}$, то есть, N_{ii} образует полугруппу относительно сложения. Докажем полезное свойство таких числовых множеств.

Лемма 1. *Пусть множество M натуральных чисел содержит сумму любых двух своих элементов, и d — наибольший общий делитель чисел из M. Тогда M содержит все достаточно большие числа, кратные d.*

Замечание 2. Выражение "все достаточно большие числа" означает: все числа, начиная с некоторого числа.

Доказательство. Существуют такие числа $a_1, \ldots, a_k \in M$, что d равно их НОД. Из элементарной теории делимости известно[1], что НОД набора целых чисел можно представить в виде

$$d = c_1a_1 + c_2a_2 + \ldots + c_ka_k,$$

где $c_1, \ldots, c_k$ — подходящие целые числа. Сложив отдельно положительные и отрицательные члены этой комбинации, получим представление

$$d = b - c, \quad b, c \in M.$$

Теперь рассмотрим число h, кратное d и записанное в виде

$$h = qc + r, \quad 0 \le r \le c - 1.$$

Поскольку h и c делятся на d, то r тоже делится, значит $r = fd$, $0 \le f \le c - 1$. Следовательно,

$$h = qc + r = qc + fd = qc + f(b - c) = (q - f)c + fb.$$

Если $h \ge (c-1)c$, то $q \ge c-1$, а это значит, что $q - f \ge 0$ и $h \in M$. $\square$

Предложение 6. *Пусть индекс импримитивности графа равен d. Найдется такое число q_0, что при любом $q \geqslant q_0$ для любых вершин i, j существует (i,j)-путь длины $qd + t_{ij}$.*

[1] См., напр., Калужнин Л.А. Введение в общую алгебру. — М.: Наука, 1973.

Д о к а з а т е л ь с т в о . Согласно лемме 1 для каждого i найдется такое l_i, что $ld \in N_{ii}$ при всех $l \geqslant l_i$. Положим $l_0 = \max\limits_i l_i$. Тогда $ld \in N_{ii}$ при всех $l \geqslant l_0$ и любом i (напомним, что по предложению 1 НОД множеств N_{ii} совпадают и равны d).

Рассмотрим простой (i, j)-путь. Его длина может быть представлена в виде $q'd + t_{ij}$, где q' зависит от i, j, но ввиду простоты пути всегда не больше, чем n. Ясно, что при любом $q \geqslant l_0 + q'$ существует (i, j)-путь длины $qd+t_{ij}$. Учитывая неравенство $q' \leq n$, окончательно получаем: *для любых вершин* i, j и любого $q \geqslant q_0 = n + l_0$ существует (i, j)-путь длины $qd + t_{ij}$. □

Ниже значок $[i]$ употребляется для обозначения циклического класса, содержащего вершину i.

Теорема 2. *Существует такое число k_0, что для любого $k \geq k_0$ и любых классов $[i], [j]$ верно одно из двух утверждений:*

1) из любой вершины и класса $[i]$ можно перейти за k шагов в любую вершину v класса $[j]$,

2) ни из какой вершины $u \in [i]$ невозможен переход за k шагов ни в какую вершину $v \in [j]$.

Первое случится, если $k \equiv t_{ij}$ $(mod\ d)$, второе — в противоположном случае.

Д о к а з а т е л ь с т в о. Прежде всего заметим, что $t_{uv} = t_{ij}$ в силу предложения 4. Теперь для доказательства первого утверждения достаточно положить $k_0 = q_0 d$ и применить предложение 6. Второе утверждение вытекает непосредственно из определения числа t_{ij}. □

Замечание 3. В дальнейшем обозначение k_0 закрепим за наименьшим из чисел, существование которых доказано в теореме 2.

Рассмотрим последовательность степеней произвольной булевской матрицы A. Поскольку множество булевских матриц порядка n конечно, то существует наименьшее p с тем свойством, что $A^p = A^{p+r}$ для некоторого $r > 0$. Число p называется *предпериодом* матрицы A, а наименьшее r с указанным свойством — её *периодом.* Таким образом, последовательность степеней матрицы A имеет вид

$$A, ..., A^{p-1}|A^p, ..., A^{p+r-1}|A^p, ..., A^{p+r-1}|...$$

Вернёмся к рассмотрению последовательности степеней матрицы смежности. Мы остановились на том, что при нормальной нумерации вершин графа степени его матрицы смежности имеют периодически повторяющуюся блочную структуру. Теорема 2 в переводе на язык булевых матриц утверждает, что при $k \geq k_0$ каждый блок $A_{pq}^{(k)}$ матрицы A^k либо целиком состоит из единиц, либо целиком из нулей.[1)]

Матрицы такого типа полностью определяются расположением ненулевых блоков, для них совпадение блочных структур означает равенство самих матриц. Поэтому, начиная с A^{k_0}, в последовательности степеней матрицы A происходит периодическое повторение с периодом, равным d. Мы пришли к следующему выводу.

Теорема 3. *Пусть A — нормальная форма матрицы смежности сильно связного графа. Тогда при $k \geq k_0$ ненулевые блоки матрицы A^k целиком состоят из единиц. Число k_0 равно предпериоду матрицы A, а индекс импримитивности d графа — периоду этой матрицы.*

Поскольку любая неприводимая булева матрица A может рассматриваться как матрица смежности сильно связного графа, теорема 3 допускает следующую переформулировку.

Следствие 1. *Предпериод k_0 неприводимой булевой матрицы есть наименьшее число, такое, что для графа этой матрицы при $k \geq k_0$ верно утверждение теоремы 2. Период неприводимой булевой матрицы совпадает с индексом импримитивности графа этой матрицы.*

Дадим определение *нормальной формы неприводимой булевой матрицы*, не использующее явно графовых терминов. Скажем, что неприводимая булева матрица находится в нормальной форме, если она имеет блочный вид (4), причём ненулевые блоки в степенях этой матрицы, начиная с некоторой степени, целиком состоят из единиц.

Упражнение 2. Докажите, что неприводимая булева матрица блочного вида (4) находится в нормальной форме в точности тогда, когда её блочный порядок равен её периоду.

Как и в случае приводимых матриц, полезно распространить определение нормальной формы на комплексные матрицы и считать, что *неприводимая комплексная матрица имеет нормальную форму*, если её портрет имеет нормальную форму. Стандартным рассуждением, как в §2, обосновывается

Предложение 7. *Любая неприводимая булева (комплексная) матрица некоторым перестановочным подобием приводится к нормальной форме.*

Из теоремы 3 и предложения 2 §2 следует

[1)]Точнее, $A_{pq}^{(k)} = I \Leftrightarrow p + k \equiv q (mod\ d)$.

Предложение 8. *Если A — нормальная форма неприводимой неотрицательной матрицы, то, начиная с некоторого показателя k_0, ненулевые блоки в матрице A^k состоят из положительных чисел.*

Пример 1. Рассмотрим граф

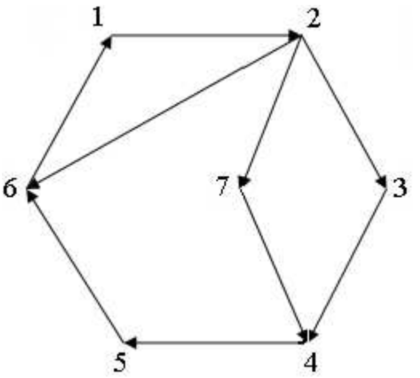


Рис. 5

Он содержит три простых контура длины 3, 6, 6, следовательно, число циклических классов d равно 3. Пользуясь определением (2), вычисляем циклические классы: $C_0 = \{1, 4\}$, $C_1 = \{2, 5\}$, $C_2 = \{3, 6, 7\}$. После перенумерации $1 \mapsto 1,\ 4 \mapsto 2,\ 2 \mapsto 3,\ 5 \mapsto 4,\ 3 \mapsto 5,\ 6 \mapsto 6,\ 7 \mapsto 7$ матрица смежности приобретает нормальную форму

$$A = \left(\begin{array}{cc|cc|ccc} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Непосредственные вычисления показывают, что её предпериод k_0 равен 6. Общий способ вычисления циклических классов и оценка сверху для предпериода k_0 произвольной булевой матрицы приведены в §5.

§ 4. Примитивные графы и матрицы

Примитивный граф определён как сильно связный граф, у которого НОД длин контуров равен единице. Разумеется, доказанные предыдущем параграфе теоремы верны и для примитивных графов, но их утверждения либо становятся тривиальными, либо приобретают настолько специальный вид, что заслуживают отдельных формулировок. Мы приведём эти формулировки, а читателям рекомендуем перечитать предыдущий параграф и обдумать, во что превращаются содержащиеся в нём определения и утверждения в случае, когда индекс импримитивности равен единице.

Предложение 1. *Граф примитивен тогда и только тогда, когда существует такое число q_0, что при $k \geqslant q_0$ из любой вершины можно перейти в любую другую вершину за k шагов.*

Неприводимая булева матрица называется *примитивной*, если её период равен 1. Другими словами, булева матрица примитивна, если примитивен её граф. Простейший пример — матрица I. Матрица с неотрицательными элементами называется примитивной, если примитивен её булевский портрет, или, эквивалентно, если примитивен её граф.

Предложение 2. *Булева матрица примитивна тогда и только тогда, когда её степени, начиная с некоторой, равны I.*

Предложение 3. *Если A — нормальная форма булевой матрицы с периодом d, то диагональные блоки матрицы A^d — примитивные матрицы.*

Последние два предложения очевидным образом переформулируются для неотрицательных матриц.

§ 5. Некоторые алгоритмические вопросы

В этом параграфе для сильно связного графа
1) приводится способ вычисления циклических классов,
2) даются оценки параметра k_0.

Предложение 1. *Вершины i, j тогда и только тогда принадлежат одному циклическому классу, когда существуют пути одинаковой длины из этих вершин в одну и ту же вершину или, как мы будем говорить, из них синхронно достижима одна и та же вершина.*

Д о к а з а т е л ь с т в о. Если вершины i, j лежат в одном классе, то согласно предложению 6 §3 при достаточно большом l существуют (i, i)-путь и (j, i)-путь длины ld.

Обратно, если вершины i, j принадлежат разным циклическим классам, то из них синхронно можно перейти лишь в разные классы, следовательно, в разные вершины. Это видно из циклического характера движения из класса в класс, описанного в теореме 1 §3. □

Лемма 1. *Если из двух вершин можно синхронно перейти в одну вершину за k шагов, то это можно сделать и за $k+1$ шагов.*

Доказательство. Пути длины k, ведущие в общую вершину, можно продолжить, добавив к ним любую дугу. □

Переведём лемму 1 на язык булевой матрицы смежности A. Наличие путей длины k из вершин i, j в некоторую общую вершину записывается так:

$$a_{il}^{(k)} = a_{jl}^{(k)} = 1 \text{ для некоторого } l, \text{ или, равносильно, } (A^k(A^k)^T)_{ij} = 1.$$

Введём сокращение $A^k(A^k)^T = B_k$. Лемма 1 тогда означает, что $(B_k)_{ij} = 1 \Rightarrow (B_{k+1})_{ij} = 1$. Выразим это свойство как неравенство $B_k \leq B_{k+1}$. Рассмотрим цепочку неравенств

$$B_1 \leq \ldots \leq B_k \leq \ldots .$$

В этой цепочке каждая матрица получается из предыдущей по правилу $B_{k+1} = AB_kA^T$. Поэтому как только в цепочке встретятся равные матрицы, то и дальше будут только равенства. А поскольку матрицы B_k — симметричные и с единицами на главной диагонали, то различных матриц в цепочке не больше, чем $\frac{n(n-1)}{2}$. Отсюда вытекает критерий, позволяющий вычислить циклические классы с помощью умножения булевых матриц.

Предложение 2. *Вершины i, j сильно связного графа с матрицей смежности A принадлежат одному циклическому классу тогда и только тогда, когда (ij)-элемент матрицы $A^{\frac{n(n-1)}{2}}(A^{\frac{n(n-1)}{2}})^T$ равен 1.*

Перейдём ко второму вопросу. Как видно из теоремы 3 §3, параметр k_0 сильно связного графа с матрицей смежности A равен предпериоду последовательности $A, A^2, \ldots$. Ясно, что для любой матрицы порядка n это число не больше числа 2^{n^2} всех булевых $n \times n$-матриц. Но это — грубая оценка. Известна более точная оценка:

Предложение 3. *Для любой булевой матрицы порядка n*

$$k_0 \leq \frac{n^2}{d} - 2n + 3d.$$

Достаточно сложное доказательство этого неравенства мы не приводим, однако выведем оценку для k_0 в важном частном случае, когда матрица примитивна.

Предложение 4. *Если матрица A порядка n примитивна, то*

$$k_0 \leq s(n-2)+n,$$

где s — длина самого короткого контура в графе матрицы A.

Доказательство. Для любой вершины i в графе матрицы A существует путь длины $n-s$ из i в некоторую вершину j кратчайшего контура. Теперь рассмотрим примитивную матрицу A^s. В графе этой матрицы вокруг вершины j есть петля. Используя при необходимости эту петлю, построим путь длины $n-1$ из j в произвольно выбранную вершину l. В графе A ему соответствует (j,l)-путь длины $s(n-1)$. Соединяя упомянутые пути в графе матрицы A, получаем (i,l)-путь длины $(n-s)+s(n-1) = s(n-2)+n$ для произвольно взятых вершин i,l. $\square$

Итак, если булева матрица A примитивна, то $A^{n^2-2n+2} = I$. Выведите самостоятельно два следствия из предложения 4.

Следствие 1. *Для любой примитивной матрицы порядка n*

$$k_0 \leq n^2-2n+2.$$

Следствие 2. *Если на диагонали примитивной матрицы порядка n есть единица, то*

$$k_0 \leq 2n-2.$$

§ 6. Цепи Маркова

Представим себе случайную динамическую систему, состояния которой изменяются по следующему закону: если в некоторый момент времени система находится в состоянии u, то в следующий момент она переходит в состояние v с вероятностью, которая зависит лишь от u и v и не зависит от предыдущих состояний.

Переходя на более точный язык, назовём *марковской системой* пару (S,p), где S — конечное непустое множество, а p — такая функция на множестве $S \times S$ с неотрицательными значениями, что $\sum_{v \in S} p(u,v) = 1$ для всех $u \in S$. Число $p(u,v)$ понимается как вероятность перехода системы из состояния u в состояние v за один шаг. Графом марковской системы назовём орграф с множеством S вершин, в котором дуга $u \to v$ существует, если $p(u,v) > 0$.

Если считать, как обычно делают, что $S = \{1,2,\ldots,n\}$, то марковские системы вполне описываются *стохастическими матрицами* переходных вероятностей. Неотрицательная матрица $P = (p_{ij})$ порядка n называется стохастической, если

$$\sum_{j=1}^{n} p_{ij} = 1,\ i = 1, \ldots, n. \tag{1}$$

Упражнение 1. Докажите, что произведение стохастических матриц и любая степень стохастической матрицы — стохастические матрицы.

Полезно следующее образное представление: по графу стохастической матрицы блуждает частица, переходя из вершины i в вершину j с вероятностью p_{ij}. Больше того, можно описать марковскую систему без явного задания стохастической матрицы, с помощью *стохастического графа*, то есть графа, каждой дуге $i \to j$ которого приписано неотрицательное число p_{ij}, причём выполняются равенства (1).

Пример 1. Марковские системы, определяемые стохастическими матрицами

$$\begin{pmatrix} \alpha & \beta & 0 \\ 0 & \alpha & \beta \\ \beta & 0 & \alpha \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 & 0 \\ \beta & 0 & \alpha & 0 \\ 0 & \beta & 0 & \alpha \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad (0 < \alpha, \beta < 1,\ \alpha + \beta = 1),$$

описываются также стохастическими графами, изображёнными на рис. 6 и 7.

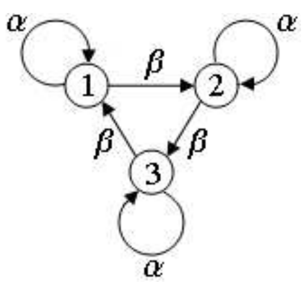


Рис. 6

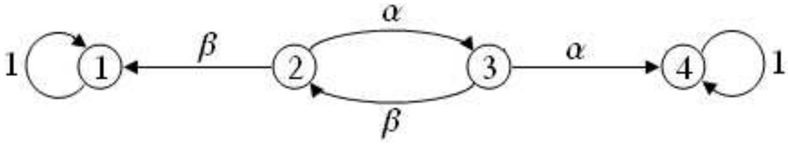


Рис. 7

Предположим, что в начальный момент времени система находится в состоянии i. В рамках описываемой модели произведение

$$p_{il_1}p_{l_1l_2}\dots p_{l_{k-1}j}$$

понимается как вероятность того, что система (частица), находясь вначале в состоянии (вершине) i, пройдёт путь $i, l_1, \dots, l_{k-1}, j$. Тогда число

$$\sum_{l_1,\dots,l_{k-1}} p_{il_1}p_{l_1l_2}\dots p_{l_{k-1}j}, \tag{2}$$

где суммирование производится по всевозможным (i,j)-путям длины k, равно вероятности перехода системы из состояния i в состояние j за k шагов.

Случайные процессы такого типа впервые изучались А.А. Марковым в начале 20-го века и по традиции называются цепями Маркова.[1)]

Цепь Маркова определяется, стало быть, если фиксировано начальное состояние системы. Можно сказать, что марковская система есть совокупность марковских цепей, которые получаются при различном выборе начального состояния системы.

Докажите простой, но решающий для применимости стохастических матриц в теории цепей Маркова факт:

Предложение 1. *Вероятность (2) перехода из состояния i в состояние j за k шагов равна $p_{ij}^{(k)}$, то есть, равна (ij)-элементу матрицы P^k.*

Иногда бывает нужно знать вероятность перехода за k шагов из состояния i в какое-либо из состояний множества α. Из предложения 1 вытекает, что эта вероятность равна $\sum_{j\in\alpha} p_{ij}^{(k)}$.

Мы рассмотрим, в качестве приложения теории графов, одну из основных задач теории цепей Маркова — задачу об асимптотическом поведении переходных вероятностей $p_{ij}^{(k)}$ при $k\to\infty$. Для двух типов марковских систем будет доказано существование пределов $\lim\limits_{k\to\infty} p_{ij}^{(k)}$ для любых состояний i, j. Другими словами, будет доказано, что последовательность $P^k, k = 1, 2, \dots$ поэлементно сходится к некоторой предельной матрице P^∞ (очевидно, стохастической).

Соответственно нашему подходу состояния марковской системы отождествляются с вершинами графа этой системы, понятия компоненты, циклического класса и т.п. переносятся на состояния системы.

1) Для углублённого знакомства см., например, Кемени Дж., Снелл Дж. Конечные цепи Маркова. — М.: Наука, 1970, или Ширяев А.Н. Вероятность (в 2-х кн.). — М.: МЦНМО, 2004.

1. Поглощающие марковские системы Нормальная форма матрицы приводимой марковской системы имеет вид

$$\begin{pmatrix} P_1 & & & \\ & \ddots & & \\ & & P_m & \\ Q_1 & \dots & Q_m & R \end{pmatrix}. \tag{3}$$

Матрица P_l состоит из вероятностей перехода внутри l-ой минимальной компоненты системы, матрица Q_l — из вероятностей перехода из множества невозвратных состояний в состояния l-ой минимальной компоненты, а матрицу R составляют вероятности перехода внутри класса невозвратных состояний.

Рассмотрим частный, но важный случай *поглощающих* марковских систем. Марковская система называется называется поглощающей, если её минимальные компоненты одноэлементны. Как видно, свойство быть поглощающей системой есть свойство графа системы. Дадим описание таких систем, не использующее явно понятие минимальной компоненты. Вначале определим *поглощающее* состояние марковской системы как такое состояние i, из которого нельзя выйти, то есть, $p_{ii} = 1$. На графе системы поглощающее состояние выделяется тем, что из него выходит единственная дуга — петля.

Упражнение 2. Поглощающая марковская система — это система, в которой имеются поглощающие состояния, причём из любого состояния достижимо поглощающее состояние.

Нормальная форма матрицы для поглощающей системы имеет совсем простой вид: блоки $P_1, \dots, P_m$ — это единичные матрицы порядка 1, а блоки $Q_1, \dots, Q_m$ представляют собой столбцы. Объединив их в матрицу Q, получим матрицу вида

$$P = \begin{pmatrix} E & 0 \\ Q & R \end{pmatrix}, \text{ соответственно, } P^k = \begin{pmatrix} E & 0 \\ Q^{(k)} & R^k \end{pmatrix}. \tag{4}$$

Для вероятностей перехода за k шагов из невозвратных состояний в поглощающие состояния, составляющих матрицу $Q^{(k)}$, имеет место формула

$$Q^{(k)} = (\sum_{h=0}^{k-1} R^h)Q, \tag{5}$$

которая доказывается индукцией по k.

Рассмотрим функцию $\|A\| = \max_{i=1}^n \sum_{j=1}^n |a_{ij}|$, определённую на комплексных матрицах порядка n. Она является матричной нормой и потому обладает кольцевым свойством $\|AB\| \le \|A\|\|B\|$ и, следовательно, свойством $\|A^k\| \le \|A\|^k$.[1)]

Лемма 1. *Если $\|A^q\| = \theta < 1$ для некоторого показателя q, то $A^k \to 0$ при $k \to \infty$.*

В формулировке леммы и дальше сходимость последовательности матриц понимается поэлементно: $A_k \to B$ означает, что $(A_k)_{ij} \to (B)_{ij}$ для всех i, j.

Доказательство. Из равенства $A^k = A^{lq}A^r$ $(0 \le r \le q-1)$ и кольцевого свойства следует, что $\|A^k\| \le \theta^l\|A^r\|$. Если $k \to \infty$, то $l \to \infty$ и обе части неравенства стремятся к 0. Тогда из определения функции $\|\cdot\|$ следует, что A^k сходится к нулевой матрице. □

Лемма 2. *Если $A^k \to 0$ при $k \to \infty$, то матрица $E - A$ обратима.*

Доказательство. Из условия леммы вытекает, что собственные значения $\lambda_1, \ldots, \lambda_n$ матрицы A по модулю меньше единицы, следовательно, среди собственных значений $1-\lambda_1, \ldots, 1-\lambda_n$ матрицы $E - A$ нет нулей, следовательно, эта матрица обратима. □

Теорема 1. *Пусть переходные вероятности поглощающей марковской системы заданы матрицами (4). Тогда предельные переходные вероятности существуют и выражаются матричными равенствами:*

$$\lim_{k\to\infty}\begin{pmatrix} E & 0 \\ Q^{(k)} & R^k \end{pmatrix} = \begin{pmatrix} E & 0 \\ Q^{(\infty)} & 0 \end{pmatrix}, \text{ где } Q^{(\infty)} = \lim_{k\to\infty} Q^{(k)} = (E-R)^{-1}Q.$$

Доказательство. Согласно предложению 3 §1 конец максимального простого пути с началом в невозвратном состоянии является поглощающим состоянием и, разумеется, единственно возможное продолжение этого пути состоит в повторении этого состояния. Поскольку длина простого пути не более $n-1$, то имеется положительная вероятность попадания за $n-1$ шагов в множество поглощающих состояний, в каком бы из невозвратных состояний ни стартовала система. Другими словами, вероятность остаться в классе невозвратных состояний после $n-1$ шагов меньше единицы, то есть, строчные суммы матрицы R^{n-1} меньше единицы. Таким образом, для матрицы R^{n-1} выполнено условие леммы 1 и мы можем заключить, что $R^k \to 0$ при $k \to \infty$.

[1)] О нормах см., например, Хорн Р., Джонсон Ч. Матричный анализ — М.:Мир,1989. Впрочем, кольцевое свойство этой функции нетрудно доказать и непосредственно.

Далее, из формулы (5) видно, что элементы матрицы $Q^{(k)}$ с ростом k могут лишь возрастать, а поскольку они, будучи вероятностями, ограничены сверху числом 1, то матрицы $Q^{(k)}$ сходятся при $k \to \infty$ к некоторому пределу $Q^{(\infty)}$.

Мы доказали, что $\lim\limits_{k\to\infty} P^k$ существует и имеет форму, указанную в теореме. Чтобы вывести формулу для $Q^{(\infty)}$, заметим, что равенство $PP^k = P^{k+1}$ при $k \to \infty$ переходит в равенство $PP^\infty = P^\infty$. Приравнивая левые нижние блоки этих блочных матриц, имеем $Q + RQ^{(\infty)} = Q^{(\infty)} \;\Rightarrow\; Q = (E-R)Q^{(\infty)}$. Поскольку матрица R удовлетворяет условию леммы 3, то из последнего равенства получаем $Q^{(\infty)} = (E-R)^{-1}Q$. □

Задача 1. Найдите предельные вероятности перехода для поглощающей марковской системы из примера 1.

Замечание 1. Результаты относительно поглощающих систем применимы к приводимым марковским системам общего вида. Поясним это применение неформально. Составим поглощающую систему, заменив в матрице (3) блоки $P_1, \ldots, P_m$ единицами, а блоки $Q_1, \ldots, Q_m$ — суммами столбцов этих блоков. Эта упрощённая (укрупнённая) модель отражает поведение исходной системы: вероятность перехода за k шагов из невозвратного состояния в i-ую компоненту равна вероятности для новой системы оказаться в i-ом поглощающем состоянии. Равны и соответствующие предельные вероятности.

2. Эргодические марковские системы. Марковская система с матрицей переходных вероятностей $P = (p_{ij})$ называется *эргодической*, если для всех i, j существуют положительные пределы

$$\lim_{k\to\infty} p_{ij}^{(k)} = \pi_j, \tag{6}$$

не зависящие от i.

Другими словами, марковская система с матрицей P переходных вероятностей — эргодическая, если последовательность P^k при $k \to \infty$ сходится к стохастической матрице вида

$$P^\infty = \begin{pmatrix} \pi_1 & \ldots & \pi_n \\ .. & \ldots & .. \\ \pi_1 & \ldots & \pi_n \end{pmatrix}$$

с положительными элементами.

Мы покажем, что наличие или отсутствие свойства эргодичности определяется графом матрицы. Но для полного исследования потребуются некоторые свойства стохастической матрицы как линейного оператора на пространстве $\mathbb{R}^n$ столбцов.

Введём *коэффициент эргодичности* стохастической матрицы. Вначале заметим, что для любых двух строк стохастической матрицы $P = (p_{ij})$ верны равенства

$$0 = \sum_j (p_{i_1 j} - p_{i_2 j}) = \sum_j{}^{+} (p_{i_1 j} - p_{i_2 j}) + \sum_j{}^{-} (p_{i_1 j} - p_{i_2 j}), \qquad (7)$$

где знаком "+" отмечена сумма положительных разностей, а знаком "−" — сумма отрицательных. Коэффициентом эргодичности стохастической матрицы $P = (p_{ij})$ называется число

$$\delta(P) = \max_{i_1, i_2} \sum_j{}^{+} (p_{i_1 j} - p_{i_2 j}).$$

Упражнение 3. Докажите,что если P — положительная стохастическая матрица, то $\delta(P) < 1$.

Пусть $x = (x_j)$ — столбец с вещественными элементами. Обозначим через $[x]$ отрезок $[\min_j x_j, \max_j x_j]$ вещественной прямой, а через $d(x)$ — длину этого отрезка.

Лемма 3. *Для любой стохастической матрицы* $P = (p_{ij})$
1) $[Px] \subseteq [x]$,
2) $d(P^k x) \le \delta(P^k) d(x), \ k = 1, 2, \ldots$.

Доказательство. 1) Действительно,

$$\max_i (Px)_i = \max_i (\sum_j p_{ij} x_j) \le (\sum_j p_{ij}) \max_j x_j = \max_j x_j.$$

Аналогично доказывается, что и левый конец отрезка $[Px]$ лежит в $[x]$.

Пункт 2) достаточно доказать для $k = 1$, общий случай получается индукцией по k. Пусть максимальный элемент столбца (Px) имеет номер i_1, а минимальный — i_2. Тогда

$$d(Px) = (Px)_{i_1} - (Px)_{i_2} = \sum_j p_{i_1 j} x_j - \sum_j p_{i_2 j} x_j =$$
$$= \sum_j (p_{i_1 j} - p_{i_2 j}) x_j = \sum_j{}^{+} (p_{i_1 j} - p_{i_2 j}) x_j + \sum_j{}^{-} (p_{i_1 j} - p_{i_2 j}) x_j \le$$
$$\le \sum_j{}^{+} (p_{i_1 j} - p_{i_2 j}) \max_j x_j + \sum_j{}^{-} (p_{i_1 j} - p_{i_2 j}) \min_j x_j = (\text{см. } (7))$$
$$= \sum_j{}^{+} (p_{i_1 j} - p_{i_2 j}) d(x) \le \delta(P) d(x). \ \square$$

Теорема 2. *Марковская система с матрицей переходных вероятностей $P = (p_{ij})$ тогда и только тогда является эргодической, когда граф цепи примитивен.*

Д о к а з а т е л ь с т в о. Если положительный предел $\lim\limits_{k\to\infty} P^k = P^\infty$ существует, то для достаточно больших k матрицы P^k обязаны быть положительными. Но это и означает, что граф цепи примитивен.

Обратно, пусть граф цепи примитивен и, следовательно, P — примитивная матрица. В силу п.1 леммы 3 для любого вещественного столбца x имеем последовательность вложенных отрезков

$$[x] \supseteq [Px] \supseteq ... \supseteq [P^k x] \supseteq \ \ldots\ .$$

Докажем, что эта последовательность стягивается в точку и, следовательно, элементы столбцов $P^k x$ сходятся при $k \to \infty$ к общему пределу.

Действительно, пусть m — показатель, для которого $P^m > 0$ (в силу следствия 1 предложения 5 §5 можно положить $m = n^2 - 2n + 2$). Согласно упражнению 3 имеем $\delta(P^m) < 1$, а применяя п.2 леммы 3, получаем

$$d(P^{ml}x) \leq \delta^l(P^m) d(x) \to 0 \text{ при } l \to \infty\,.$$

Таким образом, невозрастающая последовательность $d(P^k x)$ содержит подпоследовательность, стремящуюся к нулю. Значит, и вся последовательность сходится к нулю и желаемое доказано. Теперь положим $x = e_j$, где e_j — столбец с единицей на j-м месте и нулями на прочих. Столбец $P^k e_j$ равен j-му столбцу матрицы P^k, его элементы сходятся при $k \to \infty$ к некоторому общему пределу π_j. □

Вычисление вектора предельных вероятностей $\pi = (\pi_1, ..., \pi_n)$ основано на простых сведениях из линейной алгебры. Переходя к пределу в равенстве $P^k P = P^{k+1}$, получаем равенство $P^\infty P = P^\infty$, то есть, $\pi P = \pi$. Задача свелась к вычислению собственного вектора, отвечающего данному собственному значению. В нашем случае собственное подпространство, отвечающее собственному значению 1, одномерно (докажите это) и, следовательно, содержит единственный стохастический вектор.

На основе теоремы 1 легко понять асимптотику переходных вероятностей и в том случае, когда граф марковской системы сильно связен, но не примитивен, то есть, $d > 1$. Предположим, что матрица P системы находится в нормальной форме и рассмотрим матрицу

P^d :

$$P = \begin{pmatrix} 0 & P_{12} & 0 & \dots & 0 \\ 0 & 0 & P_{23} & \dots & 0 \\ .. & .. & .. & .. & .. \\ 0 & 0 & 0 & 0 & P_{d-1,d} \\ P_{d1} & 0 & 0 & 0 & 0 \end{pmatrix}, \ P^d = \begin{pmatrix} P_{11}^{(d)} & & \\ & \ddots & \\ & & P_{dd}^{(d)} \end{pmatrix}.$$

Стохастические матрицы $P_{11}^{(d)}, \dots, P_{dd}^{(d)}$ — см. §4, — примитивны. Рассматривая их как матрицы новых эргодических марковских систем, видим, что существует

$$\lim_{l\to\infty} P^{dl} = (P^d)^\infty = \begin{pmatrix} (P_{11}^{(d)})^\infty & & \\ & \ddots & \\ & & (P_{dd}^{(d)})^\infty \end{pmatrix}$$

с положительными равнострочными стохастическими матрицами на диагонали. Обозначим

$$((P^d)^\infty)_{jj} = \lim_{l\to\infty} \ p_{jj}^{(dl)} = \pi_j.$$

Число π_j равно предельной вероятности перехода в j из j (а также из любого состояния циклического класса $[j]$) через число шагов, кратное d. Существуют также пределы

$$\lim_{l\to\infty} P^{dl+r} = P^r (P^d)^\infty, \ r = 1, \dots, d-1.$$

Матрица $P^r(P^d)^\infty$ имеет ту же блочную структуру, что и P^r, причём с учётом теоремы 2 §3

$$(P^r(P^d)^\infty)_{ij} = \lim_{l\to\infty} p_{ij}^{(dl+r)} = \pi_j, \text{ если } t_{ij} = r,$$

$$(P^r(P^d)^\infty)_{ij} = p_{ij}^{(dl+r)} = 0, \ l = 1, 2, \dots, \text{ если } t_{ij} \neq r.$$

Таким образом, для любых состояний i, j последовательность $p_{ij}^{(k)}$ стремится к периодическому повторению с периодом d, при котором серия $d-1$ нулей сменяется числом π_j.

Задача 2. Докажите, что марковская система с тремя состояниями из примера 1 — эргодическая и найдите предельные вероятности состояний.

Глава 3

Задача о максимальном потоке в сети

§ 1. Основные леммы о потоках и разрезах в сети

Сеть — это орграф, в котором выделены две вершины — *источник* s и *сток* t. По определению из источника дуги могут лишь выходить, а в сток — лишь входить. Каждой дуге (i, j) сопоставлено положительное целое число c_{ij} — *пропускная способность* дуги.

Потоком в сети называется функция f, заданная на дугах сети, принимающая целые значения и удовлетворяющая условиям:

1) $0 \leqslant f_{ij} \leqslant c_{ij}$,

2) $\sum\limits_{i:i \to j} f_{ij} = \sum\limits_{l:j \to l} f_{jl}$, для любой промежуточной вершины j (то есть, неравной s, t).

Число f_{ij} называется *величиной потока по дуге* (i, j).

Пример 1. Рассмотрим сеть, изображённую на рис. 1.

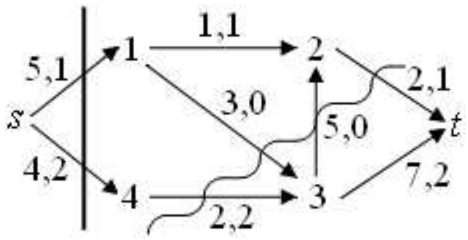


Рис. 1

Первое из чисел, приписанных дуге, — пропускная способность, второе — величина потока по этой дуге.

Разрезом сети называется любое разбиение $(X, \bar{X})$ множества вершин на две части, такое, что $s \in X$, $t \in \bar{X}$. *Пропускной способностью разреза* называется сумма пропускных способностей дуг, ведущих из X в $\bar{X}$, то есть, число

$$c(X, \bar{X}) = \sum_{\substack{i \to j \\ i \in X, j \in \bar{X}}} c_{ij}.$$

Величиной потока через разрез называется сумма величин потока на дугах, ведущих из X в $\bar{X}$, минус сумма значений потока на дугах, ведущих в обратном направлении, то есть, число

$$f(X,\bar{X}) = \sum_{\substack{i\to j\\ i\in X,\, j\in\bar{X}}} f_{ij} \; - \sum_{\substack{i\to j\\ i\in\bar{X},j\in X}} f_{ij}.$$

На рис. 1 указаны два разреза сети — вертикальной чертой и волнистой линией. Для первого разреза пропускная способность равна 9, величина потока равна 3. Для второго, соответственно, 7 и 3.

Лемма 1. *Величина потока через любой разрез одна и та же.*

Доказательство. Пусть $(X,\bar{X})$ — произвольный разрез, j — любая вершина из X, но не источник. Докажем, что через разрез $(X\setminus\{j\},\bar{X}\cup\{j\})$ протекает поток той же величины, что и через $(X,\bar{X})$. После перемещения вершины j инцидентные ей дуги иначе влияют на величину потока через разрез. А именно,

$$f\Big(X\setminus\{j\},\bar{X}\cup\{j\}\Big) = f(X,\bar{X}) + \sum_{\substack{i\to j\\ i\in X}} f_{ij} + \sum_{\substack{i\to j\\ i\in\bar{X}}} f_{ij} - \sum_{\substack{j\to l\\ l\in X}} f_{jl} - \sum_{\substack{j\to l\\ l\in\bar{X}}} f_{jl}.$$

Однако, в силу свойства 2) потока сумма последних четырех слагаемых в правой части равенства равна нулю. Удалив из X еще одну вершину k и рассуждая вполне аналогично, получим

$$f\Big(X\setminus\{j,k\},\bar{X}\cup\{j,k\}\Big) = f(X,\bar{X}).$$

Продолжая перемещения вершин, на некотором шаге получим равенство

$$f(X,\bar{X}) = f(\{s\},\bar{\{s\}}) = \sum_{l:\ s\to l} f_{sl}.$$

Итак, величина потока через любой разрез равна суммарной величине потока на дугах, выходящих из источника. Этим завершено доказательство леммы. □

В частности,

$$\sum_{l:\ s\to l} f_{sl} = \sum_{m:\ m\to t} f_{mt},$$

то есть, сумма потоков на дугах, выходящих из источника, равна сумме потоков на дугах, входящих в сток. Лемма 1 делает законным следующее определение.

Величиной $val(f)$ потока f называется величина потока через (любой) разрез сети.

Лемма 2. *Для любого потока f и любого разреза $(X, \bar{X})$*

$$val(f) \leqslant c(X, \bar{X}).$$

Доказательство следует из определений соответствующих величин.

Лемма 3. *Если для некоторого потока f_0 и некоторого разреза $(X_0, \bar{X}_0)$*

$$val(f_0) = c(X_0, \bar{X}_0), \tag{1}$$

то величина потока f_0 — максимально возможная, а пропускная способность разреза $(X_0, \bar{X}_0)$ — наименьшая из пропускных способностей разрезов сети.

Доказательство. Действительно, для произвольного потока f и произвольного разреза $(X, \bar{X})$ из леммы 2 и равенства (1) следует:

$$val(f) \leqslant c(X_0, \bar{X}_0) = val(f_0) \leqslant c(X, \bar{X}). \ \square$$

Для краткости поток максимальной величины называют *максимальным* потоком, а разрез с минимальной пропускной способностью — *минимальным* разрезом.

§ 2. Нахождение максимального потока: алгоритм и теорема

В этом параграфе излагается алгоритм построения максимального потока и теорема о максимальном потоке и минимальном разрезе (Форд и Фалкерсон, 1955).

Пусть некоторый поток f в сети уже имеется, например, поток с нулевыми значениями на всех дугах. Алгоритм Форда—Фалкерсона состоит из двух чередующихся процедур — помечивания вершин и изменения потока.

Помечивание вершин. Вершины снабжаются метками, состоящими из двух элементов. Источник s получает условную метку $(-, \infty)$. Теперь пусть имеется некоторое множество помеченных вершин. Выбираем любую из них и обрабатываем ее. Обработка i-ой вершины с меткой (x, ε) состоит в помечивании *из вершины* i смежных непомеченных вершин по следующему правилу:

если $i \to j$ и $f_{ij} < c_{ij}$, то вершине j присваивается метка $(i^+, \min(\varepsilon,\ c_{ij} - f_{ij}))$;

если $i \leftarrow j$ и $f_{ji} > 0$, то вершине j присваивается метка $(i^-,\ \min(\varepsilon, f_{ji}))$.

Затем обрабатывается другая помеченная вершина и так далее.

Процесс помечивания заканчивается в двух случаях:

1) Ни одну вершину больше нельзя пометить, но сток не помечен. Тогда алгоритм останавливается. 2) Сток помечен. Тогда производится изменение потока.

Изменение потока. Пусть сток получил метку (m^+, δ). Тогда прибавляем δ к f_{mt} и переходим в вершину m. Общий шаг: если мы находимся в вершине j с меткой (i^+, x), то прибавляем δ к f_{ij} и переходим в i. А если метка j равна (i^-, x), то вычитаем δ из f_{ij} и переходим в i. Заметим, что правило формирования меток таково, что после прибавления δ новое значение потока не превышает пропускной способности дуги, а при вычитании δ не получается отрицательной величины. Продолжаем изменение потока, пока не достигнем источника. Почему это непременно случится? Представим себе, что мы нумеруем вершины по мере их помечивания. Тогда вершины, помеченные из v, получают номер, бóльший, чем номер вершины v. При изменении потока переход совершается наоборот, от вершины с бóльшим номером к вершине с меньшим номером. Пока не достигнем вершины номер один — источника.

Следует еще убедиться в том, что при изменении потока не нарушается условие 2) в определении потока. При прохождении вершины j возможны четыре случая:

$$\begin{aligned}
&i \xrightarrow{f_{ij}+\delta} j \xrightarrow{f_{jk}+\delta} k\\
&i \xrightarrow{f_{ij}+\delta} j \xleftarrow{f_{kj}-\delta} k\\
&i \xleftarrow{f_{ji}-\delta} j \xleftarrow{f_{kj}-\delta} k\\
&i \xleftarrow{f_{ji}-\delta} j \xrightarrow{f_{jk}+\delta} k
\end{aligned}$$

Легко проверить, что условие 2) во всех случаях по-прежнему выполняется.

Величина измененного потока на $\delta \geqslant 1$ больше, чем у исходного потока. Теперь снова переходим к помечиванию. Схема алгоритма такова:

$$\begin{tikzcd}
\text{ПОМ} \arrow[r, shift left, "\text{если сток пометился}"] \arrow[d, "\text{если сток не пометился}"'] & \text{ИЗМ} \arrow[l, shift left] \\
\text{КОНЕЦ} &
\end{tikzcd}$$

Поскольку поток увеличивается не меньше, чем на единицу, а величина потока не может превышать пропускной способности любого разреза, то алгоритм останавливается после конечного числа шагов.

Докажем теорему о результатах работы алгоритма.

Теорема 1. *Поток f, вычисленный алгоритмом, имеет максимальную величину, а разрез $(X, \bar{X})$, где X — множество вершин, помеченных при последнем помечивании, имеет минимальную пропускную способность.*

Д о к а з а т е л ь с т в о. Для всякой дуги (i, j), где $i \in X$, $j \in \bar{X}$, верно $f_{ij} = c_{ij}$. Действительно, если допустить, что $f_{ij} < c_{ij}$, то j должна была быть помечена из i по первому правилу помечивания. А для всякой дуги (j, i), где $i \in X$, $j \in \bar{X}$, верно $f_{ji} = 0$, поскольку при $f_{ji} > 0$ вершина j должна была быть помечена из i по второму правилу помечивания. В силу этих свойств

$$f(X, \bar{X}) = c(X, \bar{X}),$$

и утверждение теоремы следует из леммы 3 §1. $\square$

Если опустить детали, связанные с конкретным алгоритмом (существуют различные алгоритмы вычисления максимального потока), теорему 1 можно сформулировать так:

Теорема 2 (о максимальном потоке и минимальном разрезе). *В любой сети существует максимальный поток. Величина максимального потока равна пропускной способности минимального разреза.*

Пример 2. Требуется построить максимальный поток в сети

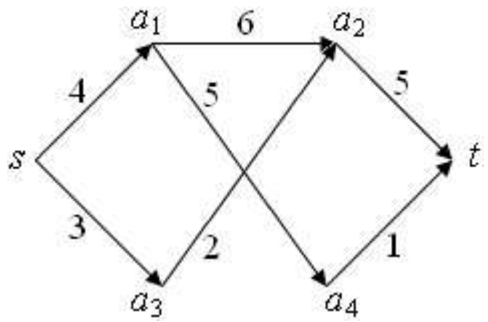


Рис. 2

Решение. (Ниже по технической причине вместо c_{ij}, f_{ij} пишем $c(i, j), f(i, j)$.) Начальный поток возьмем нулевой.

Помечивание вершин. Источник s получает метку $(-, \infty)$. Для смежной с s вершины a_1 имеем $c(s, a_1) - f(s, a_1) = 4 - 0 = 4 > 0$, поэтому a_1 получает метку $(s^+, 4)$. Аналогично, для вершины a_3 имеем $c(s, a_2) - f(s, a_2) = 3 - 0 = 3 > 0$, поэтому a_3 получает метку $(s^+, 3)$. Теперь вершина s обработана. Переходим к помеченной вершине a_1. Поскольку $c(a_1, a_2) - f(a_1, a_2) = 6 - 0 = 6 > 0$, можно пометить вершину a_2: она получает метку $(a_1^+, 4)$, где $4 = \min(6, 4)$. Аналогично, приписываем вершине a_4 метку $(a_1^+, 4)$. Этим завершена обработка вершины a_1. Переходя к обработке помеченной вершины a_2, замечаем, что из неё можно пометить сток t, так как $c(a_2, t) - f(a_2, t) = 5 - 0 = 5 > 0$, его метка — $(a_2^+, 4)$, где $4 = \min(5, 4)$. Сеть с указанными метками изображена на рис. 3. Поскольку сток получил метку, переходим к изменению потока.

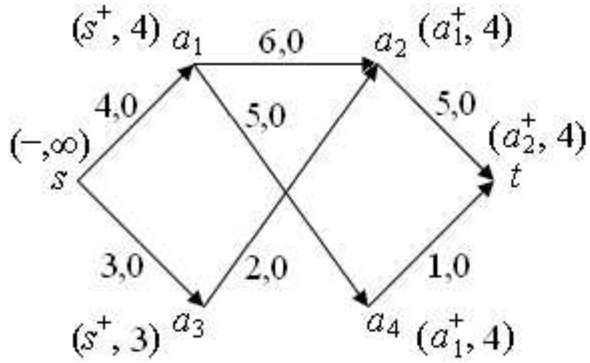


Рис. 3

Изменение потока. Сток t получил метку $(a_2^+, 4)$, поэтому меняем $f(a_2, t) = 0$ на $f(a_2, t) = 0 + 4 = 4$ и переходим к вершине a_2. Её метка — $(a_1^+, 4)$, значит, $f(a_1, a_2) = 0$ следует заменить на $f(a_1, a_2) = 0 + 4 = 4$. Наконец, для вершины a_1 с меткой $(s^+, 4)$ меняем $f(s, a_1) = 0$ на $f(s, a_1) = 0 + 4 = 4$. Стираем все метки. Теперь величина потока равна 4.

Помечивание вершин. Снова приписываем s метку $(-, \infty)$. У вершины s есть две смежных вершины — a_1 и a_3. Вершину a_1 из s пометить нельзя, так как $c(s, a_1) - f(s, a_1) = 4 - 4 = 0$, а a_3 — можно, поскольку $c(s, a_3) - f(s, a_3) = 3 - 0 = 3 > 0$. Следовательно, a_3 получает метку $(s^+, 3)$. Вершина s обработана, переходим к помеченной вершине a_3. С a_3 смежна одна непомеченная вершина — a_2. Из a_3 в a_2 ведёт дуга, причём $c(a_3, a_2) - f(a_3, a_2) = 2-0=2>0$, поэтому a_2 получает метку $(a_3^+, 2)$, где $2 = \min(2, 3)$. Обработка вершины a_3 завершена, переходим к помеченной вершине a_2. Поскольку $c(a_2, t) - f(a_2, t) = 5-4=1>0$ стоку t приписываем метку $(a_2^+, 1)$, где $1 = \min(1, 2)$. Сеть с новыми метками изображена на рис. 4.

Изменение потока. Двигаясь по меткам вершин от t к s, устанавливаем $f(a_2, t) = 4 + 1 = 5$, $f(a_3, a_2) = 0 + 1 = 1$, $f(s, a_3) = 0 + 1 = 1$. Стираем метки. Величина нового потока равна 5.

Помечивание вершин. Приписываем s метку $(-, \infty)$, вершине a_3 — метку $(s^+, 2)$, вершине a_2 — метку $(a_3^+, 1)$. Сток из a_2 пометить нельзя, поскольку $c(a_2, t) - f(a_2, t) = 5 - 5 = 0$, но можно пометить верши-

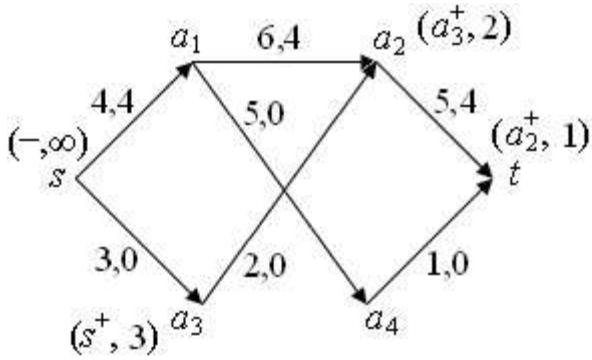


Рис. 4

ну a_1. В самом деле, из a_1 в a_2 ведет дуга, причем $f(a_1, a_2) = 4 > 0$, следовательно, вершина a_1 получает метку $(a_2^-, 1)$, где $1 = \min(4, 1)$. Теперь вершине a_4 приписываем метку $(a_1^+, 1)$, а стоку t — метку $(a_4^+, 1)$.

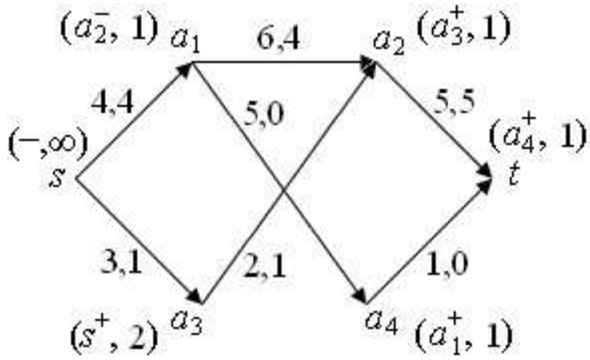


Рис. 5

Изменение потока. Положив $f(a_4, t) = 0 + 1 = 1$, $f(a_1, a_4) = 0 + 1 = 1$, $f(a_1, a_2) = 4 - 1 = 3$, $f(a_3, a_2) = 1 + 1 = 2$ и $f(s, a_3) = 1 + 1 = 2$, получаем поток величины 6 (см. рис. 6).

Помечивание вершин. Приписав источнику s метку $(-, \infty)$, а вершине a_3 — метку $(s^+, 1)$, убеждаемся в том, что дальнейшее помечивание невозможно. Таким образом, сток остался непомеченным. Работа алгоритма закончена, поток максимальной величины 6 построен. Заметим, что пропускная способность разреза $(X, \bar{X})$, где $X = \{s, a_3\}$, тоже равна 6 и является минимальной согласно теореме 1.

Интуиция, возможно, подсказывает читателю, что должна быть связь между потоками в сети и путями из источника в сток. Действительно, справедливо следующее

Предложение 1. *Если в сети существует (s, t)-путь, то в ней имеется поток положительной величины. Обратно, если в сети*

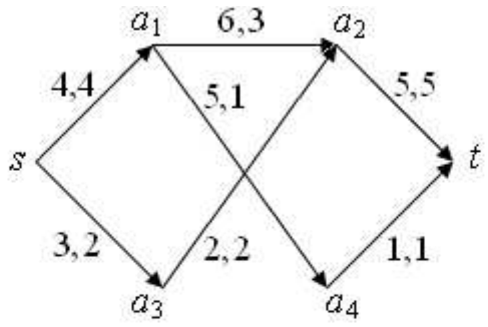


Рис. 6

задан поток положительной величины, то существует (s,t)-путь, на дугах которого поток положителен.

Д о к а з а т е л ь с т в о. Первое утверждение оставим в виде несложного упражнения. Чтобы доказать второе, обозначим через X множество вершин, достижимых из s путями, на дугах которых поток положителен. Предположим, что t не содержится в X, и рассмотрим разрез $(X,\bar{X})$. Из определения X следует, что не существует дуг с началом в X и концом в $\bar{X}$, несущих положительный поток. Но тогда величина потока через разрез $(X,\bar{X})$ оказывается равной нулю, что противоречит данному условию. □

Упражнение 1. Примените описанный выше алгоритм для нахождения максимального потока и минимального разреза в сети из примера 1.

§ 3. Приложения теоремы о потоках

С помощью теоремы о максимальном потоке и минимальном разрезе и соответствующего алгоритма можно решать разнообразные задачи теории графов.

1. *Достижимость вершин.* Пусть i,j — различные вершины орграфа. Достижима ли вершина j из i, то есть, существует ли в графе (i,j)-путь?

Если существует (i,j)-путь, то существует и простой (i,j)-путь, поэтому ответ на наш вопрос не изменится, если удалить дуги, входящие в i, и дуги, выходящие из j. Пропускные способности всех дуг положим равными 1 и получим сеть с источником i и стоком j, или, короче, (i,j)-сеть. Отметим, что для (i,j)-сети пропускная способность разреза $(X,\bar{X})$ равна числу элементов в множестве $E_X = \{k \to l : k \in X, l \in \bar{X}\}$. Следующее утверждение прямо следует из предложения 1 §2.

Предложение 1. *Вершина j достижима из вершины i тогда и только тогда, когда максимальный поток в (i, j)-сети имеет положительную величину.*

2. *Непересекающиеся пути и теорема Менгера.* Пусть i, j — различные вершины орграфа. Два (i, j)-пути называются *непересекающимися по дугам*, если они не имеют общих дуг. Каково максимальное количество непересекающихся по дугам простых (i, j)-путей? Для решения этой задачи снова воспользуемся понятием (i, j)-сети из п. 1.

Предложение 2. *Максимальное количество непересекающихся по дугам простых (i, j)-путей равно величине максимального потока в (i, j)-сети.*

Доказательство. Рассмотрим некоторое множество W непересекающихся по дугам простых (i, j)-путей, содержащее максимальное число w элементов. Пусть через произвольно взятую вершину m проходят k путей из W. Поскольку каждый из них (в силу простоты) проходит через m единственный раз, и пути не имеют общих дуг, то в m входит ровно k дуг и выходит из m ровно k дуг, принадлежащих путям из W. Отсюда следует, что функция, равная единице на дугах путей из W и нулю на прочих дугах, является потоком. Величина его, очевидно, равна w.

Обратно, пусть в сети имеется поток величины w. В силу предложения 1 §2 существует простой (i, j)-путь, на дугах которого поток равен 1. Удалив дуги этого пути, получим сеть с потоком величины $w - 1$. Будем повторять процедуру удаления путей, пока не получим сеть с потоком величины 0. Легко заметить, что удаляемые пути не имеют общих дуг и всего таких путей w.

Мы установили, что в (i, j)-сети тогда и только тогда имеется w-элементное множество простых (i, j)-путей, непересекающихся по дугам, когда в этой сети есть поток величины w. Отсюда немедленно следует доказываемое утверждение. □

Из предложения 2 и теоремы о потоках вытекает

Следствие 1. *Максимальное количество непересекающихся по дугам простых (i, j)-путей равно наименьшей из пропускных способностей разрезов (i, j)-сети.*

Множество C дуг графа называется (i, j)-*разделяющим*, если любой (i, j)-путь содержит дугу из C. Нетрудно понять, что количество непересекающихся по дугам простых (i, j)-путей не может превышать числа дуг любого (i, j)-разделяющего множества. Теперь заметим, что для любого разреза $(X, \bar{X})$ (i, j)-сети множество дуг

$E_X = \{k \to l : k \in X, l \in \bar{X}\}$ является (i,j)-разделяющим, причём его пропускная способность равна $|E_X|$. Отсюда и из следствия 1 вытекает утверждение, известное как теорема Менгера:

Следствие 2. *Максимальное число непересекающихся по дугам простых (i,j)-путей равно минимальному числу элементов (i,j)-разделяющего множества дуг.*

3. *Граничный ранг $(0,1)$-матрицы и теорема Холла.* Сопоставим $(0,1)$-матрице $A = (a_{ij})$ размеров $m \times n$ двудольный граф с вершинами $s_1, \ldots, s_m$; $t_1, \ldots, t_n$, положив, что $s_i \to t_j$, если $a_{ij} = 1$. Добавим источник s и сток t, и положим, что $s \to s_i$, $t_j \to t$ для всех i, j. Пропускные способности дуг считаем равными 1. Полученную сеть назовём A-сетью.

В A-сети элементу $a_{ij} = 1$ соответствует простой (s,t)-путь $s \to s_i \to t_j \to t$, причём это соответствие между единицами матрицы A и простыми (s,t)-путями в A-сети взаимно однозначно. Как легко видеть, элементы $a_{ij} = 1$ и $a_{uv} = 1$ стоят на разных линиях (то есть, $i \neq u, j \neq v$) в точности тогда, когда соответствующие им пути не имеют общих дуг. Следовательно, граничный ранг матрицы A равен максимальному количеству непересекающихся по дугам простых (s,t)-путей. Применяя предложение 2, получаем

Предложение 3. *Граничный ранг матрицы A равен минимальной пропускной способности разреза A-сети (или, что то же, величине максимального потока в A-сети).*

Как следствие получим простое доказательство теоремы Холла в её нетривиальной части.

Следствие 3. *Если любые k юношей $(k = 1, \ldots, m)$ знакомы в совокупности не менее, чем с k девушками, то задача о свадьбах разрешима.*

Д о к а з а т е л ь с т в о. Пусть задача о свадьбах задаётся матрицей A. Как известно, граничный ранг матрицы A можно трактовать как максимально возможное количество бракосочетаний. Согласно предложению 3 для доказательства следствия достаточно проверить, что минимальная пропускная способность A-сети равна равна числу m юношей. Рассмотрим любой разрез $(X, \bar{X})$. Пусть в X входят k вершин, соответствующих юношам, и l вершин, соответствующих девушкам. Тогда в множество E_X входят: а) $m - k$ дуг, ведущих из источника к вершинам-юношам из $\bar{X}$, б) l дуг, ведущих от вершин-девушек из X к источнику, в) по условию следствия, не меньше, чем $k-l$ дуг, ведущих от вершин-юношей из X к вершинам-девушкам из $\bar{X}$. Таким образом, $|E_X| \geq (m-k)+l+(k-l)=m$. Равенство $|E_X|=m$ достигается при $X=\{s\}$. $\square$

Глава 4

Теория автоматов

§ 1. Буквы, слова, языки, автоматы.

Алфавитом называется любое конечное непустое множество X. Его элементы называются *буквами.* Конечные последовательности букв называются *словами.* Например,

а) $X = \{0, 1\}$, слова — двоичные последовательности 0, 1, 00, $1101, \ldots$;

б) $X = \{$а, б, в, …, я$\}$ — русский алфавит, слово – при нашем определении – любая последовательность букв;

в) $X = \{\cdot, -\}$ — алфавит азбуки Морзе.

В примерах мы обычно пользуемся алфавитом, составленным из первых букв латинского алфавита:

г) $X = \{a, b\}$.

Длина $|p|$ слова p равна количеству его букв, причем каждая буква считается столько раз, сколько она встречается в слове, так что, например, длина слова 00100 равна 5. *Пустое слово* обозначаем буквой e. Его длина равна 0. Множество всех слов в алфавите X обозначается через X^*.

Если $p = x_1 \ldots x_k$, $q = y_1 \ldots y_m$, то то *конкатенацией* слов p и q называется слово $pq = x_1 \ldots x_k y_1 \ldots y_m$. Для пустого слова e и любого слова p полагаем $ep = pe = p$. Докажите в виде упражнения

Предложение 1. *Конкатенация — ассоциативная операция с нейтральным элементом (единицей), равным e.*

Таким образом, множество X^* относительно операции конкатенации образует моноид, то есть, полугруппу с единицей.

Языком над алфавитом X называется любое множество слов $L \subseteq X^*$. Например,

$L_1 = \{e, ab, a^2b^2\}$;

$L_2 = \{a^k b^l : \ k, l = 1, 2, \ldots\}$

(используются обозначения $p^k = \underbrace{pp\ldots p}_{k \text{ раз}}$, $p^0 = e$);

$L_3 = \{a^k b^k : \ k = 1, 2, \ldots\}$;

$L_4 = \{a^{k^2} : \ k = 1, 2, \ldots\}$.

Интерес представляют не любые множества слов, а те, которые возникают при решении задач математики, лингвистики, программирования — задач, связанных с переработкой информации. Такими множествами слов являются, в частности, языки, распознаваемые конечными автоматами.

Автомат — это совокупность (X, S, δ), где X — алфавит, S — непустое множество, элементы которого называются *состояниями* автомата, δ — функция из $S \times X$ в S, она называется *функцией перехода*.

Запись $\delta(s, x) = s'$ читается так: автомат под действием сигнала (буквы) $x \in X$ переходит из состояния s в состояние s'.

Автомат называется *конечным*, если множество его состояний конечное. Конечный автомат можно задать таблицей. Например, таблица

	a	b
1	2	3
2	4	5
3	3	3
4	3	6
5	3	3
6	3	7
7	3	3

задает автомат с алфавитом $X = \{a, b\}$ и множеством состояний $S = \{1, 2, 3, 4, 5, 6, 7\}$, переходы которого определены очевидным образом: например, $\delta(2, b) = 5$, $\delta(5, a) = 3$ и так далее. Автомат удобно изображать в виде орграфа. При этом вершины обозначают состояния, и дуга ведёт из s в $s' \Leftrightarrow \delta(s, x) = s'$ для некоторой буквы x. Все такие буквы считаются метками этой дуги.

Автомат, заданный выше таблицей, изображается графом

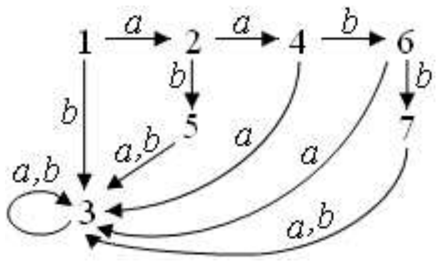
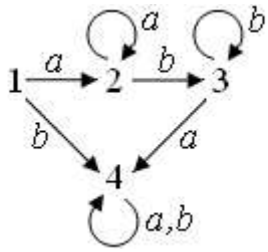


Рис. 1

Заметим, что любой граф с множеством вершин S, дуги которого помечены буквами алфавита X так, что для любой пары $(s, x) \in S \times X$ имеется ровно одна дуга с началом s, помеченная буквой x, определяет автомат. Действительно, будем считать, что конец этой дуги указывает на состояние, в которое автомат переходит из s под действием x. Этим функция перехода полностью определяется.

Вот еще один пример графа автомата и соответствующей ему таблицы:



	a	b
1	2	4
2	2	3
3	4	3
4	4	4

Рис. 2

Перейдем к более удобной форме записи: вместо $\delta(s,x) = s'$ будем писать $s\delta(x) = s'$. Тем самым каждой букве x сопоставляется преобразование $\delta(x) : S \to S$. Слову $p = x_1x_2\ldots x_k$ сопоставим произведение преобразований

$$\delta(p) = \delta(x_1)\delta(x_2)\ldots\delta(x_k).$$

Положим, что $\delta(e)$ — тождественное преобразование. Таким образом, конкатенации pq слов p и q отвечает произведение соответствующих преобразований, то есть, $\delta(pq) = \delta(p)\delta(q)$. Тем самым, множество $\{\delta(p), p \in X^*\}$ преобразований множества S образует мультипликативный моноид. Назовём его *моноидом автомата* (X, S, δ).

Фактически выше мы определили гомоморфизм δ моноида слов X^* в моноид преобразований, совершаемых этими словами в пространстве состояний автомата.

Если автомат находится в состоянии s и получает на входе слово $p = x_1x_2\ldots x_k$, то он переходит в состояние $s\delta(x_1)$, затем в $s\delta(x_1)\delta(x_2) = s\delta(x_1x_2)$ и так далее, наконец, в состояние $s\delta(x_1)\delta(x_2)\ldots\delta(x_k) = s\delta(p)$. На графе автомата состоянию s и слову p отвечает единственный путь

$$s, s\delta(x_1), \ldots, s\delta(p). \tag{1}$$

Назовем автомат (X, S, δ) *настроенным*, если в нем выделено некоторое *начальное* состояние s_0 и некоторое подмножество $F \subseteq S$ *допускающих* состояний. Мы будем рассматривать и “полунастроенные” автоматы вида (X, S, δ, F) с фиксированным множеством F допускающих состояний, но без выделенного начального состояния. Говорят, что настроенный автомат (X, S, δ, s_0, F) *распознаёт язык* $L \subseteq X^*$, если

$$s_0\delta(p) \in F \;\Leftrightarrow\; p \in L.$$

Таким образом, распознаваемый язык состоит из тех и только тех слов, которые переводят автомат из начального в какое-либо допускающее состояние.

На графе автомата словам распознаваемого языка отвечают в точности те пути (1), которые начинаются в вершине s_0, а заканчиваются в одной из вершин множества F.

Вернувшись к примерам автоматов, приведенным выше, нетрудно увидеть, что первый из них с настройкой $s_0 = 1$, $F = \{1, 5, 7\}$ распознаёт язык L_1, а второй автомат с настройкой $s_0 = 1$, $F = \{3\}$ распознаёт язык L_2. Однако существуют языки, нераспознаваемые никакими конечными автоматами.

Предложение 2. *Язык* $L_3 = \{a^k b^k : \ k = 1, 2, \dots\}$ *не распознаётся никаким конечным автоматом.*

Доказательство. Предположим противное: язык L_3 распознаётся некоторым автоматом (X, S, δ, s_0, F) с n состояниями. Рассмотрим состояния

$$s_0\delta(a), \ s_0\delta(a^2), \ \dots, \ s_0\delta(a^{n+1}).$$

В этой последовательности $n+1$ элементов, значит, для некоторых неравных k, l имеем $s_0\delta(a^k) = s_0\delta(a^l)$. Но тогда, с одной стороны,

$$(s_0\delta(a^k))\delta(b^k) = s_0\delta(a^k b^k) \in F, \tag{2}$$

а с другой,

$$(s_0\delta(a^l))\delta(b^k) = s_0\delta(a^l b^k) \notin F. \tag{3}$$

Но состояния в (2) и (3) равны. Полученное противоречие завершает доказательство. □

Возникает естественный вопрос: как устроены языки, распознаваемые конечными автоматами?

§ 2. Критерий распознаваемости языка конечным автоматом

Будем говорить, что слова $p, q \in X^*$ *различимы словом* $r \in X^*$ *относительно языка* $L \subseteq X^*$, если $pr \in L$, $qr \notin L$ или $pr \notin L$, $qr \in L$. Если для p и q различающих слов не существует, то будем говорить, что слова p и q *неразличимы относительно языка* L и писать $p \sim q(L)$ или $p \sim q$, когда язык L фиксирован. Таким образом,

$$p \sim q(L) \quad \Leftrightarrow \quad \forall r \in X^* \ (pr \in L \Leftrightarrow qr \in L).$$

Лемма 1. *Отношение неразличимости слов относительно языка рефлексивно, симметрично и транзитивно, то есть является отношением эквивалентности.*

Д о к а з а т е л ь с т в о. Первые два свойства очевидны, поясним транзитивность. Пусть $p \sim q$, $q \sim w$ и r — произвольное слово. Транзитивность следует из соотношений:

$$pr \in L \ \Leftrightarrow\ qr \in L \ \Leftrightarrow\ wr \in L. \ \square$$

Лемма 2. *Отношение $\sim$ устойчиво справа относительно конкатенации, а именно,*

$$p \sim q \ \Rightarrow\ \forall w \in X^* \ pw \sim qw.$$

Д о к а з а т е л ь с т в о. Действительно, для любого слова r имеем

$$(pw)r = p(wr) \in L \ \Leftrightarrow\ (qw)r = q(wr) \in L. \ \square$$

Задача 1. Для любого $p \in X^*$ определим язык $L_p = \{q : pq \in L\}$. Доказать, что $p \sim q(L) \Leftrightarrow L_p = L_q$.

Рангом языка L называется число $\operatorname{rk} L$ классов эквивалентности отношения $\sim (L)$. Другими словами, ранг — это максимальное количество слов, попарно различимых относительно языка. Если число классов эквивалентности бесконечно, то полагают $\operatorname{rk} L = \infty$.

Обозначим через $[w]$ класс неразличимости, содержащий слово w. Поскольку каждое слово над X лежит точно в одном классе, то запись $[w]$ однозначно фиксирует класс.

Лемма 3. *Если язык $L \subseteq X^*$ распознается конечным настроенным автоматом (X, S, δ, s_0, F) с n состояниями, то $\operatorname{rk} L \leqslant n$.*

Д о к а з а т е л ь с т в о. Пусть $p_1, \ldots, p_{n+1}$ — произвольно выбранные слова. Среди состояний $s_0\delta(p_1), \ldots, s_0\delta(p_{n+1})$ есть равные, пусть $s_0\delta(p_i) = s_0\delta(p_j)$. Тогда $(s_0\delta(p_i))\delta(r) = (s_0\delta(p_j))\delta(r)$, то есть, $s_0\delta(p_i r) = s_0\delta(p_j r)$ для любого слова r. Это значит, что слова p_i, p_j неразличимы относительно L. Итак, среди любых $n+1$ слов по меньшей мере два слова неразличимы. Следовательно, $\operatorname{rk} L \leqslant n$. $\square$

Критерий распознаваемости языка конечным автоматом был найден Майхиллом и Нероудом в середине прошлого века.

Теорема 1. *Язык $L \subseteq X^*$ распознается конечным автоматом тогда и только тогда, когда он имеет конечный ранг. Если $\operatorname{rk} L = n$, то существует автомат с n состояниями, распознающий L, причем никакой автомат, распознающий L, не может иметь состояний меньше, чем n.*

Д о к а з а т е л ь с т в о. Построим автомат $\widetilde{A}(L)$, распознающий язык L. Его состояниями считаем классы неразличимости. Определим функцию перехода. Для любого класса $[w]$ и любой буквы x положим $[w]\widetilde{\delta}(x) = [wx]$. Корректность этого определения обеспечивается леммой 2. Для слова $p = x_1 x_2 \ldots x_k$ имеем:

$$[w]\widetilde{\delta}(p) = [w]\widetilde{\delta}(x_1)\widetilde{\delta}(x_2)\ \ldots\ \widetilde{\delta}(x_k) =$$

$$[wx_1]\widetilde{\delta}(x_2)\ \ldots\ \widetilde{\delta}(x_k) = [wx_1x_2\ldots x_k] = [wp].$$

В качестве начального состояния возьмем класс $[e]$. Класс $[w]$ считаем допускающим состоянием, если $w \in L$, то есть, $[w] \in \widetilde{F} \Leftrightarrow w \in L$. Это определение корректно, так как слова из класса неразличимости либо все принадлежат L, либо все не принадлежат. Из предыдущей цепочки равенств при $w = e$ для произвольного слова p получаем $[e]\widetilde{\delta}(p) = [p]$ и, следовательно, $[e]\widetilde{\delta}(p) \in \widetilde{F} \Leftrightarrow p \in L$. Итак, автомат $\widetilde{A}(L)$ распознает язык L. Если число $\operatorname{rk} L$ его состояний конечно, то любой автомат, распознающий L, не может иметь меньше состояний по лемме 3. Если же $\operatorname{rk} L = \infty$, то в силу той же леммы язык L не распознается никаким конечным автоматом. $\square$

Задача 2. Докажите, что любой конечный язык распознаётся некоторым конечным автоматом.

Пример 1. Пусть язык L состоит из слов, начинающихся с буквы a и заканчивающихся буквой b, то есть

$$L = \{apb:\ p \in X^*\}.$$

Рассмотрим слова e, a, b, ab. Легко заметить, что слово e отличимо от слов a и ab словом b, поскольку $eb = b \notin L$, $ab \in L$, $abb = ab^2 \in L$. Слова e и b различимы словом ab, так как $eab = ab \in L$, $bab \notin L$. Рассуждая аналогично, можно установить, что любые два слова из указанной четверки различимы. Теперь заметим, что любое слово, длина которого больше единицы, имеет один из трех типов: apa, apb, bp. Слова первого типа эквивалентны a, второго — ab, третьего — b. Следовательно, классов отношения $\sim (L)$ всего четыре: $[e]$, $[a]$, $[b]$, $[ab]$. Соответствующий автомат изображён на рис. 3. Начальное состояние — $[e]$, допускающее состояние одно — $[ab]$.

Опишем некоторый систематический способ построения автомата $\widetilde{A}(L)$. Пусть дан язык $L \subseteq X^*$. Множество слов $W \subseteq X^*$ называется *базисом отношения* $\sim (L)$, если

а) любые два слова из W различимы относительно L,

б) любое слово из X^* эквивалентно некоторому слову из W.

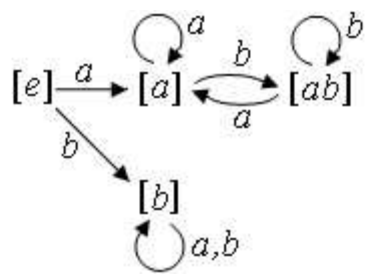


Рис. 3

Теорема 2. *Пусть множество W попарно различимых относительно L слов обладает свойствами:*

1) $e \in W$,

2) $\forall p \in W\ \forall x \in X\ \exists q \in W \quad px \sim q(L)$.

Тогда W — базис отношения $\sim (L)$.

Доказательство. Достаточно показать, что любое слово эквивалентно некоторому слову из W. Применим индукцию по длине слов. Для слов длины 0, то есть, для e, утверждение верно в силу 1). Пусть оно верно для слов длины k. Рассмотрим произвольное слово p длины $k+1$. Представим его в виде $p = p'x$, $x \in X$. По предположению индукции $p' \sim q$ для некоторого $q \in W$. Из леммы 2 выводим $p = p'x \sim qx$. Для слова qx ввиду 2) найдется слово $r \in W$ такое, что $qx \sim r$. Поскольку отношение $\sim$ транзитивно, то

$$p \sim qx,\ qx \sim r \ \Rightarrow\ p \sim r \in W,$$

что и завершает доказательство. $\square$

Если базис вычислен и каждому базисному слову p и каждой букве $x \in X$ сопоставлено базисное слово q, где $q \sim px$, то тем самым определена функция перехода: $[p]\widetilde{\delta}(x) = [px] = [q]$.

Метод нахождения базиса и одновременного построения таблицы функции перехода автомата $\widetilde{A}(L)$ заключается в следующем.

Строки таблицы соответствуют базисным словам, столбцы — буквам алфавита. Вначале имеется одно базисное слово e и таблица имеет одну строку (незаполненную). Опишем общий шаг заполнения таблицы. Находим первую, считая сверху, незаполненную строку и в ней — первую, считая слева, пустую клетку. Пусть строка соответствует слову p, а клетка стоит в столбце, соответствующем букве x. Проверяем, есть ли среди слов, уже зачисленных в базис, слово, эквивалентное px. Если есть – вписываем его в указанную пустую клетку, а если нет – вписываем в неё px. Объявляем слово px базисным и добавляем соответствующую ему строку к таблице.

Если на некотором шаге все строки таблицы оказались заполненными, то построение закончено. В том случае, когда $\operatorname{rk} L = \infty$, оно будет продолжаться неограниченно долго.

Пример 2. Рассмотрим язык L, состоящий из слов вида $pabbaq$, то есть, слов, содержащих подслово $abba$. Рекомендуем читателю проверить, что описанный метод приводит к нижеследующей таблице. Рядом изображён соответствующий автоматный граф. Короткая стрелка указывает на начальное состояние, единственное допускающее состояние обведено кружком.

	a	b
e	a	e
a	a	ab
ab	a	abb
abb	$abba$	e
$abba$	$abba$	$abba$

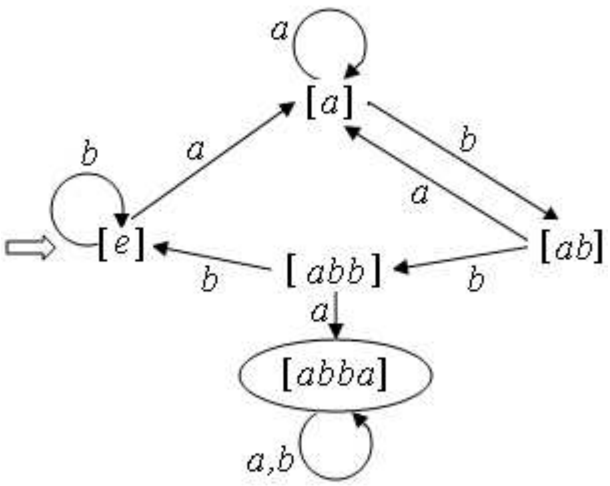


Рис. 4

§ 3. Единственность минимального автомата

Мы рассматриваем автоматы над некоторым фиксированным алфавитом X. Говорят, что состояния s и s' автомата (S, δ, F) *эквивалентны*, если для любого слова p

$$s\delta(p) \in F \Leftrightarrow s'\delta(p) \in F.$$

Настроенный автомат (S, δ, s_0, F) называется

1) *связным*, если для любого состояния $s \in S$ найдётся такое слово p, что $s_0\delta(p) = s$;

2) *приведённым*, если у него нет эквивалентных состояний.

Связный и приведённый настроенный автомат называется *минимальным*. Предлагаем доказать самостоятельно

Предложение 1. *Автомат $\widetilde{A}(L)$ — минимальный.*

Автоматы (S, δ) и (S', δ') называются *изоморфными*, если существует такая биекция $\sigma : S \to S'$, что

$$s\delta(x) = t \Leftrightarrow (s\sigma)\delta'(x) = t\sigma$$

или, в более краткой форме,

$$(s\sigma)\delta'(x) = s\delta(x)\sigma \tag{1}$$

для всех $s \in S$, $x \in X$.

Упражнение 1. Доказать, что автоматы изоморфны в том и только том случае, когда изоморфны их графы, причём соответствующие при этом изоморфизме дуги имеют одинаковые метки.

Настроенные автоматы

$$(S, \delta, s_o, F) \text{ и } (S', \delta', s_0', F') \tag{2}$$

назовём изоморфными, если дополнительно к условию (1) верно

$$s_0\sigma = s_0', \ s \in F \Leftrightarrow s\sigma \in F'. \tag{3}$$

Очевидно, что изоморфные автоматы распознают один и тот же язык. Замечательно, что для минимальных автоматов верно и обратное:

Теорема 1. *Если настроенные минимальные автоматы распознают один и тот же язык, то они изоморфны.*

Доказательство. Предположим, что автоматы (2) удовлетворяют условию теоремы. Рассмотрим множество пар состояний

$$\{(s_0\delta(p), \ s_0'\delta'(p)) : \ p \in X^*\}. \tag{4}$$

Легко усмотреть, что эти пары состоят из эквивалентных состояний. Ввиду связности первого автомата каждое из состояний $s \in S$ появляется в качестве левого элемента пары, а в силу приведённости второго автомата появляется ровно один раз. Действительно, предположим, что некоторое s входит в две пары. Тогда $s = s_0\delta(p) = s_0\delta(q)$ и $s_0'\delta(p) \neq s_0'\delta(q)$ для некоторых слов p и q. Получается, что неравные состояния второго автомата эквивалентны, поскольку они эквивалентны одному и тому же состоянию первого автомата — противоречие с приведённостью. По аналогичным причинам каждое из состояний $s' \in S'$ появляется в множестве (4) в качестве правого элемента пары ровно один раз. А это значит, что множество (4), сопоставляющее состоянию $s_0\delta(p)$ первого автомата эквивалентное ему состояние $s_0'\delta'(p)$ второго автомата, задаёт биекцию из S в S'. Обозначим её через σ и проверим выполнение условия (1). Пусть $s = s_0\delta(p)$, тогда

$$s\sigma\delta'(x) = s_0'\delta'(p)\delta'(x) = s_0'\delta'(px) = s_0\delta(px)\sigma = s_0\delta(p)\delta(x)\sigma = s\delta(x)\sigma.$$

Условия (3) выполняются очевидным образом. □

Ввиду предложения 1 доказанную теорему можно сформулировать так: *автомат* $\widetilde{A}(L)$ *— единственный, с точностью до изоморфизма, минимальный автомат, распознающий язык* L.

Из теоремы Майхилла—Нероуда, предложения 1 и теоремы 1 сразу следует, что число состояний любого конечного минимального настроенного автомата равно рангу распознаваемого им языка. Но верно и обратное утверждение.

Теорема 2. *Если число состояний конечного настроенного автомата* (S, δ, s_0, F) *совпадает с рангом распознаваемого им языка, то этот автомат — минимальный.*

Д о к а з а т е л ь с т в о. Рассмотрим состояния

$$s_0\delta(p_1),\ \ldots, s_0\delta(p_n), \tag{5}$$

где $p_1, \ldots, p_n$ — базис отношения $\sim (L)$. Допустим, что некоторые состояния $s_0\delta(p_i)$, $s_0\delta(p_j)$ эквивалентны. Но тогда, как легко проверить, $p_i \sim p_j(L)$, что невозможно для базисных слов. Следовательно, состояния последовательности (5) попарно неэквивалентны и, стало быть, попарно различны. Поскольку по условию теоремы автомат имеет $n = \operatorname{rk} L$ состояний, то все они присутствуют в (5). Мы доказали оба условия минимальности. □

§ 4. Признаки нераспознаваемости языка конечным автоматом

Лемма 1. *Пусть язык* L *распознается конечным автоматом. Тогда существует такое натуральное число* n*, что любое слово* p*,* $|p| \geqslant n$*, можно представить в виде* $p = p_1p_2p_3$*,* $p_2 \neq e$*, причем для всех* i

$$\begin{aligned} p \in L \quad &\Rightarrow \quad p_1p_2^ip_3 \in L,\\ p \notin L \quad &\Rightarrow \quad p_1p_2^ip_3 \notin L. \end{aligned}$$

Д о к а з а т е л ь с т в о. Пусть L распознается автоматом (S, δ, s_0, F) с n состояниями, $p = x_1x_2 \ldots x_n$ — произвольное слово. Среди $n+1$ состояний s_0, $s_0\delta(x_1)$, $s_0\delta(x_1x_2)$, $\ldots$, $s_0\delta(x_1 \ldots x_n)$ есть два равных. Пусть $s_0\delta(x_1 \ldots x_m) = s_0\delta(x_1 \ldots x_m x_{m+1} \ldots x_{m+l})$. Обозначим $x_1 \ldots x_m = p_1$, $x_{m+1} \ldots x_{m+l} = p_2$, $x_{m+l+1} \ldots x_n = p_3$. Не исключено, впрочем, что p_1 или p_3 могут быть пустыми. Ясно, что $s_0\delta(p_1) = s_0\delta(p_1p_2)$ влечет $s_0\delta(p_1) = s_0\delta(p_1p_2^i)$, а отсюда следует $s_0\delta(p) = s_0\delta(p_1p_2p_3) = s_0\delta(p_1p_2^ip_3)$ $\forall i \in \mathbb{N}$. Таким образом, слова

$p_1p_2^ip_3$ переводят автомат из начального в одно и то же состояние. Если оно допускающее, то все эти слова лежат в L, в противном случае все они не лежат в L. $\square$

Из леммы 1 немедленно следует признак нераспознаваемости:

Следствие 1. *Если для сколь угодно большого числа n найдется слово $p \in L$, $|p| \geqslant n$, такое, что при любом представлении $p = p_1p_2p_3$ найдется показатель i, при котором $p_1p_2^ip_3 \notin L$, то язык L не распознается никаким конечным автоматом.*

Очевидный пример — язык L_3. При любом представлении слова $p = a^kb^k \in L$ в виде $p = p_1p_2p_3$ $(p_2 \neq e)$ имеем $p_1p_2^2p_3 \notin L$. Следовательно, L_3 — нераспознаваемый язык.

Докажем нераспознаваемость языка L_4. Предположим противное. Тогда для достаточно большого k существует такое представление $a^{k^2} = a^{k_1}a^{k_2}a^{k_3}$ $(k_2 \geqslant 1)$, что слова $a^{k_1}a^{ik_2}a^{k_3}$, $a^{k_1}a^{(i+1)k_2}a^{k_3}$ принадлежат L_4 при $i = 1, 2, \dots$. Разность длин этих слов равна k_2 для любого i. Однако разность между длинами соседних слов из L_4 равна $(k+1)^2 - k^2 = 2k+1$ и она стремится к бесконечности при $k \to \infty$. Пришли к противоречию.

Теперь рассмотрим язык L_5, состоящий из слов вида a^k, где k — простое число. Он не распознается конечным автоматом, поскольку при сколь угодно большом простом k и любом разложении $a^k = a^{k_1}a^{k_2}a^{k_3}$ $(k_2 \geqslant 1)$ слово $a^{k_1}a^{ik_2}a^{k_3}$ не принадлежит L_5 при $i = k_1 + k_3 > 0$ (число $k_1 + (k_1 + k_3)k_2 + k_3$ — не простое). Если же $k_1 + k_3 = 0$, то число ik_2 — не простое при любом $i \geqslant 2$.

Еще одно сильное средство установления нераспознаваемости языков конечными автоматами состоит в следующем. Пусть L и M — непересекающиеся языки. Скажем, что слова p, q *различимы относительно языков* L и M, если для некоторого r

$$pr \in L, qr \in M, \text{ или } qr \in L, pr \in M. \tag{1}$$

Лемма 2. *Если существует бесконечное множество слов, попарно различимых относительно языков L и M, то любой язык P, такой, что*

$$L \subseteq P, \ M \cap P = \emptyset, \tag{2}$$

не распознается конечным автоматом.

Доказательство почти очевидно, поскольку если для слов p, q выполняется (1), то $pr \in P$, $qr \notin P$ или $pr \notin P$, $qr \in P$. То есть, слова, различимые относительно пары L, M, различимы относительно любого языка P, удовлетворяющего условию (2). $\square$

Рассмотрим языки $L_6 = \{a^k b a^k : \ k = 1, 2, \ldots\}$, $L_7 = \{a^k b a^l : k < l\}$. Легко доказать, что слова a^1, a^2, $\ldots$ попарно различимы относительно этой пары языков. Из леммы 2 следует, например, нераспознаваемость языка, состоящего из симметричных (одинаково читаемых слева направо и справа налево) слов.

Нами доказана нераспознаваемость языка L_3, причем двумя способами. Заметим, что первое доказательство (см. предложение 2 §1) фактически устанавливает, что слова a^1, a^2, $\ldots$ попарно различимы относительно языков L_3 и $L_7 = \{a^k b^l : \ k \neq l\}$. Из леммы 2 следует, что язык L_8, состоящий из слов, в которые буквы a и b входят в одинаковом количестве, не распознается конечным автоматом. Действительно, этот язык содержит L_3 и не пересекается с L_7.

§ 5. Свойства операций над языками

Теорема 1. *Если язык $L \subseteq X^*$ распознается конечным автоматом, то и его дополнение $\bar{L}$ распознается конечным автоматом.*

Д о к а з а т е л ь с т в о. Если автомат (S, δ, s_0, F) распознает L, то, очевидно, автомат $(S, \delta, s_0, \bar{F})$ распознает $\bar{L}$. $\square$

Прямым произведением автоматов

$$A_1 = (S_1, \delta_1) \text{ и } A_2 = (S_2, \delta_2)$$

называется автомат, обозначаемый $A_1 \times A_2$, с множеством состояний $S_1 \times S_2$ и функцией перехода $(u, v)\delta(x) = (u\delta_1(x), v\delta_2(x))$. Ясно, что для любого слова $p \in X^*$

$$(u, v)\delta(p) = (u\delta_1(p), v\delta_2(p)). \tag{1}$$

Теорема 2. *Если языки L и M распознаются конечными автоматами, то их объединение $L \cup M$ и пересечение $L \cap M$ тоже распознаются конечными автоматами.*

Д о к а з а т е л ь с т в о. Пусть автомат A_1 с настройкой s_1, F_1 распознает язык L, автомат A_2 с настройкой s_2, F_2 распознает язык M. Покажем, что автомат $A_1 \times A_2$ распознает языки $L \cup M$ и $L \cap M$ при подходящих настройках.

Положим, что (s_1, s_2) — начальное состояние автомата $A_1 \times A_2$. Вначале определим множество допускающих состояний так:

$$F = \{(u, v) : \ u \in F_1 \text{ или } v \in F_2\}. \tag{2}$$

Тогда согласно (1)

$$(s_1, s_2)\delta(p) \in F \Leftrightarrow s_1\delta_1(p) \in F_1 \text{ или } s_2\delta_2(p) \in F_2 \Leftrightarrow$$
$$p \in L \text{ или } p \in M \Leftrightarrow p \in L \cup M.$$

Таким образом, построенный автомат с множеством допускающих состояний (2) распознает язык $L \cup M$.

Теперь в качестве множества допускающих состояний возьмем множество

$$F = \{(u, v) : \ u \in F_1, v \in F_2\} = F_1 \times F_2. \tag{3}$$

Аналогично предыдущему случаю легко получаем, что с множеством допускающих состояний (3) тот же автомат распознает язык $L \cap M$. □

Прямое произведение автоматов полезно для решения и других вопросов теории автоматов. Например, даны настроенные автоматы

$$A_1 = (S_1, \delta_1, s_1, F_1), \quad A_2 = (S_2, \delta_2, s_2, F_2). \tag{4}$$

Как узнать, распознают ли эти автоматы один и тот же язык, то есть, эквивалентны ли они? Неэквивалентность означает, что некоторое слово p переводит один из автоматов в допускающее состояние, а другой автомат — в недопускающее состояние. Но это равносильно тому, что прямое произведение $A_1 \times A_2$ переводится некоторым словом p из начального состояния (s_1, s_2) в состояние $(s_1\delta_1(p), s_2\delta_2(p))$, где

$$s_1\delta_1(p) \in F_1, s_2\delta_2(p) \in \bar{F}_2 \text{ или } s_1\delta_1(p) \in \bar{F}_1, s_2\delta_2(p) \in F_2. \tag{5}$$

Свойство (5) можно выразить короче:

$$(s_1\delta_1(p), s_2\delta_2(p)) \in (F_1 \times \bar{F}_2) \cup (\bar{F}_1 \times F_2).$$

Таким образом, автоматы (4) эквивалентны тогда и только тогда, когда автомат

$$A_1 \times A_2 \text{ с настройкой } (s_1, s_2), \ (F_1 \times \bar{F}_2) \cup (\bar{F}_1 \times F_2) \tag{6}$$

распознает пустой язык. Дальше нам потребуется простая

Лемма 1. *Если язык, распознаваемый автоматом (S, δ, s_0, F) с n состояниями, не пуст, то он содержит слово длины $\leqslant n - 1$.*

Д о к а з а т е л ь с т в о. Непустота языка равносильна тому, что $s_0 \in F$ или в графе автомата есть путь с началом в s_0 и концом в множестве F. Но если такой путь существует, то существует и простой путь длины не более $n - 1$ с тем же началом и концом. Соответствующее ему слово принадлежит распознаваемому языку. □

Теорема 3. *Пусть автоматы (4) с числом состояний n_1 и n_2 распознают языки L_1 и L_2. Если слова длины $\leqslant n_1n_2 - 1$ в L_1 и L_2 одни и те же, то $L_1 = L_2$.*

Д о к а з а т е л ь с т в о. Условие теоремы означает, что не существует слова длины $\leqslant n_1n_2 - 1$, которое переводит настроенный автомат (6) в допускающее состояние. Поскольку этот автомат имеет n_1n_2 состояний, то в силу леммы 1 таких слов нет вообще, то есть, язык, распознаваемый автоматом (6), пустой. Стало быть, $L_1 = L_2$. □

Поскольку множество слов длины не более $\leqslant n_1n_2 - 1$ конечно, то теорема 3 обеспечивает возможность проверить эквивалентность автоматов за конечное число действий.

Пример 1. Рассмотрим автоматы на рис. 5. Положим, что у

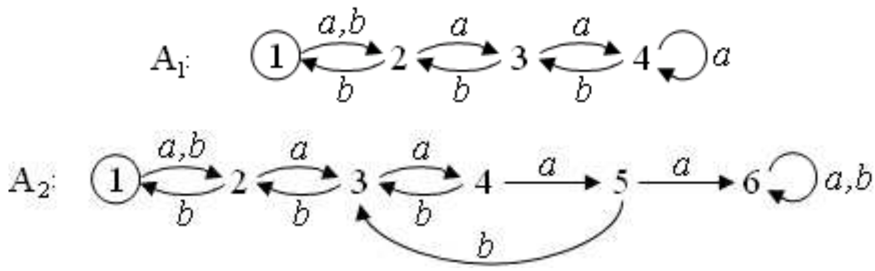


Рис. 5

обоих автоматов начальное состояние 1 является единственным допускающим состоянием. На рис. 6 изображён граф настроенного автомата (6) (проведены лишь дуги, ведущие из состояний, достижимых из начального состояния (1,1), допускающие состояния обведены кружка́ми).

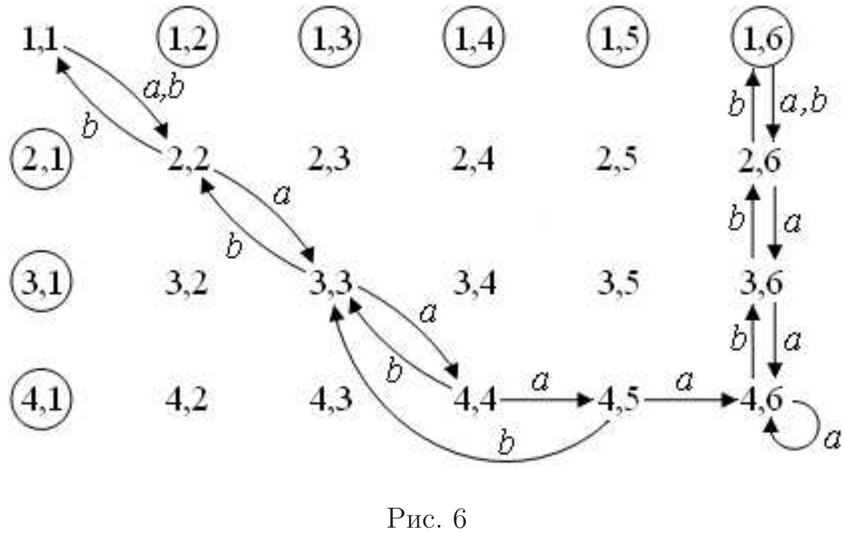


Рис. 6

Рекомендуем читателю проверить, что a^5b^3 — самое короткое слово, которое один из автоматов переводит в допускающее состояние, а другой — в недопускающее. Таким образом, языки L_1 и L_2, распознаваемые автоматами A_1 и A_2, различны, хотя слова длины не более 7 в них одни и те же.

§ 6. Синтаксический моноид

Пусть дан язык $L \subseteq X^*$. Определим на множестве X^* отношение конгруэнтности $\equiv (L)$ следующим образом. Слова $p, q \in X^*$ *конгруэнтны относительно* L, если

$$\forall r_1, r_2 \in X^* \ (r_1pr_2 \in L \Leftrightarrow r_1qr_2 \in L).$$

Прямо из определений отношений $\equiv (L)$ и $\sim (L)$ видно, что конгруэнтность — более сильное свойство, чем неразличимость:

$$p \equiv q(L) \ \Rightarrow \ p \sim q(L).$$

Лемма 1. *Отношение конгруэнтности относительно языка является эквивалентностью.*

Доказательство сходно с доказательством леммы 1 §2.

Для конгруэнтности справедливо более сильное утверждение, чем лемма 2 §2.

Лемма 2. *Если $p_1 \equiv q_1(L)$, $p_2 \equiv q_2(L)$, то $p_1p_2 \equiv q_1q_2(L)$.*

Д о к а з а т е л ь с т в о. Согласно определению конгруэнтности $p_1p_2 \equiv q_1p_2(L)$, $q_1p_2 \equiv q_1q_2(L)$. Используя транзитивность отношения $\equiv (L)$, получаем $p_1p_2 \equiv q_1q_2(L)$. $\square$

Определим *синтаксический моноид* языка $L \subseteq X^*$. Элементы синтаксического моноида — это классы конгруэнтности $\equiv (L)$. Определим умножение классов: если $\langle p\rangle$ и $\langle q\rangle$ — классы, содержащие слова p и q, то $\langle p\rangle\langle q\rangle = \langle pq\rangle$. В силу леммы 2 произведение классов не зависит от выбора их представителей и потому определено корректно. Легко проверяется ассоциативность умножения классов и то, что класс $\langle e\rangle$ — единица относительно умножения. Итак, классы конгруэнтности образуют моноид, который и называется синтаксическим моноидом языка L.

Лемма 3. *Пусть (S, δ, s_0, F) — минимальный автомат, распознающий язык L. Преобразования $\delta(p)$ и $\delta(q)$ совпадают тогда и только тогда, когда слова p и q конгруэнтны относительно L, то есть,*

$$\delta(p) = \delta(q) \Leftrightarrow p \equiv q(L).$$

Д о к а з а т е л ь с т в о. Если $\delta(p) = \delta(q)$, то при любых $r_1, r_2 \in X^*$ имеем:

$$\delta(r_1pr_2) = \delta(r_1)\delta(p)\delta(r_2) = \delta(r_1)\delta(q)\delta(r_2) = \delta(r_1qr_2).$$

Отсюда следует цепочка импликаций

$$s_0\delta(r_1pr_2) = s_0\delta(r_1qr_2) \Rightarrow (r_1pr_2 \in L \Leftrightarrow r_1qr_2 \in L) \Rightarrow p \equiv q(L).$$

Теперь пусть $\delta(p) \neq \delta(q)$, то есть, $s\delta(p) \neq s\delta(q)$ для некоторого состояния $s \in S$. Ввиду связности автомата найдётся такое слово r_1, что $s_0\delta(r_1) = s$. Следовательно, имеем

$$s_0\delta(r_1)\delta(p) = s_0\delta(r_1p) \neq s_0\delta(r_1q) = s_0\delta(r_1)\delta(q).$$

В силу приведённости автомата неравные состояния неэквивалентны. Поэтому существует такое слово r_2, что лишь одно из состояний

$$s_0\delta(r_1p)\delta(r_2),\ s_0\delta(r_1q)\delta(r_2)$$

принадлежит F. А это значит, что слова p и q не конгруэнтны относительно языка L. $\square$

Теорема 1. *Моноид минимального автомата* (S, δ, s_0, F)*, распознающего язык* L*, изоморфен синтаксическому моноиду языка* L*.*

Д о к а з а т е л ь с т в о. Определим отображение моноида автомата в синтаксический моноид, положив $\delta(p) \mapsto \langle p\rangle$. Корректность определения этого отображения и то, что оно инъективно, следует из леммы 3, а сюръективность очевидна. Осталась лёгкая проверка условия гомоморфизма:

$$\delta(p)\delta(q) = \delta(pq) \mapsto \langle pq\rangle = \langle p\rangle\langle q\rangle.$$

Итак, отображение, сопоставляющее преобразованию $\delta(p)$ класс конгруэнтности $\langle p\rangle$, в случае минимального автомата является изоморфизмом моноидов. $\square$

Следствие 1. *Синтаксический моноид языка* L *содержит конечное число элементов тогда и только тогда, когда язык* L *распознаётся конечным автоматом. При этом, если ранг языка равен* n*, то число элементов синтаксического моноида не больше* n^n*.*

Д о к а з а т е л ь с т в о. Предположим, что синтаксический моноид языка L содержит m элементов, то есть, имеется m классов конгруэнтности $\equiv (L)$. Тогда среди любых $m+1$ слов из X^* есть хотя бы два слова, которые конгруэнтны и, значит, неразличимы относительно L. Следовательно, $\mathrm{rk}\, L \leq m$.

Теперь пусть $\mathrm{rk}\, L = n$. Моноид минимального автомата состоит из некоторых преобразований n-элементного множества состояний. Всего таких преобразований n^n. Отсюда и из теоремы 1 и следует верхняя граница для мощности синтаксического моноида. $\square$

Синтаксический моноид языка можно задать с помощью графа. Впрочем, технически удобнее описывать изоморфный ему граф моноида автомата.

Предположим, что минимальный автомат, распознающий язык L, построен. Рассмотрим орграф, вершины которого — различные автоматные операторы $\delta(p)$, причём из вершины $\delta(p)$ в вершину $\delta(px)$ ведёт дуга с меткой x. Тогда путь с началом $\delta(e)$ в этом графе, помеченный словом p, ведёт в вершину $\delta(p)$. Так мы получаем удобный способ находить $\delta(p)$ для любого слова p, прослеживая путь в графе моноида, отвечающий этому слову. Для наглядного представления оператора упорядочим состояния автомата: $s_1, \ldots, s_n$. Тогда оператор $\delta(p)$ полностью определяется списком $s_1\delta(p), \ldots, s_n\delta(p)$. Этот список и будем использовать в качестве обозначения оператора $\delta(p)$.

Рекомендуем читателю убедиться в том, что на рис. 7 изображён граф моноида автомата из примера 1 §2 (квадратные скобки в обозначениях состояний опущены).

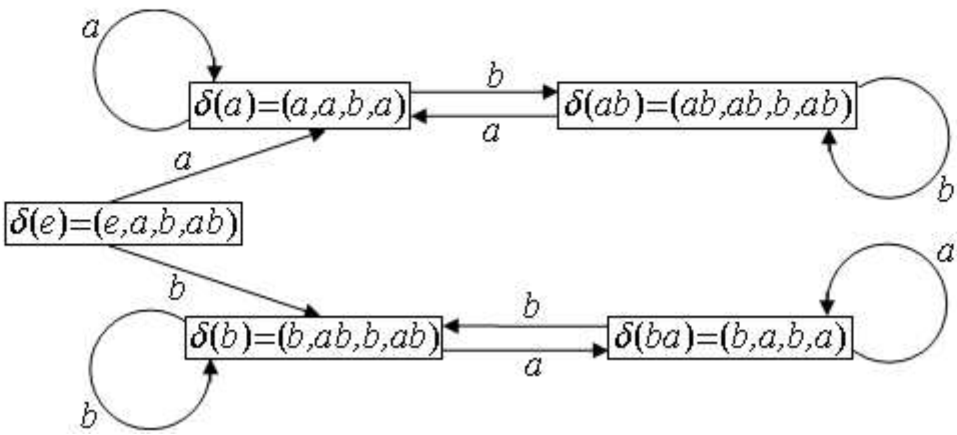


Рис. 7

Описанный способ представления синтаксического моноида удобен, если язык конечно распознаваем. Но и в противном случае устройство синтаксического моноида может быть прозрачным, как в следующей задаче.

Задача 1. Доказать, что синтаксический моноид языка L_8 (см. конец §4) изоморфен аддитивному моноиду целых чисел.

Следующие две задачи связывают свойства синтаксического моноида со свойствами его графа.

Задача 2. Мультипликативный моноид называется нильпотентным, если он содержит такой элемент 0, что $0x = x0 = 0$ для любого x, причем существует такое k, что произведение любых k элементов моноида равно 0. Доказать, что синтаксический моноид нильпотентен тогда и только тогда, когда все пути в графе моноида, длина которых не меньше числа его вершин, заканчиваются в одной и той же вершине.

Задача 3. Доказать, что синтаксический моноид тогда и только тогда является группой, когда его граф сильно связен.

Литература

1. **Оре О.** Теория графов. — М.: Наука, 1980.
2. **Харари Ф.** Теория графов — М.: Мир, 1973.
3. **Емеличев В.А., Мельников О.И., Сарванов В.И., Тышкевич Р.И.** Лекции по теории графов. — М.: Наука, 1990.
4. **Уилсон Р.** Введение в теорию графов.— М.: Мир, 1977.
5. **Асанов М.О., Баранский В.А., Расин В.А.** Дискретная математика: графы, матроиды, алгоритмы. — Москва, Ижевск: РХД, 2001.
6. **Арбиб М.А.(ред.)** Алгебраическая теория автоматов, языков и полугрупп. — М.: Статистика, 1975.
7. **Кудрявцев В.Б., Алёшин С.В., Подколзин А.С.** Введение в теорию автоматов. — М.: Наука, 1985.