

Лекция 2.3

Основные понятия и представления теории информации

(2 ч.)

План

1. Меры информации
2. Основные понятия системной классификации объектов
3. Виды системной классификации документированной информации
4. Система кодирования информации
5. Элементы теории графов

4. Система кодирования информации

Система кодирования информации

Код (франц. code) - универсальный способ, закон отображения информации при ее обработке (хранении, передаче, приеме, переработке) в виде системы однозначных соответствий между элементами сообщений и сигналами, при помощи которых эти элементы зафиксированы.

Система кодирования информации

- Код строится на базе алфавита, состоящего из букв, цифр и других символов и характеризуется:
- *значимостью* - число символов в кодовой комбинации
- *длиной* - число позиций (символов) в коде;
- *основанием* - число символов, букв однозначно различимых качественных признаков алфавита;
- *структурой* — порядок расположения в коде символов, используемых для обозначения классификационного признака;
- *весом* - число ненулевых символов.

Система кодирования информации

В зависимости от значений этих показателей бывают следующие виды кодов: двоичные, восьмеричные, шестнадцатеричные, равномерные, неравномерные, позиционные, непозиционные и др.

Процедура присвоения объекту кодового обозначения называется *кодированием* и сводится к однозначному преобразованию символов одного алфавита в другой по определенному правилу, закону, алгоритму.

Система кодирования информации

- **Первичный алфавит** - исходный, кодируемый алфавит, обладающий определенным числом качественных признаков (буквы алфавита, наборы символов, и др.), m_1 .
- **Вторичный алфавит** - набор однозначно различимых качественных признаков m_2 , обладающих необходимыми физическими свойствами для **перемещения** символов первичного алфавита в пространстве и во времени.

Закон преобразования символов первичного алфавита во вторичный можно записать в виде $m_1 \leq m_2^n$, где n - длина комбинаций кода во вторичном алфавите.

Код представляет полный набор всех возможных комбинаций символов вторичного алфавита, построенных по данному закону.

Система кодирования информации

Цели кодирования:

- преобразование информации в систему символов (кодов), обеспечивающую простоту и надежность аппаратной реализации информационных услуг и их эффективность;
- согласование свойств источника сообщений со свойствами канала связи (по Шеннону);
- эффективное кодирование, при котором путем устранения избыточности существенно снижается среднее число символов, требующихся на букву сообщения, что дает выигрыш во времени передачи или в объеме запоминающих устройств;
- сжатие входной информации;
- обеспечение заданной достоверности передачи или хранения информации путем внесения избыточности с учетом интенсивности и статистических закономерностей помехи в канале связи.

Система кодирования информации

Виды и способы кодирования

Можно выделить две группы систем кодирования:

- *классификационную систему кодирования, ориентированную на проведение предварительной классификации объектов либо на основе иерархической системы, либо на основе фасетной системы;*
- *регистрационную систему кодирования, не требующую предварительной классификации объектов.*

Система кодирования информации

Система кодирования для систем классификации информации:

Классификационная

- Последовательная
- Параллельная

Регистрационная

- Побуквенная
- Позиционная
- Порядковая
- Пословная
- Серийно-порядковая

Система кодирования информации

Классификационное кодирование применяется после проведения классификации объектов.

Последовательное (линейное) кодирование [in-line-coding] - представление алгоритма в виде последовательности не образующих циклы операторов (команд).

Система кодирования информации

Для иерархической классификационной структуры содержание такого вида кодирования заключается в последовательной записи кода старшей группировки 1-го уровня, затем кода группировки 2-го уровня, затем кода группировки 3-го уровня и т.д.

В результате получается кодовая комбинация, каждый разряд которой содержит информацию о специфике выделенной группы на каждом уровне иерархической структуры. Последовательная система кодирования обладает теми же достоинствами и недостатками, что и иерархическая система классификации.

Система кодирования информации

Пример. Проведем кодирование информации, классифицированной с помощью иерархической схемы информационной системы "факультет".

Количество кодовых группировок будет определяться глубиной классификации и равно 3. Прежде чем начать кодирование, необходимо определиться с алфавитом, т.е. какие будут использоваться символы.

Для большей наглядности выберем десятичную систему счисления - 10 арабских цифр. Анализ схемы на рис. 2.4 показывает, что длина кода определяется 3 десятичными разрядами, а кодирование группировки на каждом уровне можно делать путем последовательной нумерации слева направо. В общем виде код можно записать как XXX, где X - значение десятичного разряда.

Система кодирования информации

Рассмотрим структуру кода, начиная со старшего разряда:

- - 1-й (старший) разряд выделен для классификационного признака "название факультета" и имеет следующие значения: 1 - юридический; 2 - управленческий; 3 - для следующего названия факультета и т.д.;
- 2-й разряд выделен для классификационного признака "возраст" и имеет следующие значения: 1 - до 20 лет; 2 - от 20 до 30 лет; 3 - свыше 30 лет;
- 3-й разряд выделен для классификационного признака "пол" и имеет следующие значения: 1 - мужчины; 2 - женщины

Принятая система кодирования позволяет легко расшифровать любой код группировки, например:

- 1310 - студенты юридического факультета, свыше 30 лет, мужчины;
- 2221 - студенты управленческого факультета, от 20 до 30 лет, женщины.

Система кодирования информации

Параллельное кодирование [parallelcoding] - вид многоаспектного кодирования свойств объектов, выполняемого на основе предварительной фасетной классификации свойств в пределах каждого признака.

Содержание этого вида кодирования заключается в следующем:

- все фасеты кодируются независимо друг от друга;
- для значений каждого фасета выделяется определенное количество разрядов кода.

Параллельная система кодирования обладает теми же достоинствами и недостатками, что и фасетная система классификации.

Система кодирования информации

Пример. Проведем кодирование информации, классифицированной с помощью фасетной схемы (табл. 3).

Количество кодовых группировок определяется количеством фасетов и равно 4. Выберем десятичную систему счисления в качестве алфавита кодировки, что позволит для значений фасетов выделить один разряд и иметь длину кода, равную 4. В отличие от последовательного кодирования для иерархической системы классификации, здесь не имеет значения порядок кодировки фасетов. В общем виде код можно записать как XXXX, где X — значение десятичного разряда.

Система кодирования информации

Рассмотрим структуру кода, начиная со старшего разряда:

- 1-й (старший) разряд выделен для фасета "пол" и имеет следующие значения: 1 - мужчины; 2 - женщины;
- 2-й разряд выделен для фасета "наличие детей у женщин" и имеет следующие значения: 1 - есть дети; 2 - нет детей, 0 - для мужчин, так как подобной информации не требуется;
- 3-й разряд выделен для фасета "возраст" и имеет следующие значения: 1 - до 20 лет; 2 - от 20 до 30 лет; 3 - свыше 30 лет;
- 4-й разряд выделен для фасета "название факультета" и имеет следующие значения: 1 - радиотехнический, 2 - машиностроительный, 3 - коммерческий; 4 - информационные системы; 5 - математический и т.д.

Система кодирования информации

Для такой системы кодирования можно записать код группировки:

- 2135 - женщины в возрасте свыше 30 лет, имеющие детей и являющиеся студентами математического факультета;
- 1021 - мужчины возраста от 20 до 30 лет, являющиеся студентами радиотехнического факультета.

Система кодирования информации

- **Регистрационное кодирование** используется для однозначной идентификации объектов и не требует предварительной классификации объектов.
- **Позиционное кодирование** [positional coding] - способ кодирования реквизитов признаков, применяющих фиксированное число значений, при котором длина кодовой комбинации устанавливается равной числу возможных значений реквизита.
- **Побуквенное кодирование** - способ кодирования реквизитов, состоящий в последовательном кодировании каждого символа и применяемый при передаче сообщений по линиям телекоммуникаций. Реквизиты-признаки - нечисловые данные (цвет, марка, фамилия и др.)

Система кодирования информации

Порядковое кодирование [serialcoding] - кодирование реквизитов-признаков, при котором все кодируемые значения сведены в список и кодовой комбинацией каждого значения является его порядковый номер в списке.

Это кодирование предполагает последовательную нумерацию объектов числами натурального ряда. Такая нумерация может быть случайной или определяться после предварительного упорядочения объектов, например по алфавиту.

Порядковое кодирование применяется в том случае, когда количество объектов невелико, например кодирование названий факультетов университета, кодирование студентов в учебной группе.

Система кодирования информации

- **Пословное кодирование** [word-serialcoding] - способ кодирования реквизитов-признаков, состоящий в последовательном кодировании каждого слова (а не буквы) входного документа. Это кодирование требует семантического анализа и, как правило, выполняется вручную.
- **Серийно-порядковое** кодирование - порядковое кодирование, при котором последовательность порядковых номеров - кодов делится на группы-серии, объединяющие объекты по какому-либо признаку.

Пример. Все студенты одного факультета разбиваются на учебные группы - серии, для которых используется порядковая нумерация. Внутри каждой группы производится упорядочение фамилий студентов по алфавиту и каждому студенту присваивается номер.

5. Элементы теории графов

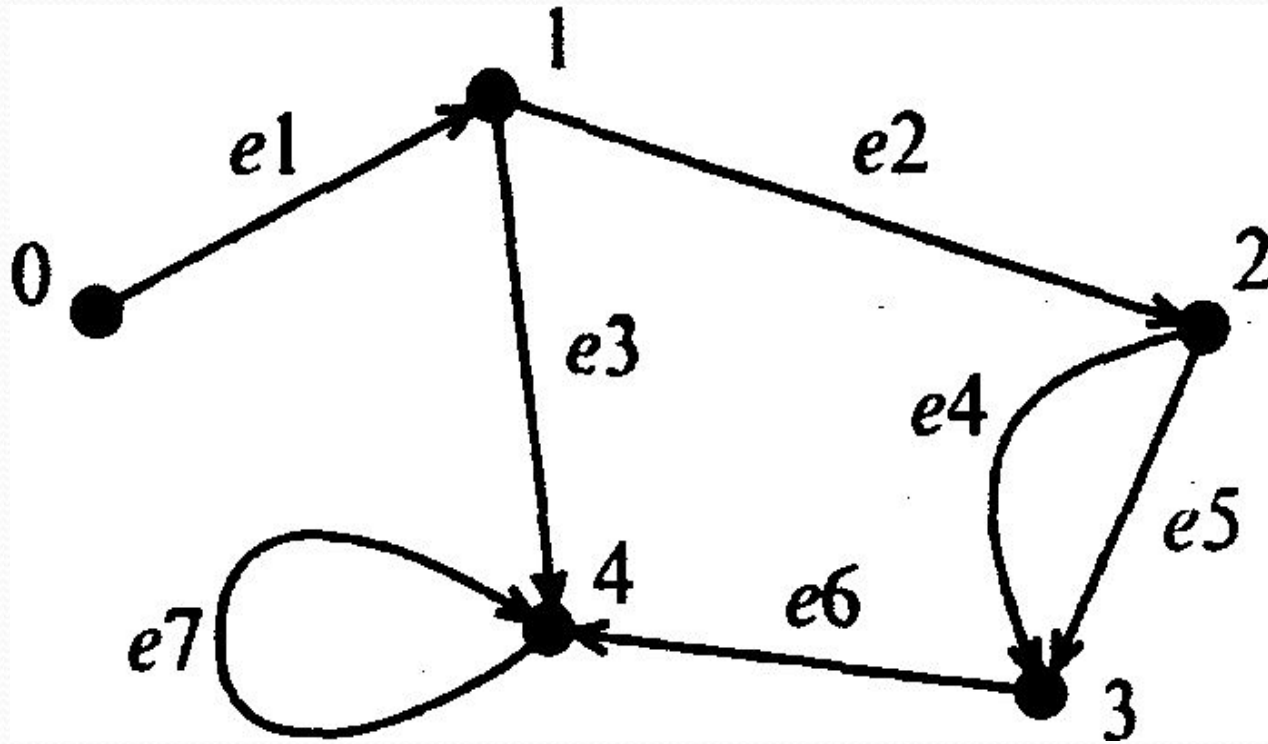
Элементы теории графов

Такая структура, как **граф** (в качестве синонима используется также термин «сеть»), имеет самые различные применения в информатике и в смежных прикладных областях.

Граф $G = (V, E)$ задается парой конечных множеств V и E . Элементы первого множества v_1, v_2, \dots, v_M называются **вершинами** графа (при графическом представлении им соответствуют точки). Элементы второго множества e_1, e_2, \dots, e_N называют **ребрами**. Каждое ребро определяется парой вершин (при графическом представлении ребро соединяет две вершины графа). Если ребра графа определяются упорядоченными парами вершин, то такой граф называют **ориентированным** (на чертеже при изображении ориентированного графа на каждом ребре ставят стрелку, указывающую его направление).

Элементы теории графов

Ориентированный граф с пятью вершинами и семью ребрами изображен на рис.



Элементы теории графов

Если две вершины соединены двумя или более ребрами, то эти ребра называют параллельными (например, ребра e_4 и e_5). Если начало и конец ребра совпадают, то такое ребро называется **петлей** (например, ребро e_7). Граф без петель и параллельных ребер называется простым.

Если ребро e_k определяется вершинами v_i и v_j (будем обозначать этот факт следующим образом: $e_k = (v_i, v_j)$), то говорят, что ребро e_k **инцидентно** вершинам v_i и v_j . Две вершины v_i и v_j называются смежными, если в графе существует ребро (v_i, v_j) .

Последовательность вершин $v_{i1}, v_{i2}, \dots, v_{ik}$, таких, что каждая пара $(v_{i, (j-1)}, v_{ij})$ при $1 < j \leq k$ определяет ребро, называется **маршрутом** в графе G . Вершины v_{i1} и v_{ik} называются **концевыми** вершинами маршрута, все остальные входящие в него вершины - **внутренними**.

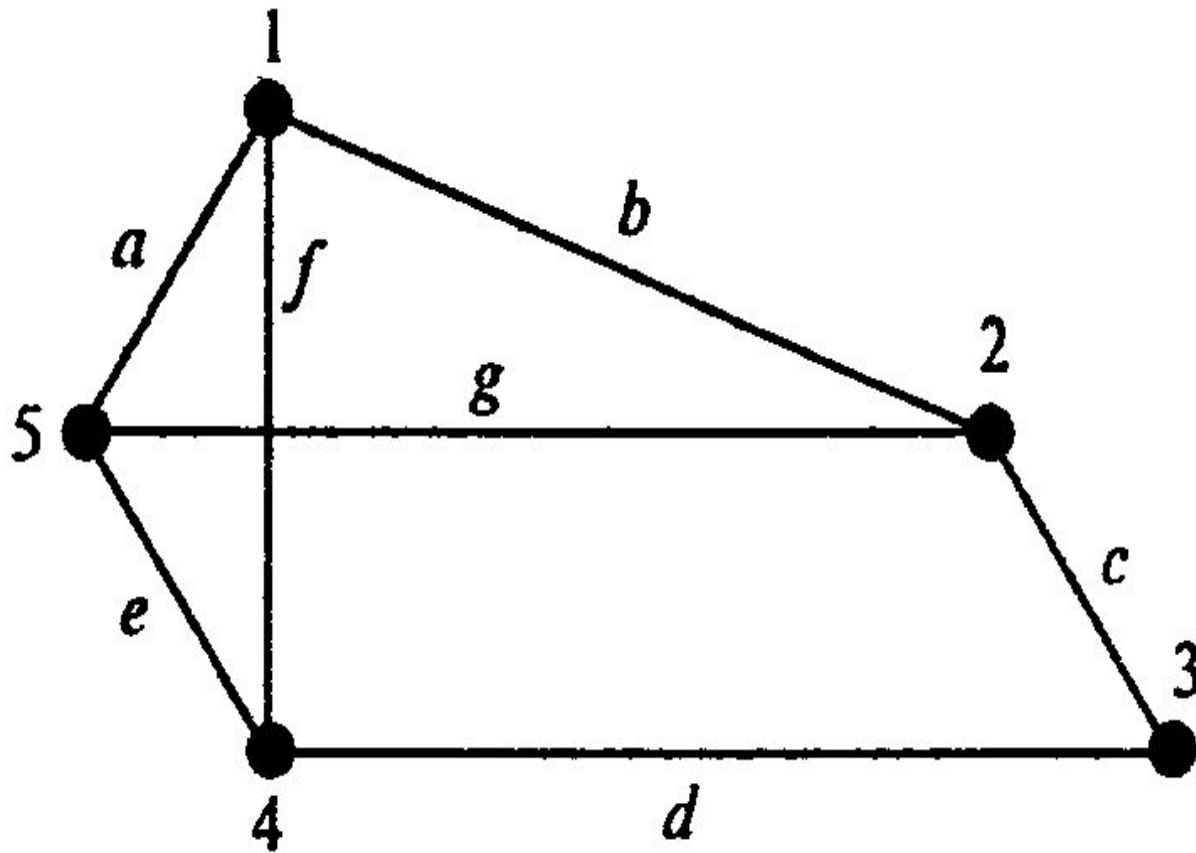
Элементы теории графов

Маршрут, в котором все определяемые им ребра различны, называют цепью. Цепь считают замкнутой, если ее концевые вершины совпадают. Замкнутая цепь, в которой все вершины (за исключением концевых) различны, называется циклом. Незамкнутая цепь, в которой все вершины различны, носит название путь. Если в ориентированном графе существует путь из v_i в v_j , то говорят, что вершина v_j достижима из вершины v_i .

Две вершины v_i и v_j называют связанными в графе G , если в нем существует путь, для которого эти вершины являются концевыми. Граф G называется связным, если каждые две вершины в нем являются связанными. На рис. изображен простой неориентированный связный граф.

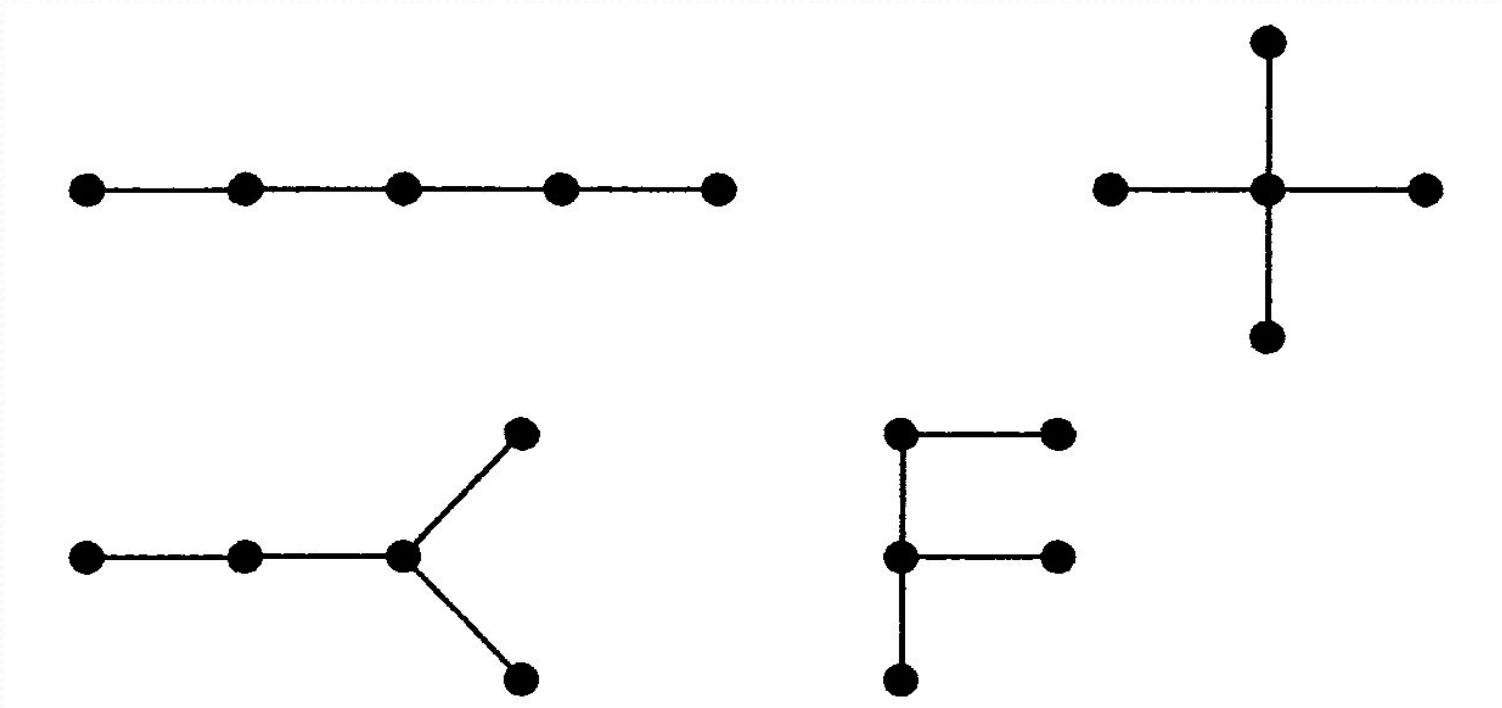
Элементы теории графов

Пример неориентированного связного графа



Элементы теории графов

Последовательность вершин v_1, v_5, v_4, v_3 , например, определяет путь, а последовательность вершин $v_1, v_5, v_4, v_3, v_1, v_1$ - цикл. **Деревом** будем называть неориентированный связный граф без циклов. **Лес** - это любой граф без циклов. На рис. показаны возможные деревья с пятью вершинами.



Элементы теории графов

Анализ приведенных здесь понятий и определений показывает, что в качестве моделей графы удобно использовать в тех случаях, когда рассматривается система каких-либо объектов, между которыми существуют определенные связи, отношения, когда изучается структура системы, возможности ее функционирования. В информатике графы используются в разделах: операционные системы, алгоритмизация, структуры данных, информационное моделирование и др.

Элементы теории графов

ПРЕДСТАВЛЕНИЕ ГРАФОВ

Важным вопросом, особенно для приложений теории графов, является определение возможных способов представления графов. Самый простой способ - полное перечисление множеств V и E . Однако, очевидно, что в этом случае выявление у графа различных характеристик и свойств будет крайне затруднительным. Граф можно представить в виде некоторого графического изображения и визуально определить некоторые свойства и характеристики заданного графа. Однако, при наличии в графе большого числа ребер и вершин этот способ также мало пригоден.

Элементы теории графов

Рассматривая различные возможные способы представления графов, мы должны иметь в виду потребность ввода соответствующей информации в компьютер. В этой связи ввод информации в числовом виде предпочтителен, хотя современные технические средства допускают ввод и графической информации (таблиц, текста, графиков, рисунков и т.д.), после чего может производиться обработка такой информации.

Элементы теории графов

Матрица смежности. Если вершины графа G помечены метками v_1, v_2, \dots, v_n , то элементы матрицы смежности $A(G)$ размера $V, \times V$ определяются следующим образом: $A(i,j) = 1$, если v_i смежна с v_j ; $A(i,j) = 0$ в противном случае (рис. а).

Матрица инцидентности. Если вершины графа G помечены метками v_1, v_2, \dots, v_m , а ребра - метками e_1, e_2, \dots, e_n , то элементы матрицы инцидентности $I(G)$ размера $M \times N$ определяются правилом: $B(i,j) = 1$, если v_i инцидентна e_j ; $B(i,j) = 0$ в противном случае (см. рис. б).

а

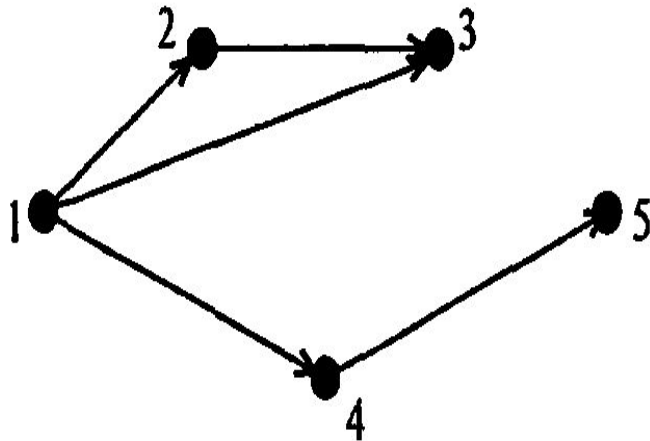
	1	2	3	4	5
1	0	1	0	1	1
2	1	0	1	0	1
3	0	1	0	1	0
4	1	0	1	0	1
5	1	1	0	1	0

б

	a	b	c	d	e	f	g
1	1	1	0	0	0	1	0
2	0	1	1	0	0	0	1
3	0	0	1	1	0	0	0
4	0	0	0	1	1	1	0
5	1	0	0	0	1	0	1

Элементы теории графов

Для ориентированного графа G , имеющего N вершин можно рассмотреть матрицу достижимости $C(G)$ размера $N \times N$, элементы которой определяются так: $C(I, J) = 1$, если вершина v_j достижима из v_i ; $C(I, J) = 0$ в противном случае. Ниже приведен пример ориентированного графа и его матрицы достижимости, рис.



$C(G)$:

	1	2	3	4	5
1	0	1	1	1	1
2	0	0	1	0	0
3	0	0	0	0	0
4	0	0	0	0	1
5	0	0	0	0	0