ВВЕДЕНИЕ В БИОИНФОРМАТИКУ

Лекция №6

Элементы теории графов

Новоселецкий Валерий Николаевич
к.ф.-м.н., доц. каф. биоинженерии
valery.novoseletsky@yandex.ru

Сайт курса http://intbio.org/bioinf2018

Теория Графов

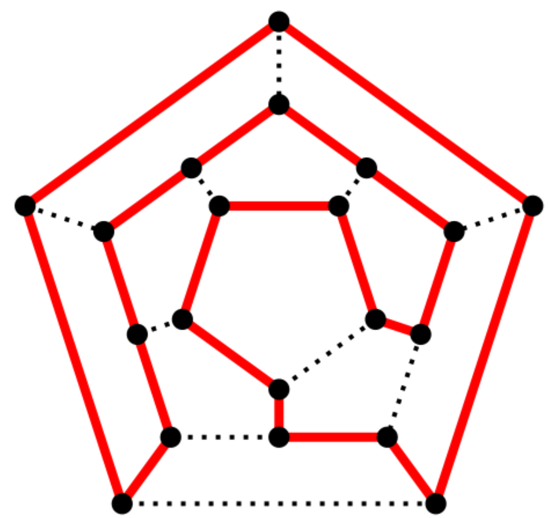
Alexander Lazarev

Institute of Control Sciences of Russian Academy of Sciences

2009-2010 учебный год

Задача коммивояжера
(Travelling salesman problem, TSP)

поиск самого выгодного маршрута, проходящего через указанные города хотя бы по одному разу с последующим возвратом в исходный город



У. Гамильтон, «Путешествие вокруг света» (1859)

Определения

Граф G = <V, E> есть совокупность множества вершин V и множества рёбер (дуг) E.

Граф называется неориентированным (неограф), если все его ребра неориентированы {x; y}, и ориентированным (орграф), если все его ребра <x; y> ориентированы. В случае произвольного графа ребра (x; y).

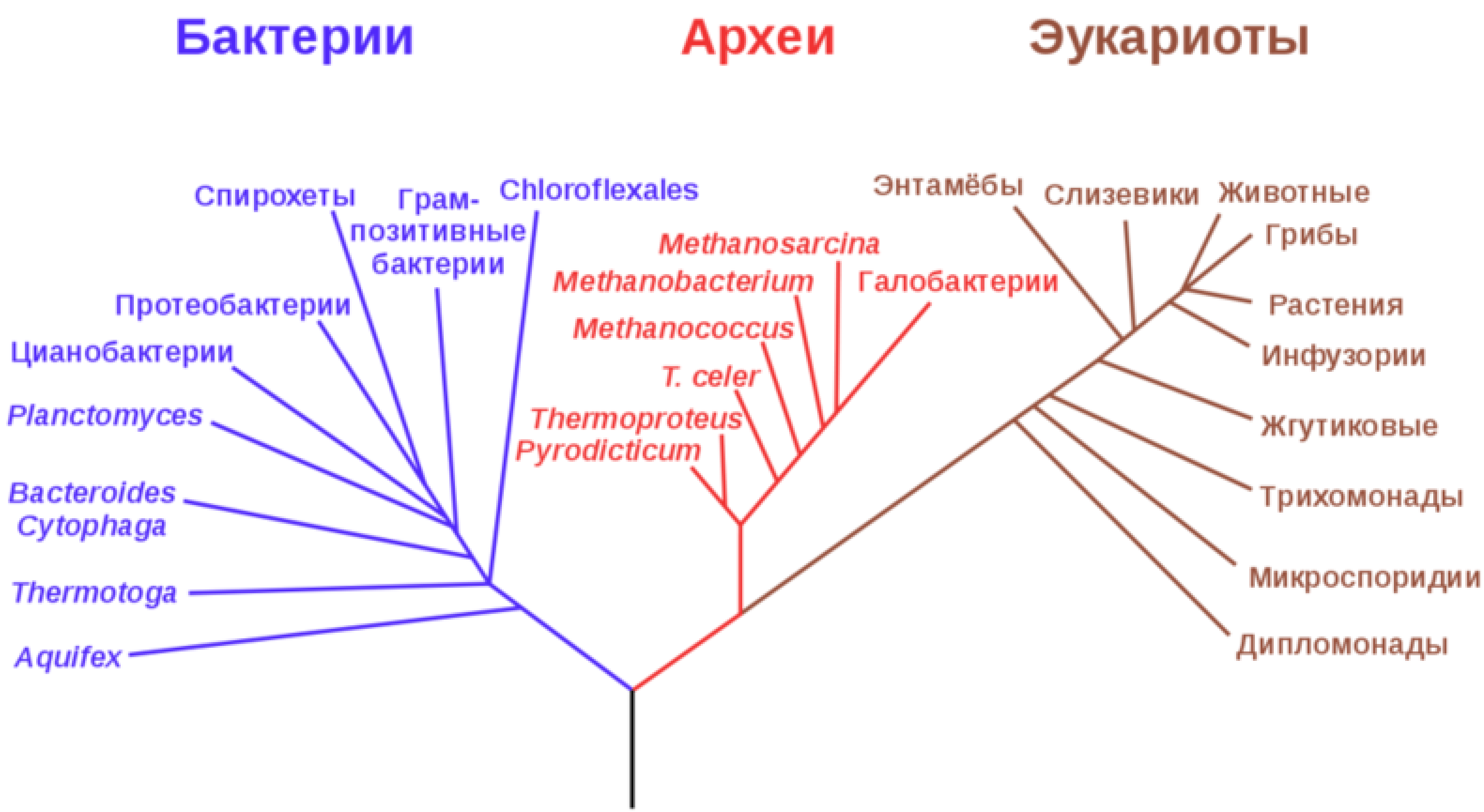
Ребро e инцидентно вершинам x и y, а вершины x и y инцидентны ребру e.

Вершины x и y смежны, если (x, y) является ребром. Два ребра смежны, если имеют общую вершину.

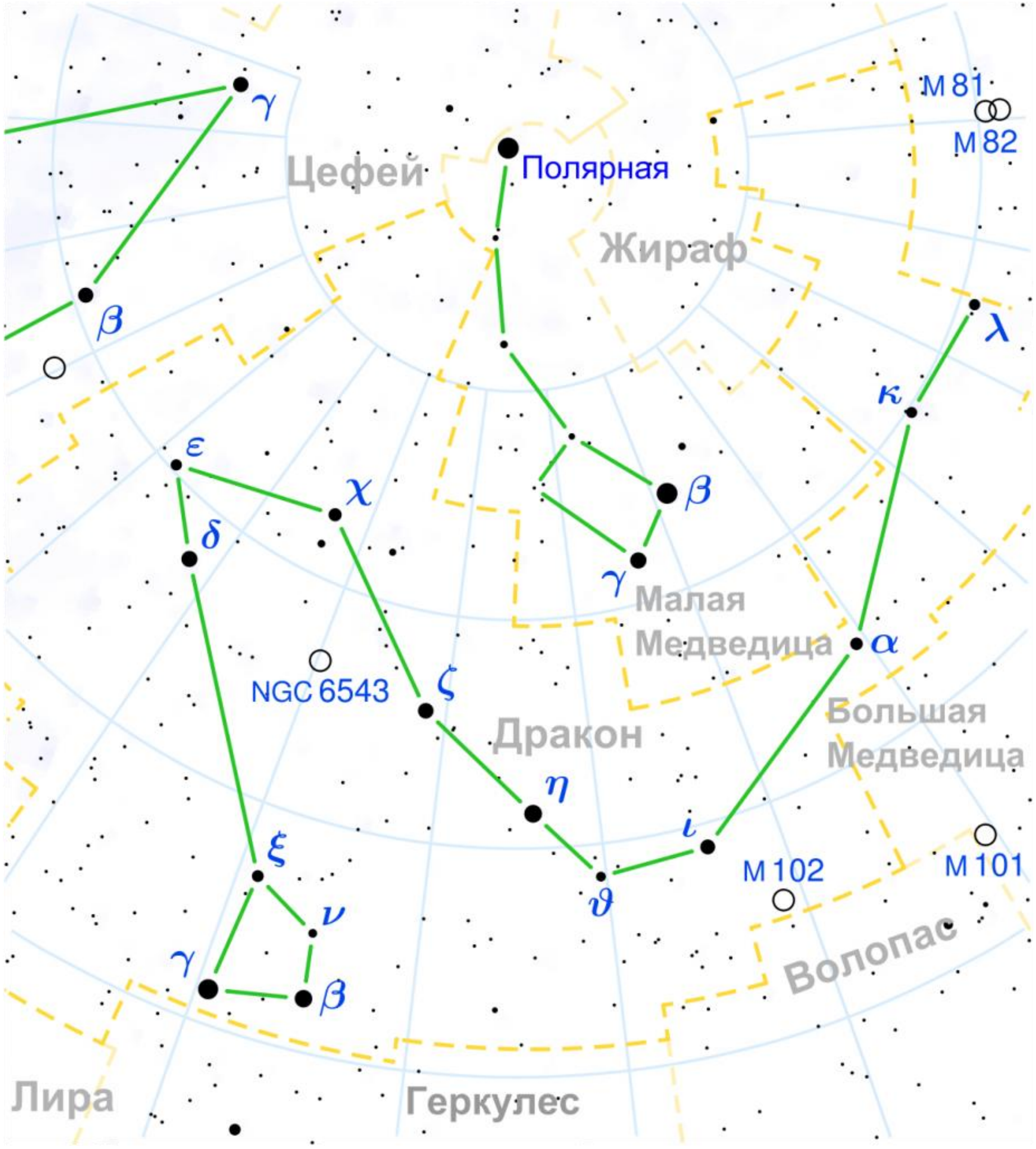
Вершина, не инцидентная никакому ребру, называется изолированной

Связный граф — граф, содержащий ровно одну компоненту связности. Это означает, что между любой парой вершин этого графа существует как минимум один путь.

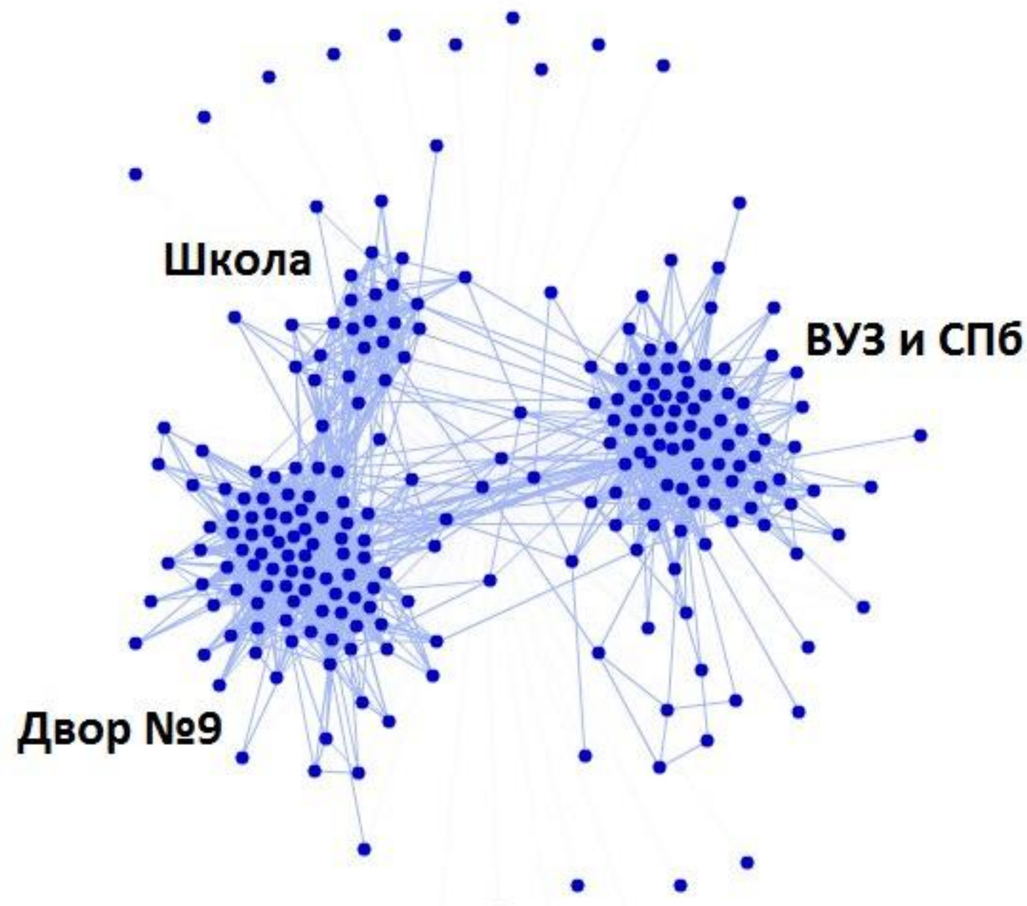
Примеры графов

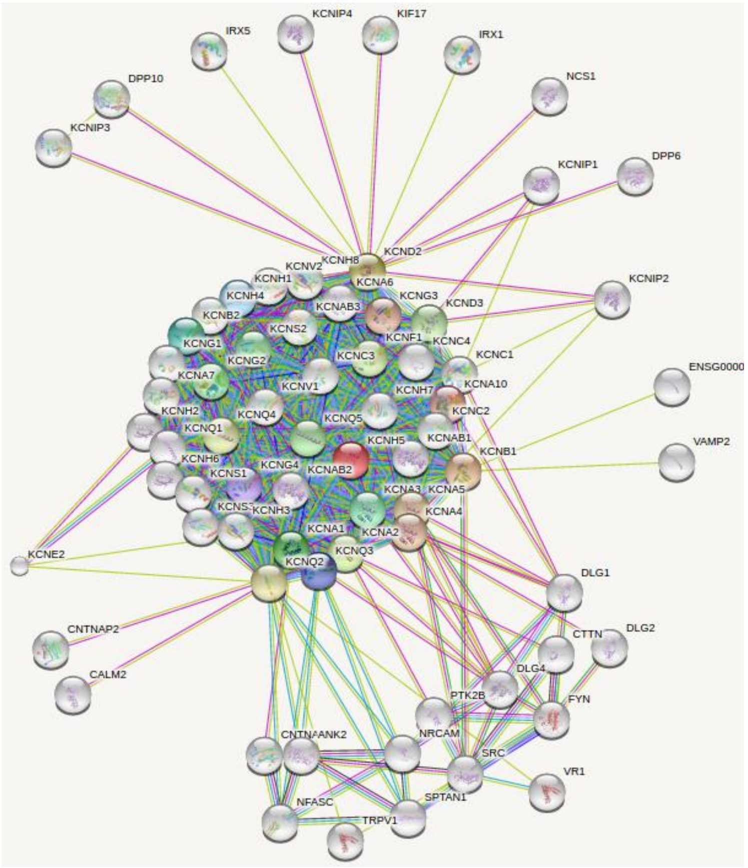


Примеры графов

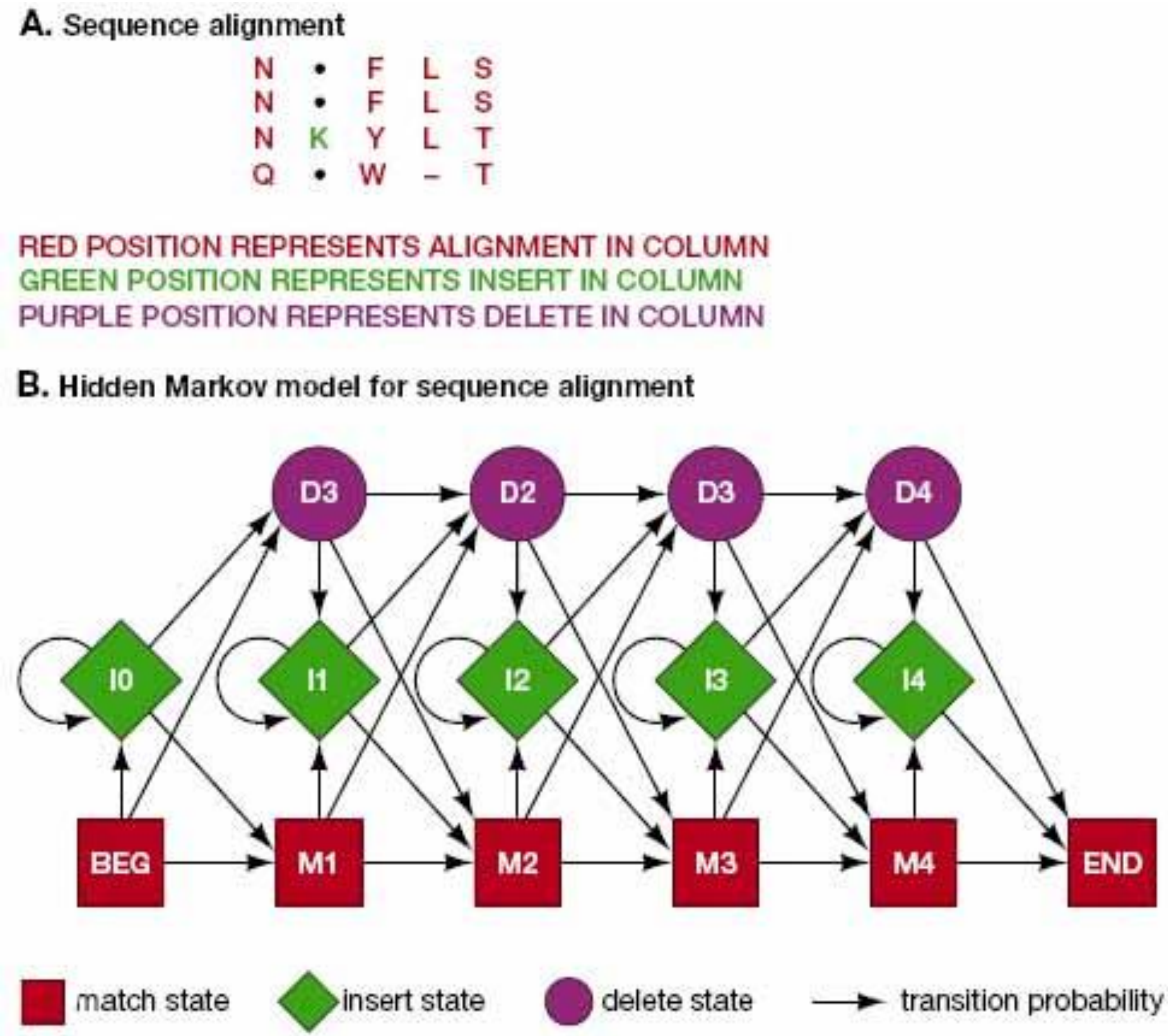


Примеры графов



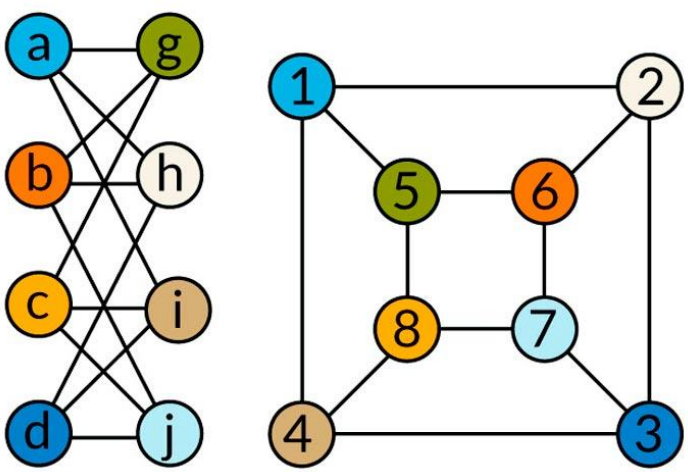


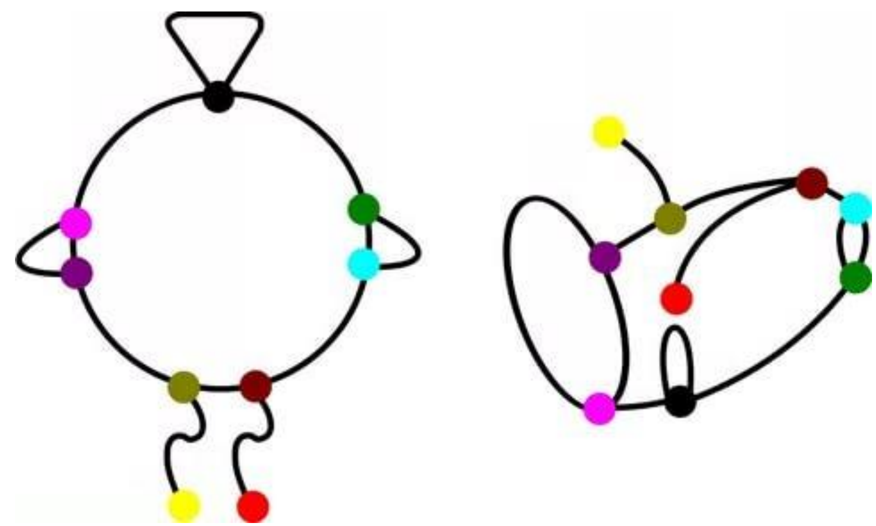
Примеры графов



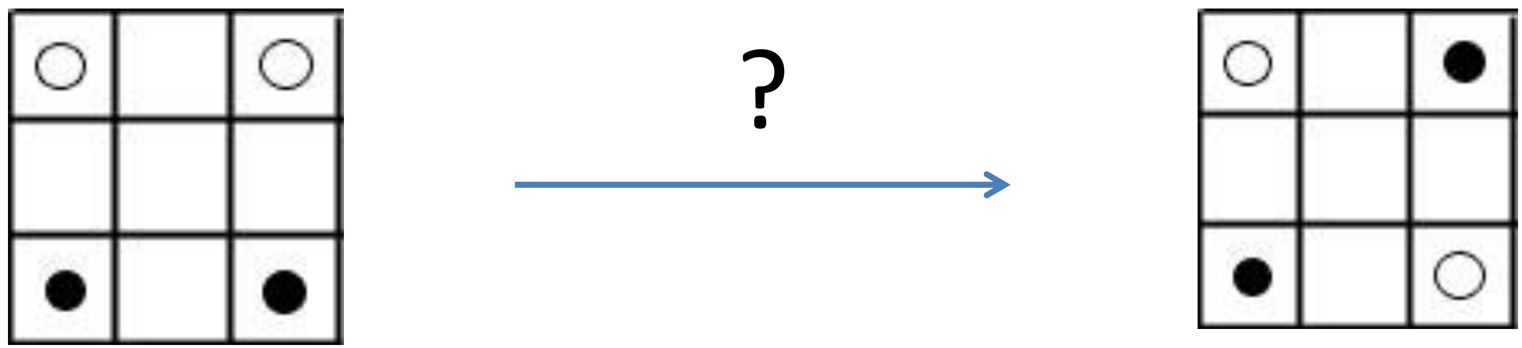
Определения

Два графа G = <V, E> и G’ = <V’, E’> изоморфны, если существует такое взаимно-однозначное соответствие между множествами их вершин V и V’, что вершины соединены ребрами в одном графе тогда и только тогда, когда соответствующие им вершины соединены ребрами в другом графе.

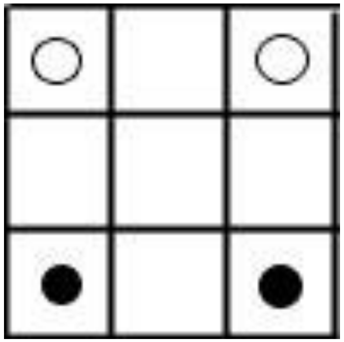




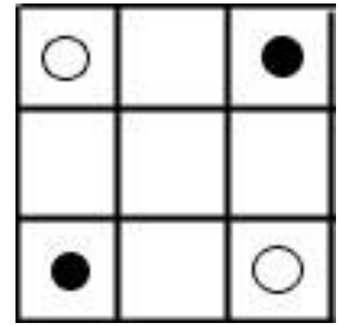
Задача о конях

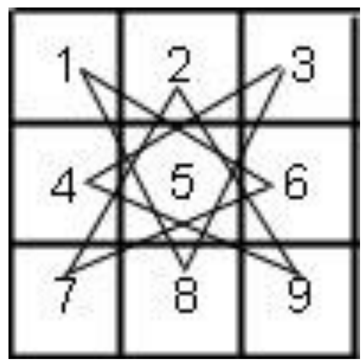


Задача о конях

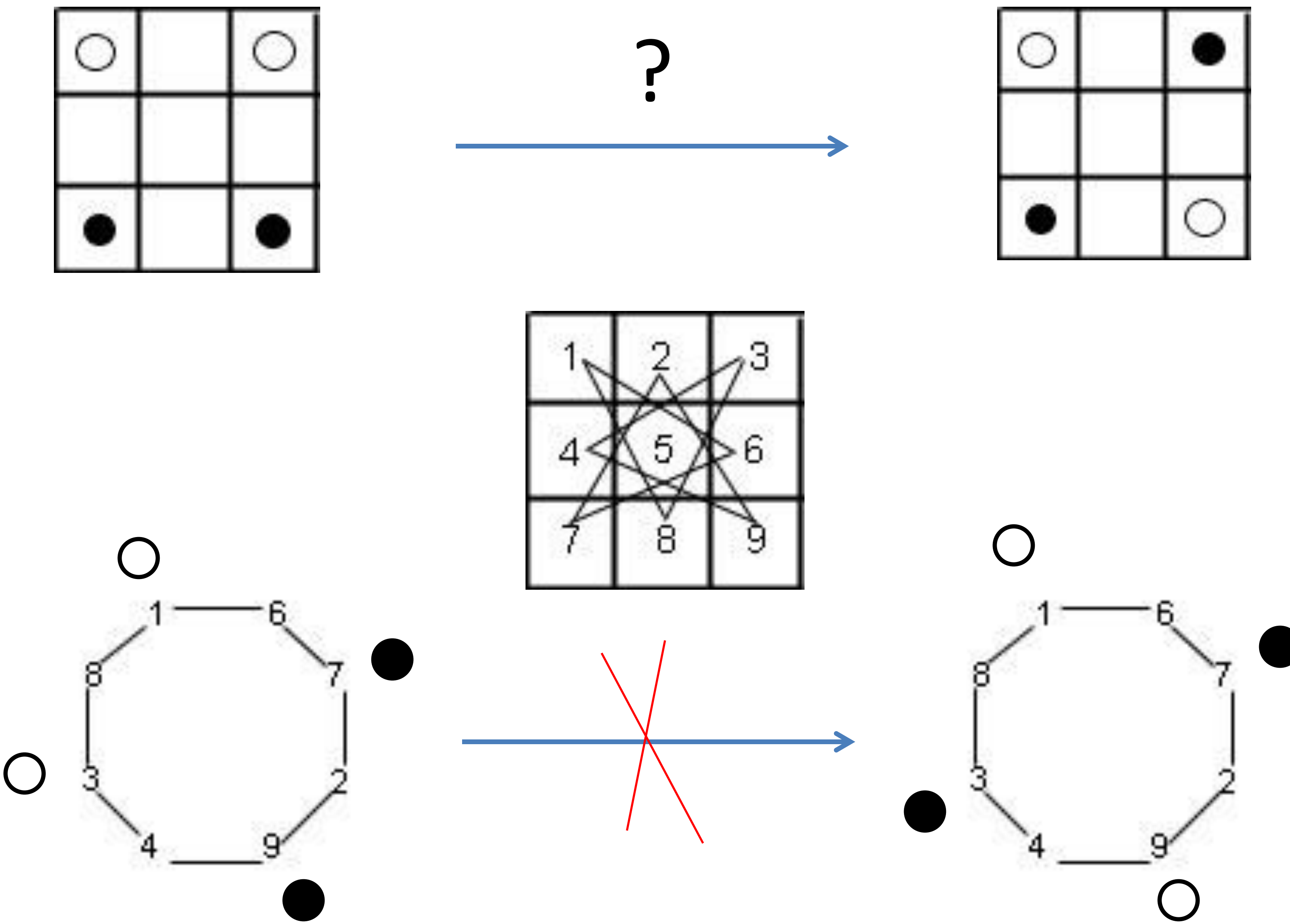


?



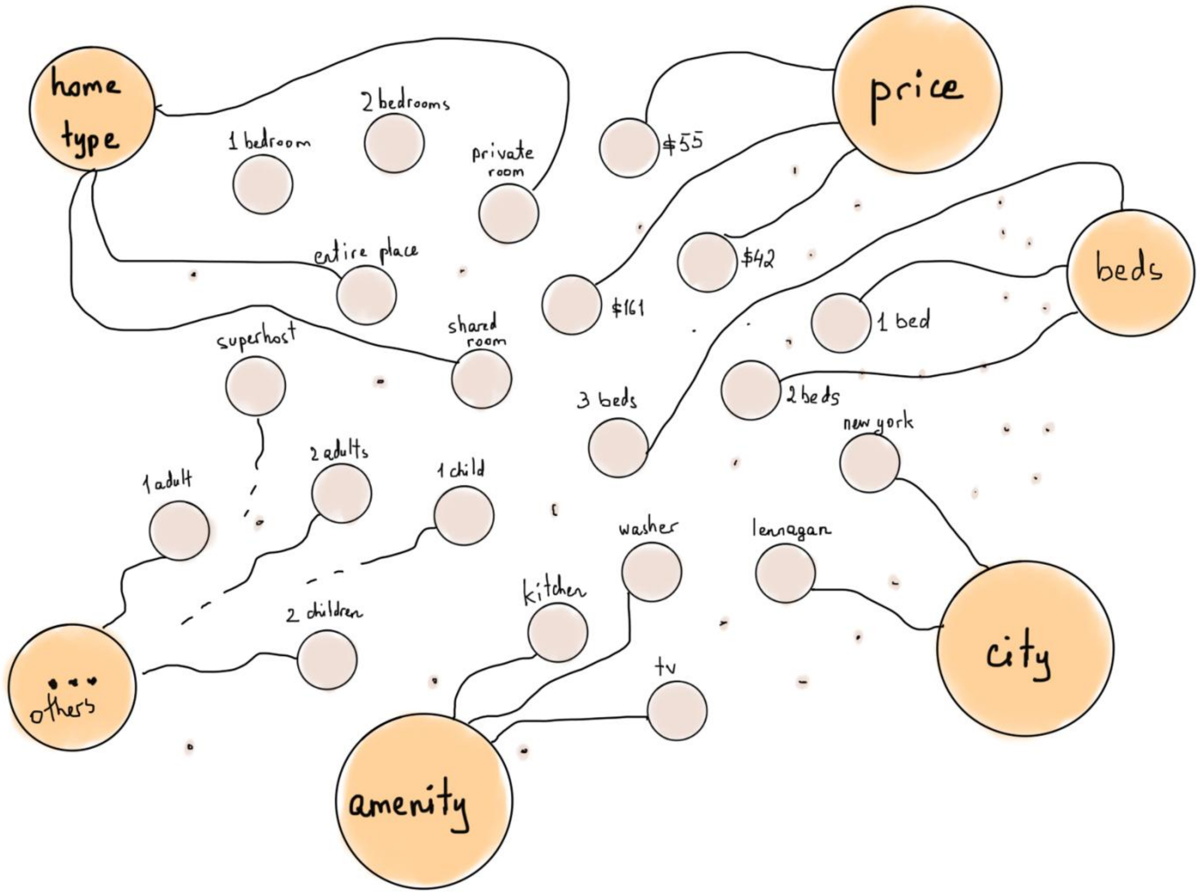


Задача о конях



Определения

Граф G = <V, E> называется двудольным, если множество его рёбер разбито на два подмножества: V = V1 U V2, V1 ∩ V2 = Ø, и рёбрами связаны только вершины из разных подмножеств.



Определения

Пусть G – неориентированный граф. Число ρ(x) ребер, инцидентных вершине x, называется степенью вершины. Отдельно следует оговаривать, считать ли петли однократными или двойными.

Пусть ρ(x, y) = ρ(y, x) – число ребер, соединяющих вершины x и y.

$$\rho(x) = \sum_{y \in V} \rho(x, y)$$

Обозначим через ne(G) число ребер в этом графе. Поскольку каждое ребро при подсчете общего числа степеней учитывается дважды (при вершине x и при вершине y), то

$$2ne(G) = \sum_{x \in V} \rho(x) = \sum_{x, y \in V} \rho(x, y)$$

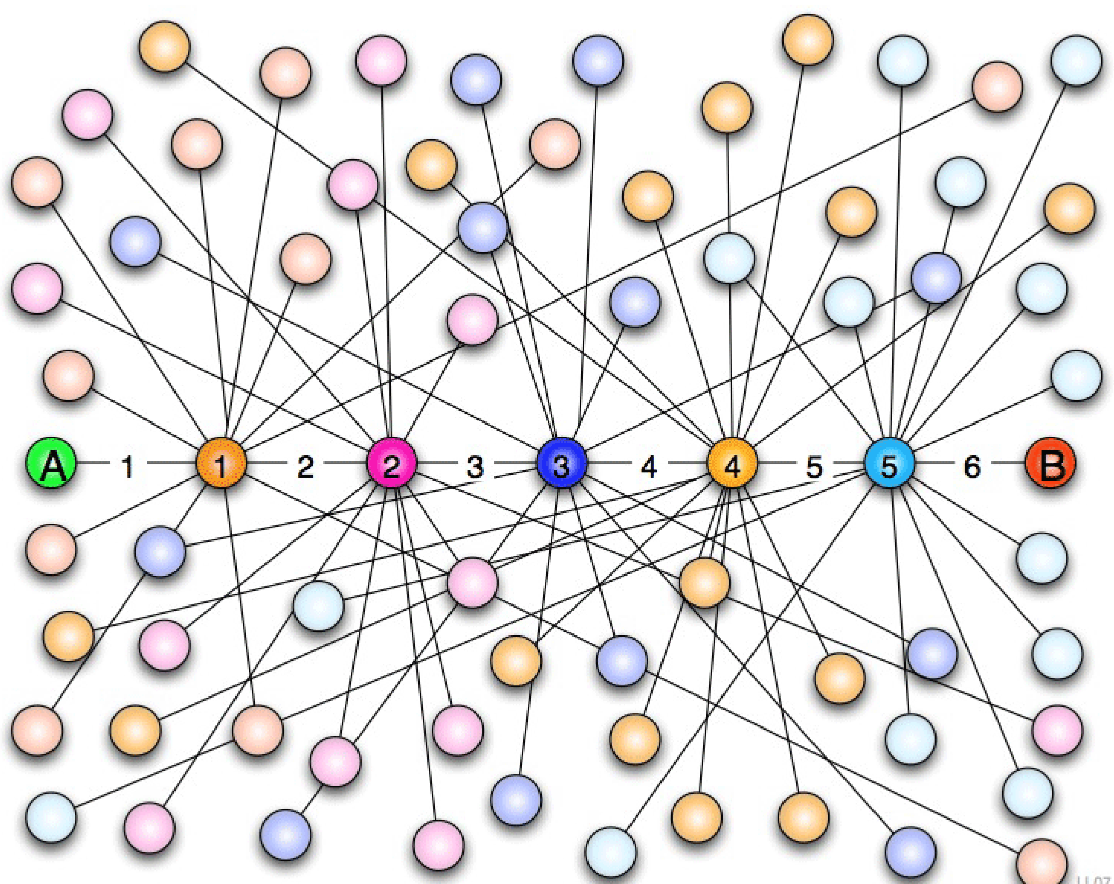
Формула справедлива и при наличии петель, если в степенях вершин учитывать их дважды.

Теорема 1 («Лемма о рукопожатиях»):

В конечном графе число вершин нечетной степени чётно.

Теория шести рукопожатий

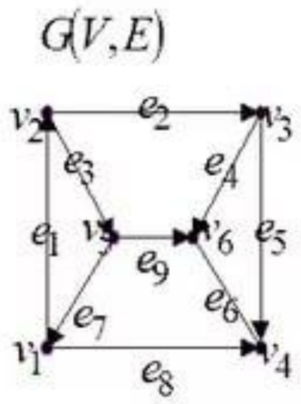
недоказанная теория, согласно которой любые два человека на Земле разделены не более чем пятью уровнями общих знакомых (и, соответственно, шестью уровнями связей).



Машинное представление графов

Матрица инцидентности – матрица со строками, соответствующими вершинам, и столбцами, соответствующим ребрам.

Для орграфов разные авторы используют разные варианты представления ребер <x; y>: либо (-1, 1), либо наоборот.



матрица инциденций графа

$$
\begin{array}{cc}
 & \begin{array}{ccccccccc} e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 & e_8 & e_9 \end{array} \\
A = \begin{array}{c} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{array} &
\begin{array}{ccccccccc}
1 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\
-1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & -1 & 0 & -1 & 0 \\
0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & -1
\end{array}
\end{array}
$$

Рисунок 3.3

Худший способ представления графа:

-требует m x n ячеек памяти

- неудобный доступ.

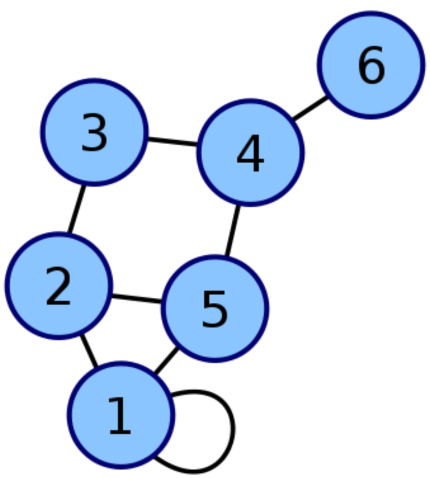
Проверка существования ребра между вершинами требует перебора всех столбцов

Машинное представление графов

Матрица смежности (вершин) – матрица n x n с индексами Bij = 1, если существует ребро из вершины i в вершину j и нулю в остальных случаях. Для неографа симметрична.

Проверка существования ребра между вершинами выполняется за один шаг.

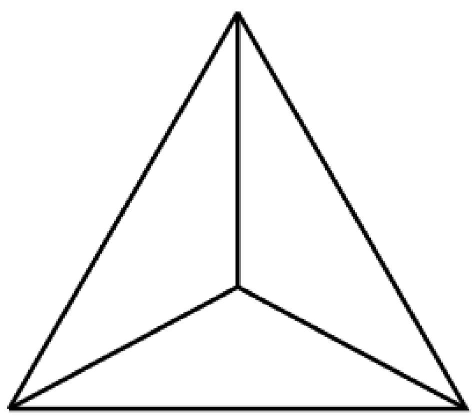
Прежний недостаток: требует n x n ячеек памяти.



$$\begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Машинное представление графов

Матрица смежности (вершин) – матрица n x n с индексами Bij = 1, если существует ребро из вершины i в вершину j и нулю в остальных случаях. Для неографа симметрична.



?

Машинное представление графов

Список инцидентности (ребер) – 2 столбца по m ячеек с указанием вершин, инцидентных данному ребру. Наиболее компактный способ представления графов.

Список смежности (вершин) – список вершин, смежных с данной. Требует O(m+n) ячеек для разреженных графов - наиболее удобный способ хранения.

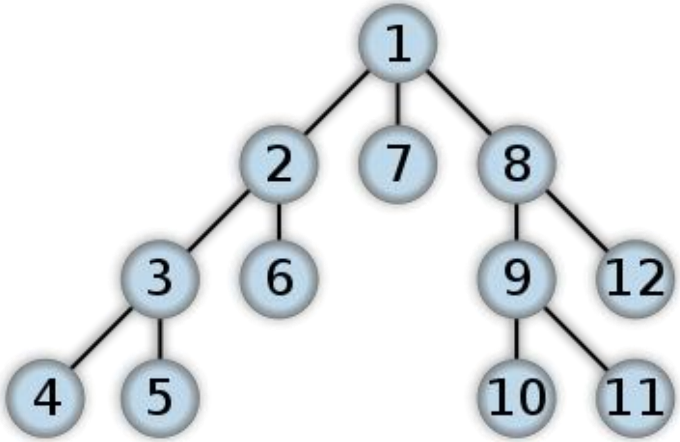
Поиск в графе

Граф называется простым, если не содержит петель и кратных рёбер.

Поиск в глубину:

Начинаем с некоторой вершины v0, затем переходим к произвольной вершине u, смежной с v0, и повторяем процесс. В общем случае, при нахождении в вершине v проверяем наличие ещё не посещённых смежных вершин. Если такие существуют, то переходим к одной из них и повторяем процедуру. В противном случае отмечаем вершину v, как посещённую, и возвращаемся в ту вершину, откуда попали в v.

Каждая вершина посещается не более одного раза и можно показать, что сложность этого алгоритма O(n + m).

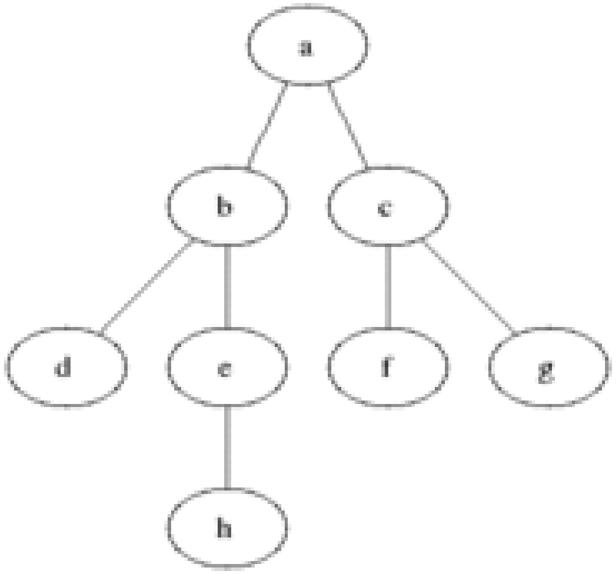


Поиск в графе

Поиск в ширину:

Начиная с узла-источника u, рассмотрим все рёбра (u,v), выходящие из узла u. Если очередной узел v является целевым узлом, то поиск завершается; в противном случае узел v добавляется в очередь. После того, как будут проверены все рёбра, выходящие из узла u, из очереди извлекается следующий узел u, и процесс повторяется. Сложность этого алгоритма также O(n + m).

Белый — вершина, которая еще не обнаружена. Серый — вершина, уже обнаруженная и добавленная в очередь. Черный — вершина, извлечённая из очереди.



Пути и циклы в графах

Путем в графе G = <V, E> называют последовательность рёбер вида <e1, e2, …, e_n-1> = S = <(v1, v2), (v2, v3), …, (v_n-1, vn)>. Говорят, что этот путь идёт из v1 в vn и имеет длину (n-1).

Если веса рёбер равны 1!

Путь называют простым, если все рёбра и вершины на нём различны, кроме быть может первой и последней.

Цикл – это простой путь длины не менее 1, который начинается и заканчивается в одной вершине. В простом неографе длина цикла не меньше 3.

Для нахождения пути между заданными вершинами u и v могут быть использованы оба вида поиска: в глубину и в ширину. При поиске в глубину последовательность вершин, определяющих путь между u и v уже содержится в памяти алгоритма на момент обнаружения вершины v. Однако этот путь может быть не кратчайшим.

Связность графов

Пусть G – неограф. Две вершины a и b называются связанными, если существует путь S с начальной вершиной a и конечной вершиной b.

Граф называется связным, если связана любая пара его вершин.

Для всякого графа существует такое разбиение множества его вершин на попарно непересекающиеся множества Vi

$$V = \bigcup_{i=1}^{k} V_i$$

что вершины в каждом Vi связаны, а вершины из различных Vi не связаны. В этом случае граф состоит из k непересекающихся связных подграфов – k компонентов связности.

Связность графов

Теорема 2: Если в конечном неориентированном простом графе G ровно две вершины a и b имеют нечетную степень, то они связаны.

По теореме 1 каждый конечный граф имеет четное число вершин нечетной степени, это же верно и для компонент.

Теорема 3: Если неориентированный простой граф G имеет n вершины и k связных компонент, то максимальное число рёбер в G

$$N(n, k) = \tfrac{1}{2} (n - k) (n - k + 1)$$

Доказательство опустим.

Теорема 4 (следствие): Простой неограф с n вершинами и числом рёбер, большим, чем $N(n, 2) = \tfrac{1}{2} (n - 2) (n - 1)$, связан.

Связность графов

Теорема 2: Если в конечном неориентированном простом графе G ровно две вершины a и b имеют нечетную степень, то они связаны.

По теореме 1 каждый конечный граф имеет четное число вершин нечетной степени, это же верно и для компонент.

Теорема 3: Если неориентированный простой граф G имеет n вершины и k связных компонент, то максимальное число рёбер в G

$$N(n, k) = \frac{1}{2}(n - k)(n - k + 1)$$

Доказательство опустим.

Теорема 4 (следствие): Простой неограф с n вершинами и числом рёбер, большим, чем $N(n, 2) = \frac{1}{2}(n - 2)(n - 1)$, связан.

Будет ли связан граф G при $n = 5$ и $m = 5$? $m = 6$? $m = 7$?

Деревья

Связный неограф называется деревом, если он не имеет циклов. Кроме того, дерево не имеет петель и кратных ребер.

Лес – упорядоченное множество упорядоченных деревьев

Дерево – связный граф, содержащий только один путь между двумя вершинами. Может быть укорененным и неукорененным.

Двоичное дерево – ориентированное дерево, в котором исходящие степени вершин (число исходящих рёбер) не превосходят 2.

Длина ребра – число, соотнесенное с каждым ребром и обозначающее, в каком-то смысле, расстояние между двумя вершинами, соединенными этим ребром.

Длина пути – сумма всех длин ребер, составляющих путь.

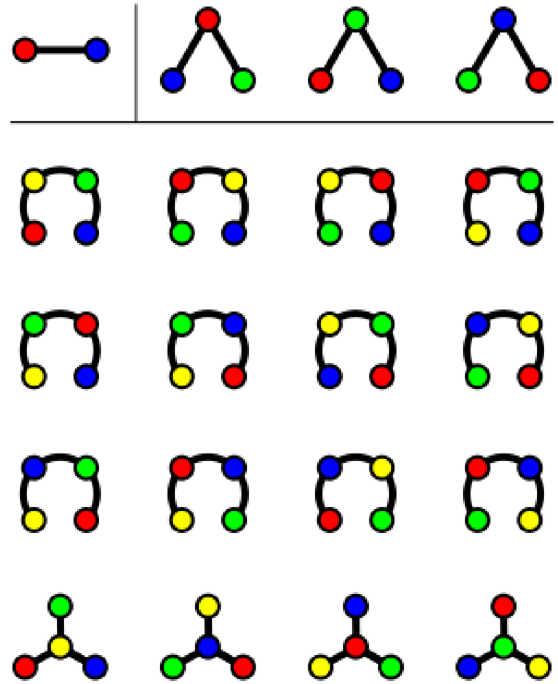
Теорема 5: В дереве любые две вершины связаны единственным простым путем.

Док-во: Если бы путей было 2, то был бы цикл.

Деревья: свойства

Теорема 7: Любое дерево с n вершинами содержит (n – 1) ребро

Док-во: По индукции по числу вершин. Для n = 1 очевидно. Пусть n > 1. Тогда в дереве существует концевая вершина v, удаляя которую вместе с инцидентным ребром (u, v), получим дерево с (n – 1) вершиной, которое, по предположению, имеет (n – 2) ребра. Значит, в исходном дереве было (n – 2 + 1) = (n – 1) ребро.

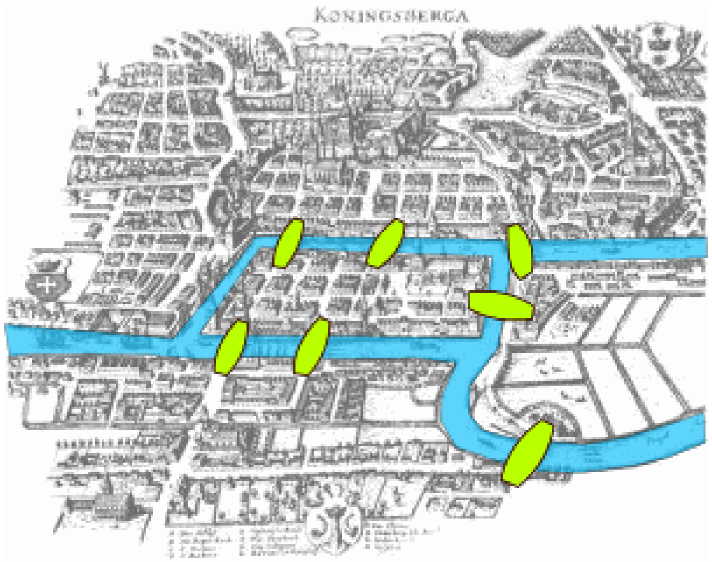


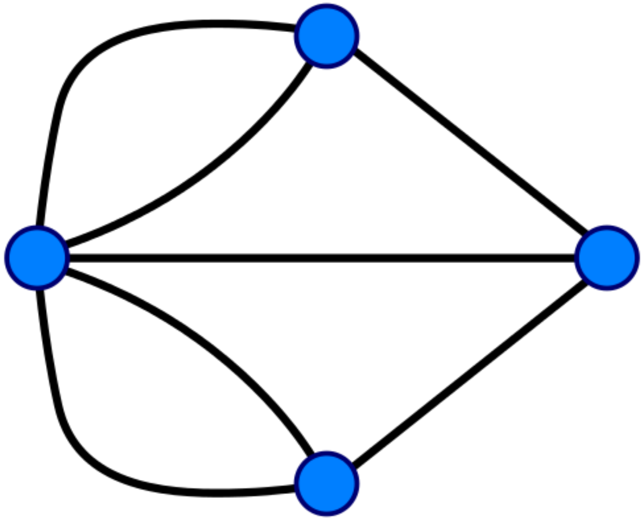
Теорема 8: Число различных деревьев, которые можно построить на n нумерованных вершинах, равно n^(n-2) (**Теорема Кэли**).

Эйлеровы пути и циклы

Эйлеровым путем в графе G называется произвольный путь, проходящий через каждое ребро графа в точности один раз. Если начальная и конечная вершины совпадают, то путь называется эйлеровым циклом.

Теорема 11: Эйлеров путь в простом неографе существует тогда и только тогда, когда граф связный и содержит не более 2 вершин нечетной степени.

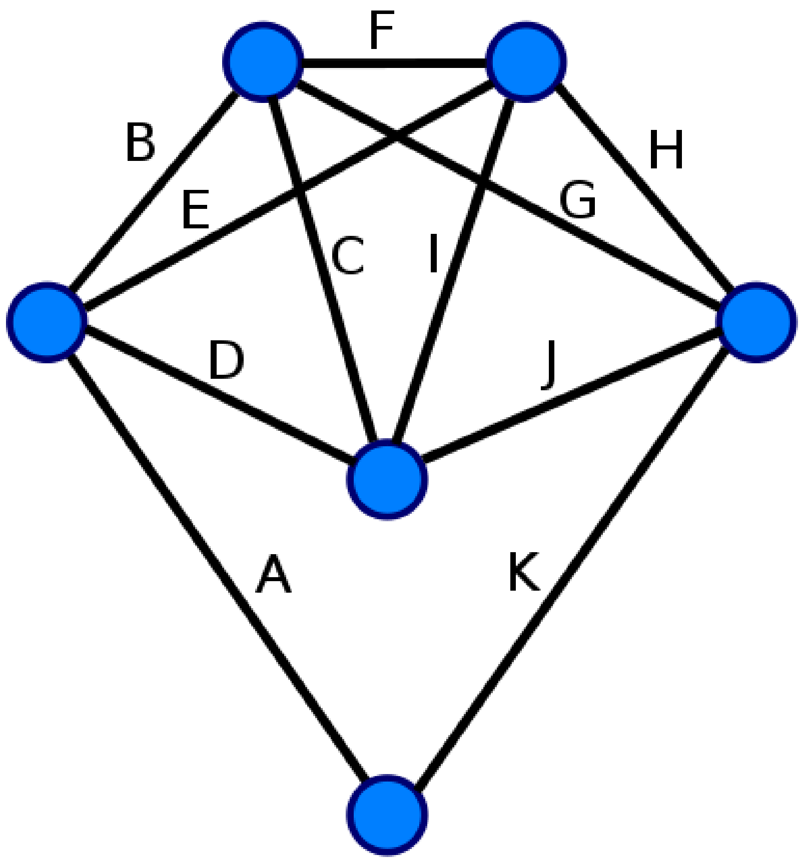




Эйлеровы пути и циклы

Теорема 12: Эйлеров цикл в простом неографе существует тогда и только тогда, когда граф связный и степени всех его вершин четные.

Доказательство традиционно опустим ☺

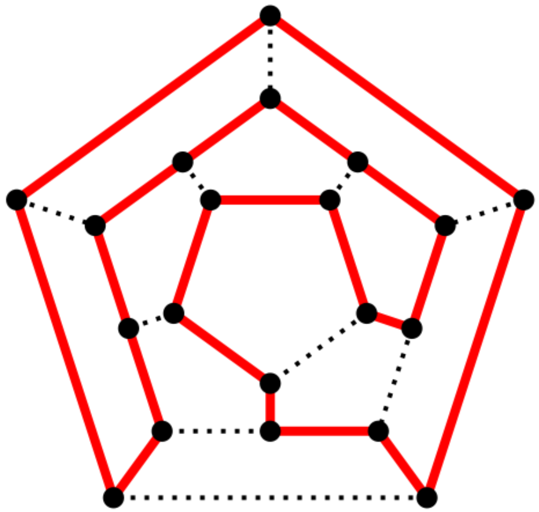


Гамильтоновы пути и циклы

Простой путь (цикл) называется гамильтоновым путем (циклом), если он проходит через каждую вершину графа ровно один раз.

В отличие от эйлеровых путей, неизвестно ни одного простого необходимого и достаточного условия для существования гамильтоновых путей и циклов.

Неизвестен и алгоритм, проверяющий существование такого пути в произвольном графе за полиномиальное время от числа вершин n (NP-полная задача).



У. Гамильтон, «Путешествие вокруг света» (1859)

Гамильтоновы пути и циклы

Теорема 17: Если в графе G с n вершинами для любой пары вершин ai и aj имеет место соотношение ρ(ai) + ρ(aj) >= (n – 1), то граф G имеет гамильтонов путь.

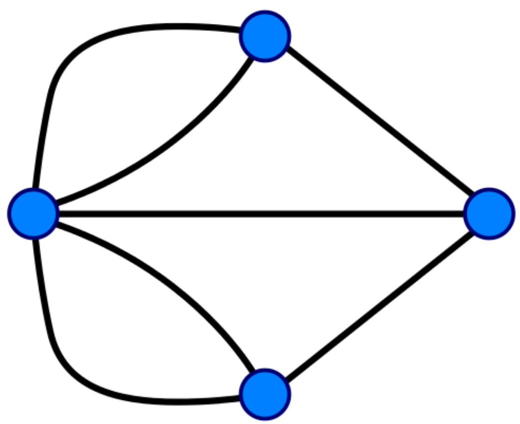
Если ρ(ai) + ρ(aj) >= n, то имеет гамильтонов цикл.

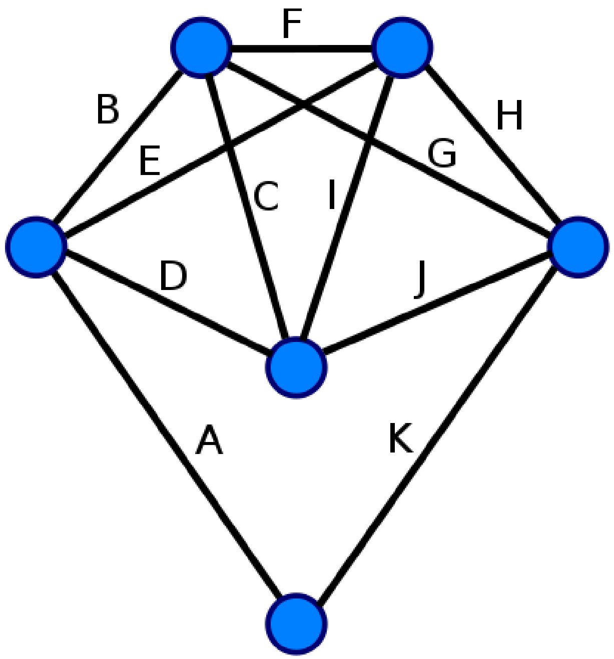
Условие Поша: Пусть граф G имеет p > 2 вершин. Если для всякого n, 0 < n < (p-1)/2, число вершин со степенями меньшими или равными n меньше, чем n, и для нечетного p число вершин со степенью (p-1)/2 не превосходит (p-1)/2, то G — гамильтонов граф. Это достаточное условие не является необходимым.

Гамильтоновы пути и циклы

Простой путь (цикл) называется гамильтоновым путем (циклом), если он проходит через каждую вершину графа ровно один раз.

Теорема 17: Если в графе G с n вершинами для любой пары вершин ai и aj имеет место соотношение ρ(ai) + ρ(aj) >= (n – 1), то граф G имеет гамильтонов путь. Если ρ(ai) + ρ(aj) >= n, то имеет гамильтонов цикл.





Задача коммивояжера
(Travelling salesman problem, TSP)

поиск самого выгодного маршрута, проходящего через указанные города хотя бы по одному разу с последующим возвратом в исходный город

~ (n – 1)! маршрутов для n городов

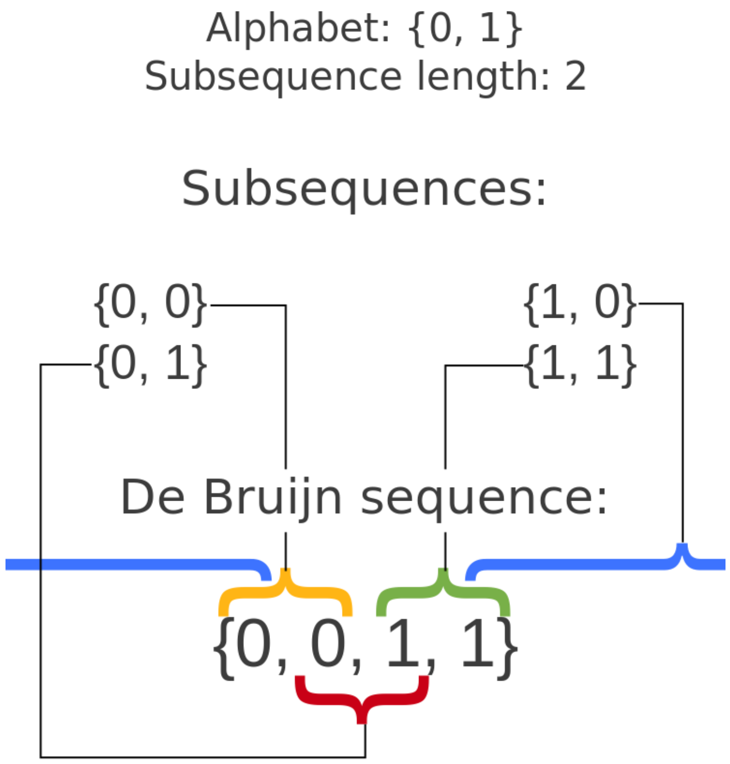
NP-полная, трансвычислительная задача

Оптимальный маршрут коммивояжёра через 15 крупнейших городов Германии. Указанный маршрут является самым коротким из всех возможных 43 589 145 600 вариантов.

Последовательности де Брёйна

Последовательностью де Брёйна порядка n на алфавите A размера k называется такая циклическая последовательность B(k, n), в которой единожды встречаются все возможные последовательности длины n составленные из букв алфавита A.

Длина такой последовательности k^n, что также является число различных строг длины n из алфавита A.



Число различных последовательностей B(k, n) =

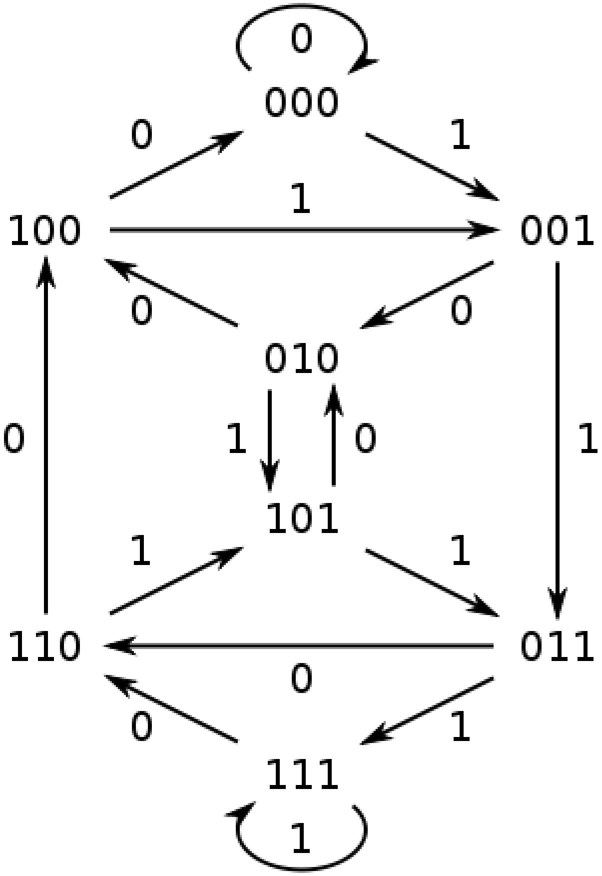
$$\frac{(k!)^{k^{n-1}}}{k^n}$$

Последовательности де Брёйна

Последовательность де Брёйна может быть получена посредством нахождения гамильтонова пути в n-мерном графе де Брёйна или, что то же, эйлерова цикла в (n – 1)-мерном графе де Брёйна.

Рассмотрим построение последовательности B(2, 4) длиной 2^4 = 16 с использованием эйлерова цикла на n – 1 = 3-мерном графе.

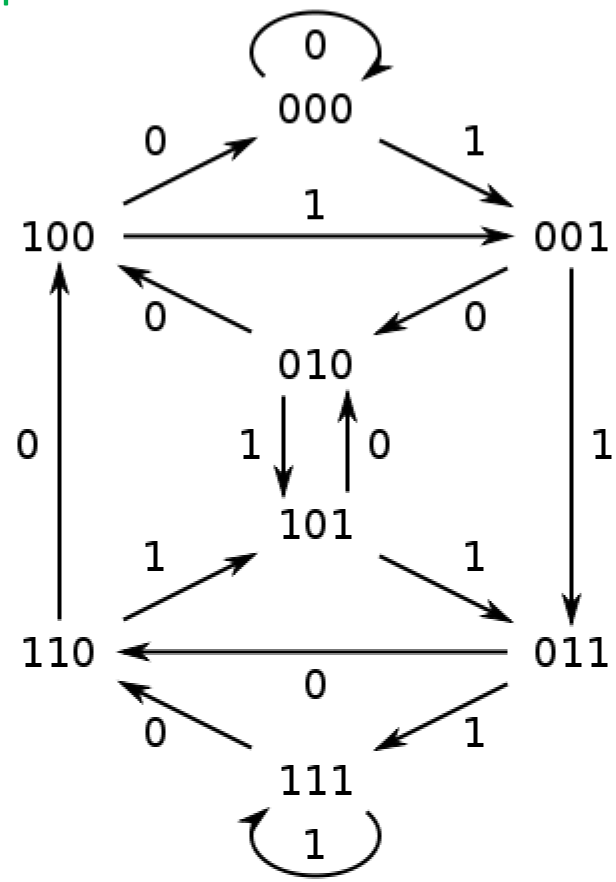
Существует ли он?



Последовательности де Брёйна

Последовательность де Брёйна может быть получена посредством нахождения гамильтонова пути в n-мерном графе де Брёйна или, что то же, эйлерова цикла в (n – 1)-мерном графе де Брёйна.

Рассмотрим построение последовательности B(2, 4) длиной 2^4 = 16 с использованием эйлерова цикла на n – 1 = 3-мерном графе.



Существует ли он?

Да! Ибо у всех вершин чётная степень

Например, список вершин может быть таков:

000, 000, 001, 011, 111, 111, 110, 101, 011, **110**, 100, 001, 010, 101, 010, 100, 000.

Тогда B(2, 4) = 0 0 0 0 1 1 1 1 0 **1 1 0** 0 1 0 1 (0 0…)

Последовательности де Брёйна

Зачем они нужны? Можно подбирать пин-код в домофоне (если тот реагирует на 4 последних нажатия) ☺

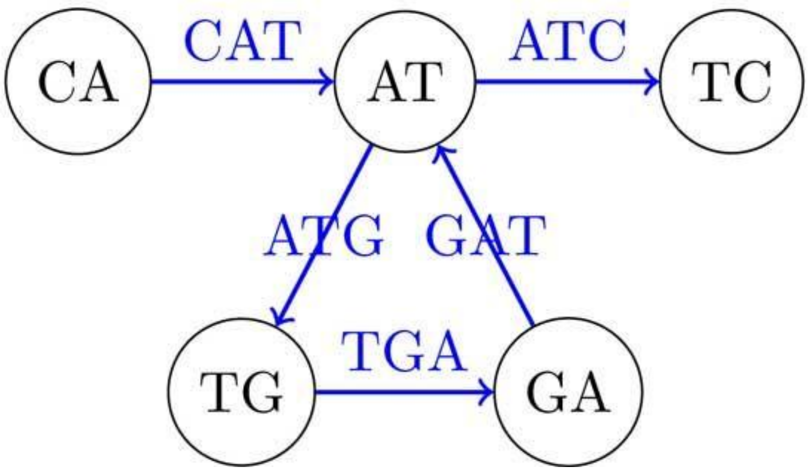
Длина (B (10, 4)) = 10^4 = 10000

Соответственно, нужно максимум 10000 + 3 нажатия для того, чтобы встретилась нужная комбинация цифр.

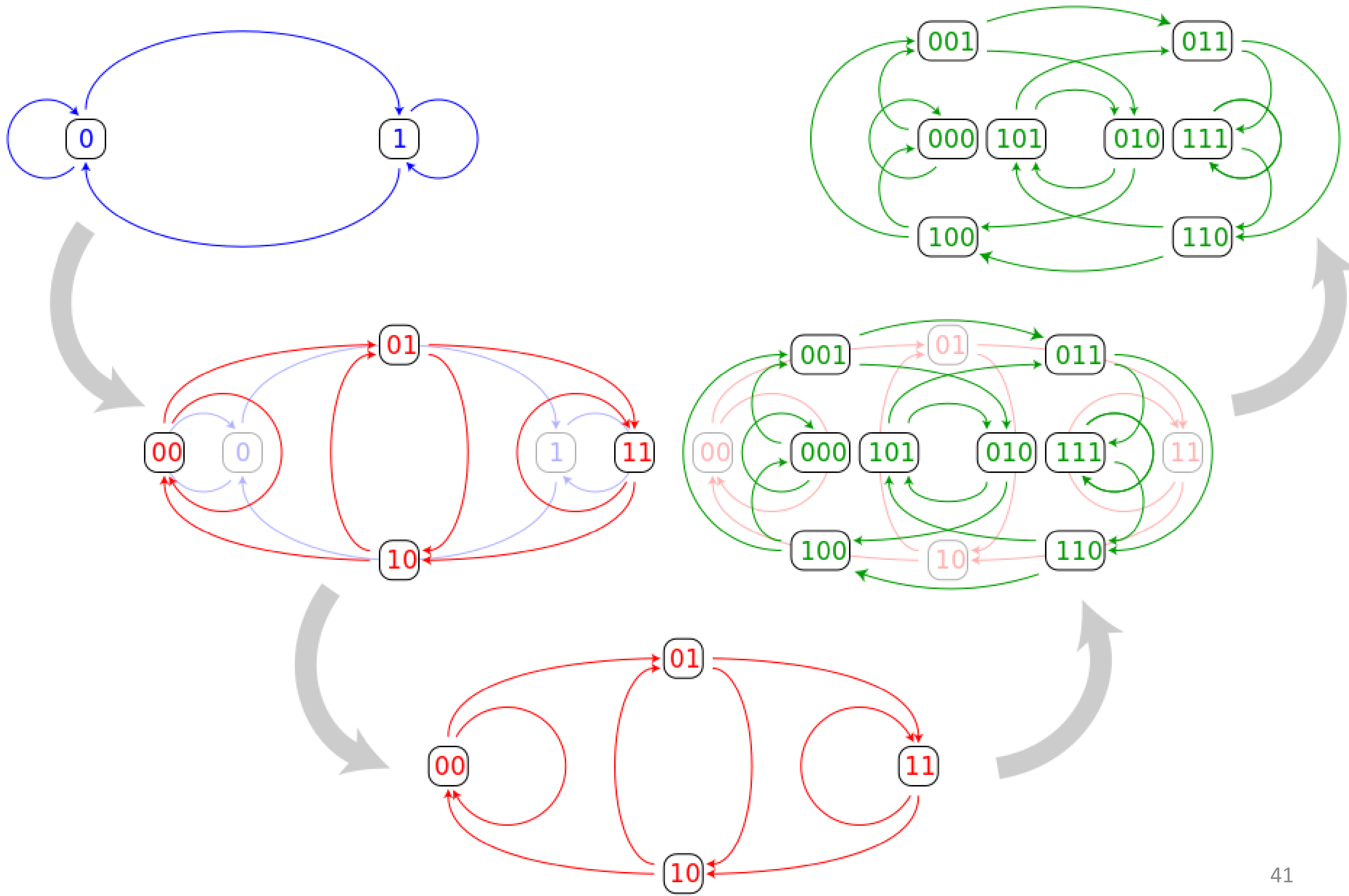
Классический подход дает 10000 * 4 = 40000 нажатий.

Граф де Брёйна — ориентированный граф с k^n вершинами, соответствующими k^n различных наборов длины n с элементами из алфавита A размера k, в котором из вершины (x1, …, x_n) в вершину (y1, …, y_n) ребро ведёт в том и только том случае, когда x_i = y_(i-1); при этом самому ребру можно сопоставить набор длины (x1, x2, …, x_n, y_n) = (x1, y1, …, y_n-1, y_n).

Графы де Брёйна



Графы де Брёйна



Правила Кирхгофа

1) алгебраическая сумма токов ветвей, сходящихся в каждом узле любой цепи, равна нулю.

2) алгебраическая сумма напряжений на резистивных элементах замкнутого контура равна алгебраической сумме ЭДС, входящих в этот контур