Эйлеровы и гамильтоновы циклы

Владимир Подольский

Факультет компьютерных наук, Высшая Школа Экономики

Сборка генома

Сборка генома

Эйлеровы циклы

Гамильтоновы циклы

Сборка генома

Сборка генома

- Геном — это строка из символов A, C, G и T

Сборка генома

- Геном — это строка из символов A, C, G и T
- Как расшифровывают геном?

Сборка генома

- Геном — это строка из символов A, C, G и T
- Как расшифровывают геном?
- Сначала секвенирование — читаем фрагменты генома

Сборка генома

- Геном — это строка из символов A, C, G и T
- Как расшифровывают геном?
- Сначала секвенирование — читаем фрагменты генома
- Но секвенирование умеет находить только короткие фрагменты генома

Сборка генома

- Геном — это строка из символов A, C, G и T
- Как расшифровывают геном?
- Сначала секвенирование — читаем фрагменты генома
- Но секвенирование умеет находить только короткие фрагменты генома
- Так что затем нужна сборка: из коротких фрагментов собрать геном

Сборка генома

Найдите строку, в которой все подстроки длины 3 — это AGC, ATC, CAG, CAT, CCA, GCA, TCA, TCC.

Сборка генома

Найдите строку, в которой все подстроки длины 3 — это AGC, ATC, CAG, CAT, CCA, GCA, TCA, TCC.

Оказывается в этой задаче полезны циклы в графах

Сборка генома

Сборка генома

Эйлеровы циклы

Гамильтоновы циклы

Сборка генома

Эйлеровы циклы

Определение

Эйлеровы циклы проходят по каждому ребру ровно один раз

Эйлеровы циклы

Определение

Эйлеровы циклы проходят по каждому ребру ровно один раз

- Определение годится и для ориентированных, и для неориентированных графов

Эйлеровы циклы

Определение

Эйлеровы циклы проходят по каждому ребру ровно один раз

- Определение годится и для ориентированных, и для неориентированных графов
- В цикле начальная и конечная вершины должны совпадать

Эйлеровы циклы

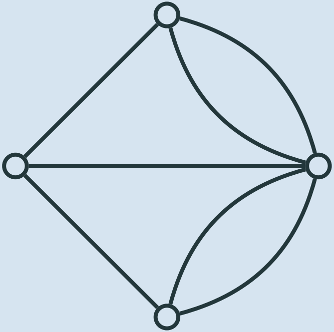
Определение

Эйлеровы циклы проходят по каждому ребру ровно один раз

- Определение годится и для ориентированных, и для неориентированных графов
- В цикле начальная и конечная вершины должны совпадать
- Можно также рассматривать эйлеровы пути, в них начальная и конечная вершины совпадать не обязаны

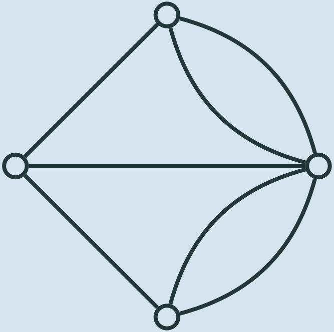
Примеры

Граф без эйлерова цикла

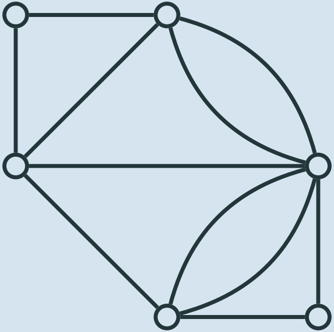


Примеры

Граф без эйлерова цикла

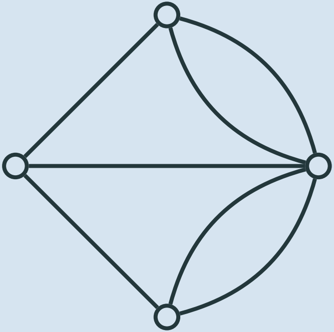


Граф с эйлеровым циклом

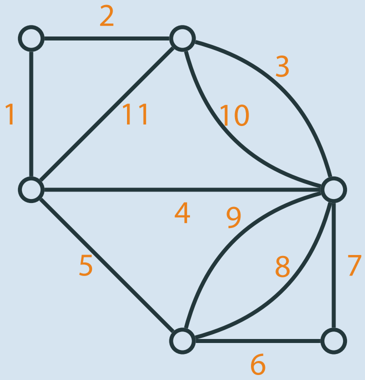


Примеры

Граф без эйлерова цикла



Граф с эйлеровым циклом



Критерий

Теорема

Связный *неориентированный* граф содержит эйлеров цикл тогда и только тогда, когда степень каждой вершины четна

Критерий

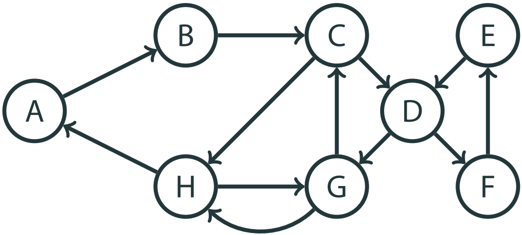
Теорема

Связный *неориентированный* граф содержит эйлеров цикл тогда и только тогда, когда степень каждой вершины четна

Теорема

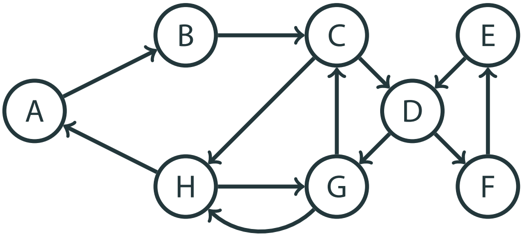
Сильно связный *ориентированный* граф содержит эйлеров цикл тогда и только тогда, когда для каждой вершины ее входная степень равна исходящей степени

Доказательство (ориентированный случай)



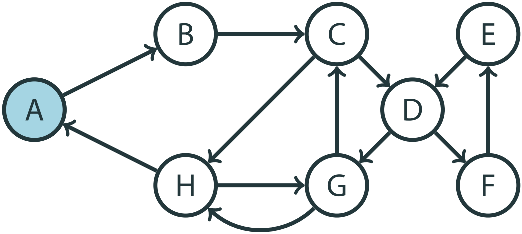
Если в какой-то вершине исходящая сте-
пень не равна входящей, то цикла точно нет

Доказательство (ориентированный случай)



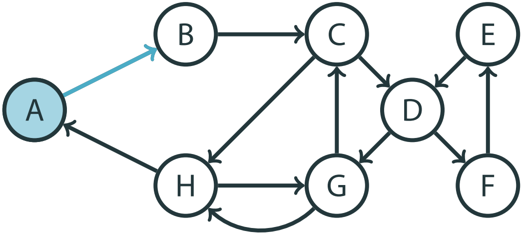
В другую сторону, пусть каждая вершина сбалансирована

Доказательство (ориентированный случай)



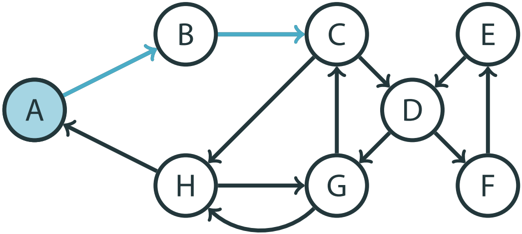
Начнем ходить из какой-то вершины

Доказательство (ориентированный случай)



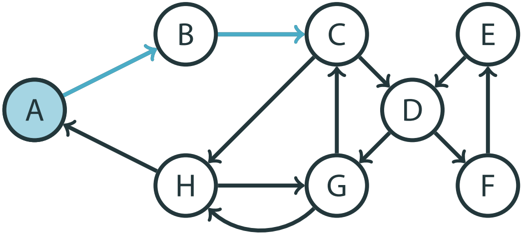
Начнем ходить из какой-то вершины

Доказательство (ориентированный случай)



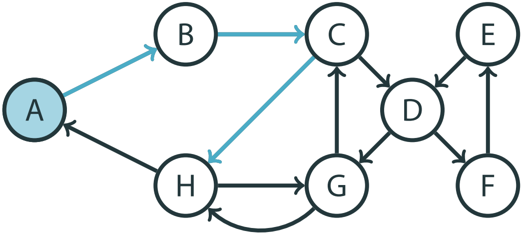
Начнем ходить из какой-то вершины

Доказательство (ориентированный случай)



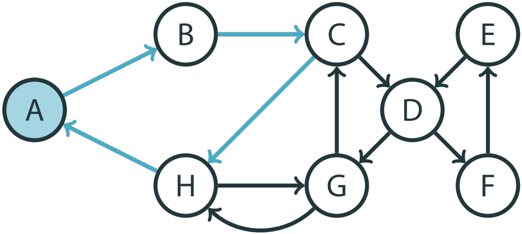
Поскольку каждая вершина сбалансирована, если мы в нее вошли, мы можем выйти

Доказательство (ориентированный случай)



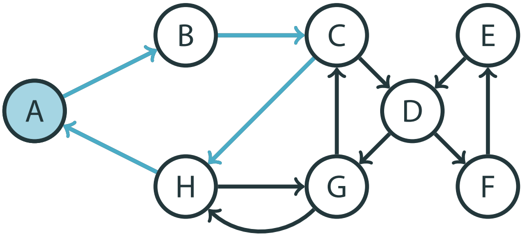
Рано или поздно мы вернемся в изначальную вершину

Доказательство (ориентированный случай)



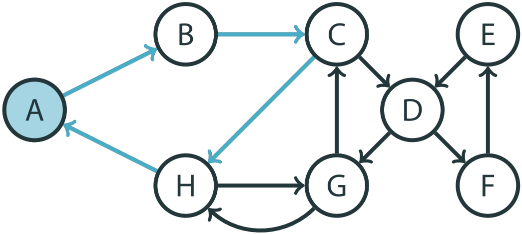
Рано или поздно мы вернем-
ся в изначальную вершину

Доказательство (ориентированный случай)



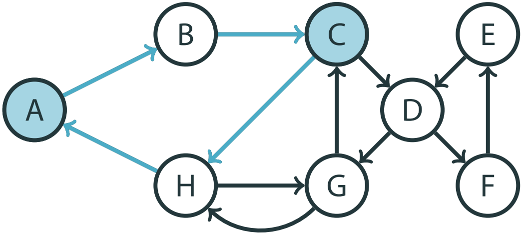
Но что делать дальше? Мы не прошли по всем ребрам и при этом попали в тупик

Доказательство (ориентированный случай)



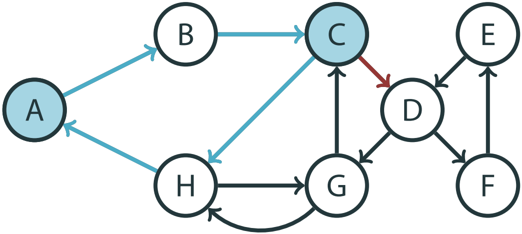
Поскольку граф сильно связный, есть непройденное ребро, которое выходит из вершины построенного цикла

Доказательство (ориентированный случай)



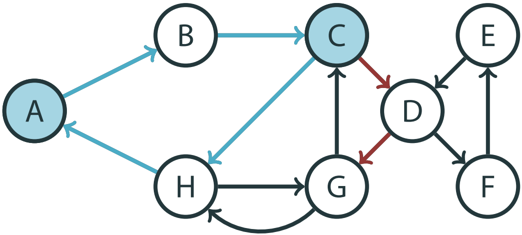
Запускаем обход из такой вершины, например, из C

Доказательство (ориентированный случай)



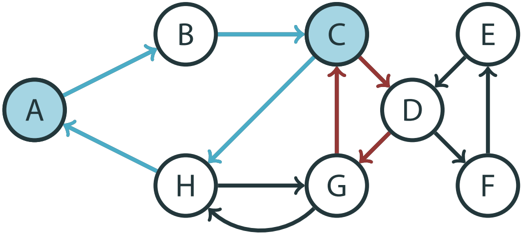
Запускаем обход из такой вершины, например, из C

Доказательство (ориентированный случай)



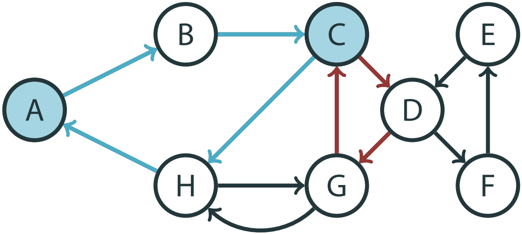
Запускаем обход из такой вершины, например, из C

Доказательство (ориентированный случай)



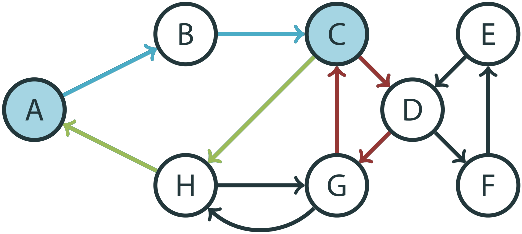
Запускаем обход из такой вершины, например, из C

Доказательство (ориентированный случай)



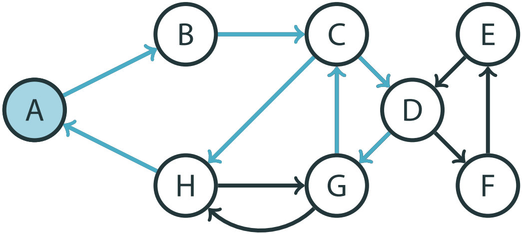
Получили второй цикл

Доказательство (ориентированный случай)



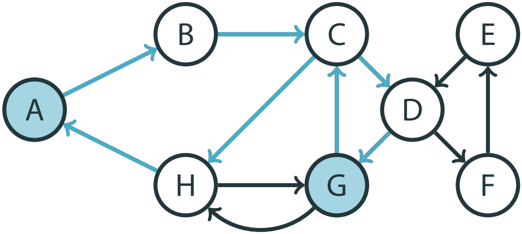
Объединим эти два цикла: сначала идем по первому, потом обходим по второму, затем заканчиваем с первым

Доказательство (ориентированный случай)



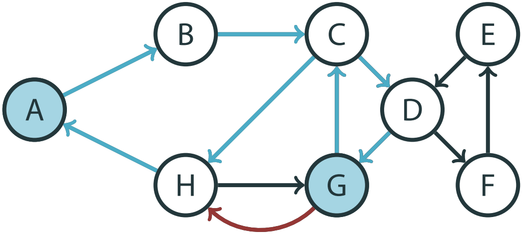
Мы получили больший цикл!

Доказательство (ориентированный случай)



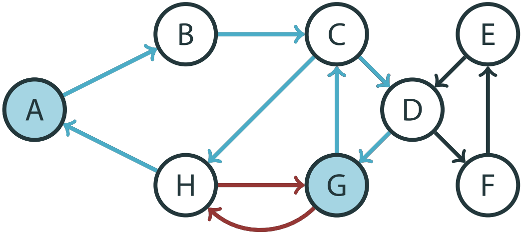
Повторим процесс с G

Доказательство (ориентированный случай)



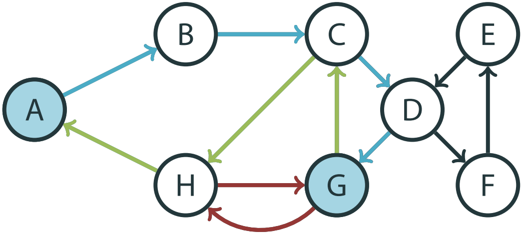
Повторим процесс с G

Доказательство (ориентированный случай)



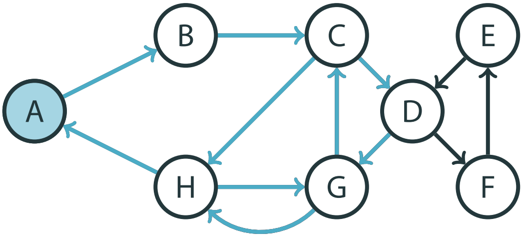
Повторим процесс с G

Доказательство (ориентированный случай)



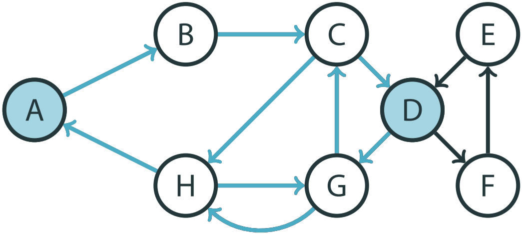
Снова объединяем циклы

Доказательство (ориентированный случай)



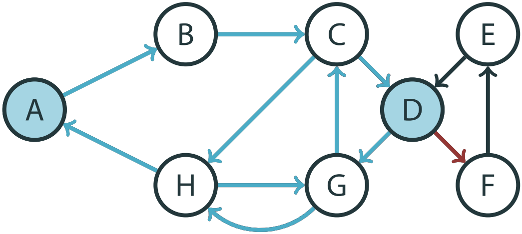
Снова объединяем циклы

Доказательство (ориентированный случай)



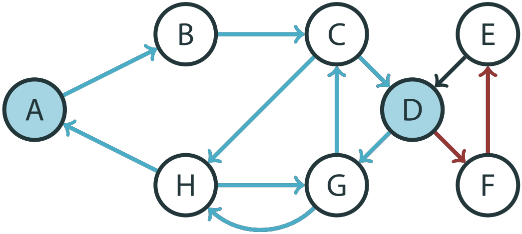
Повторим процесс из D

Доказательство (ориентированный случай)



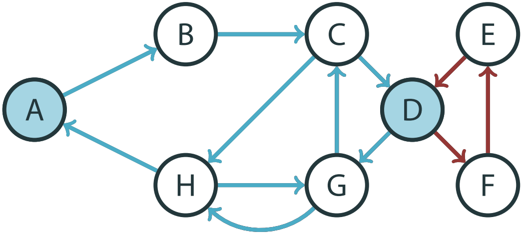
Повторим процесс из D

Доказательство (ориентированный случай)



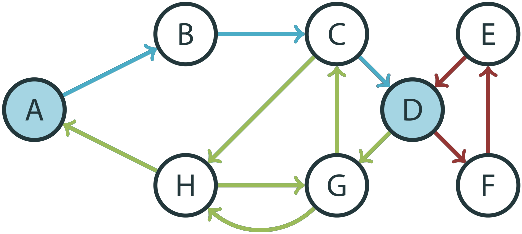
Повторим процесс из D

Доказательство (ориентированный случай)



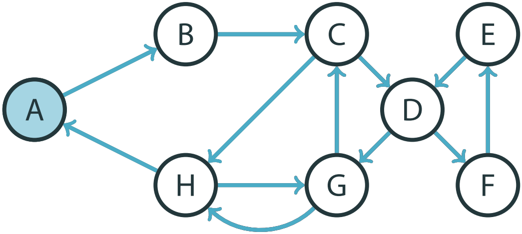
Повторим процесс из D

Доказательство (ориентированный случай)



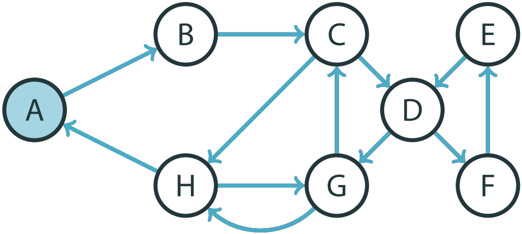
Снова объединяем циклы

Доказательство (ориентированный случай)



Снова объединяем циклы

Доказательство (ориентированный случай)

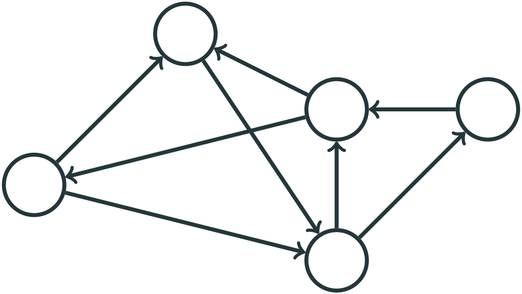


Эйлеров цикл построен

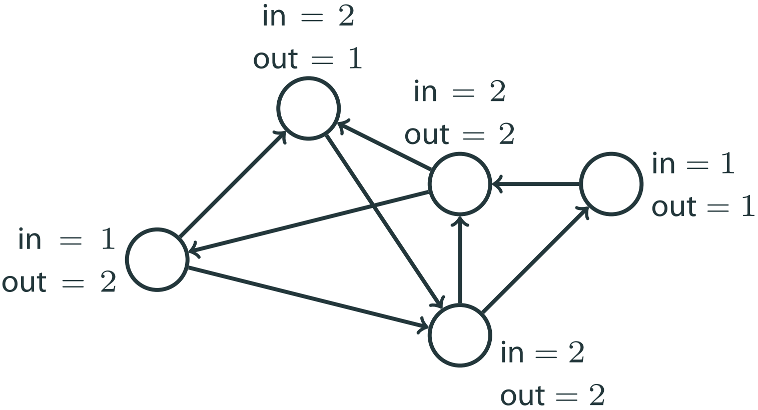
Эйлеровы пути

- Для эйлеровых путей есть аналогичный критерий
- Графу разрешается содержать две несбалансированные вершины: начальную и конечную точку в пути
- Добавляя ребро между этими вершинами мы получим граф с эйлеровым циклом

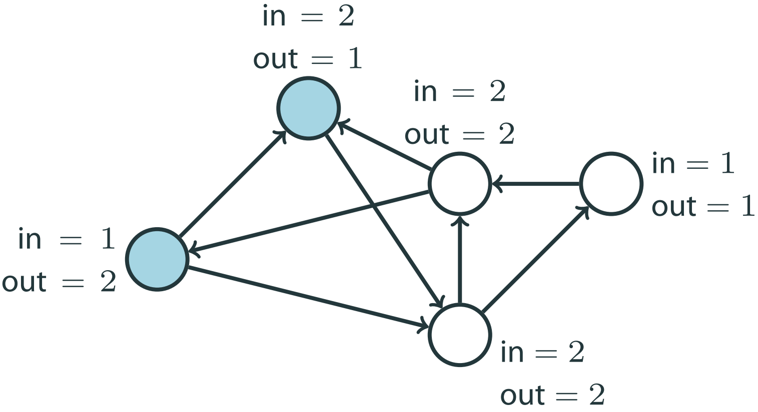
От пути к циклу



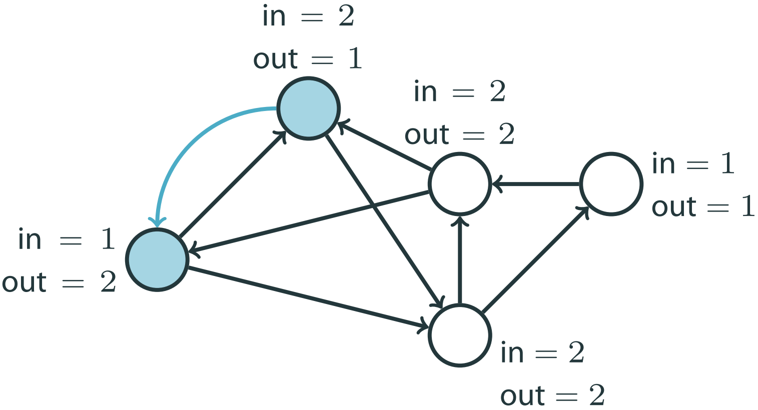
От пути к циклу



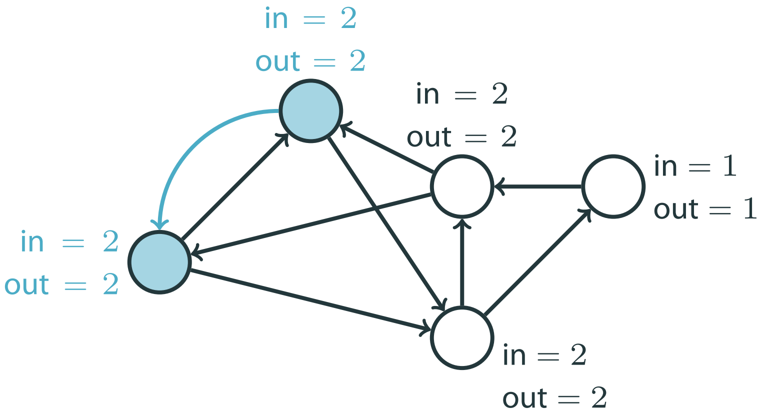
От пути к циклу



От пути к циклу



От пути к циклу



Сборка генома

Сборка генома

Эйлеровы циклы

Гамильтоновы циклы

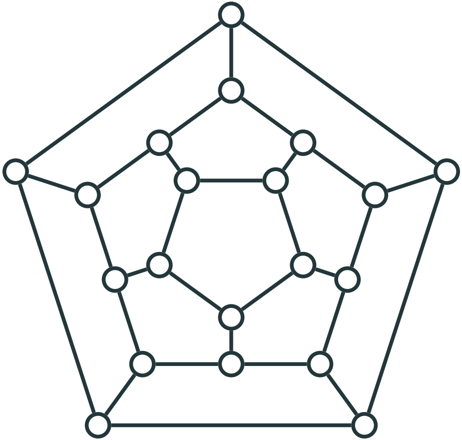
Сборка генома

Гамильтоновы циклы

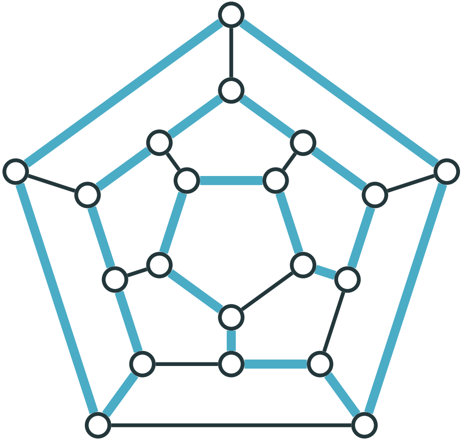
Определение

Гамильтонов цикл — это цикл в графе, проходящий по всем вершинам ровно один раз.

Пример



Пример



Простой критерий?

- Не известно простого способа проверять, есть ли в графе гамильтонов цикл

Простой критерий?

- Не известно простого способа проверять, есть ли в графе гамильтонов цикл
- Мы не знаем алгоритма, работающего за полиномиальное время

Простой критерий?

- Не известно простого способа проверять, есть ли в графе гамильтонов цикл
- Мы не знаем алгоритма, работающего за полиномиальное время
- Задача о гамильтоновом цикле относится к NP-полным задачам, так что вопрос о ее полиномиальный разрешимости — центральный открытый вопрос теоретической информатики

Сборка генома

Сборка генома

Эйлеровы циклы

Гамильтоновы циклы

Сборка генома

Сборка генома

Найдите строку, в которой все подстроки длины — это
AGC, ATC, CAG, CAT, CCA, GCA, TCA, TCC.

Сборка генома

Найдите строку, в которой все подстроки длины — это
AGC, ATC, CAG, CAT, CCA, GCA, TCA, TCC.

Оказывается в этой задаче полезны циклы в графах

Все подстроки длины 3

TCCAGCATCA
TCC
 CCA
  CAG
   AGC
    GCA
     CAT
      ATC
       TCA

Все подстроки длины 3

```
TCCAGCATCA
TCC
 CCA
  CAG
   AGC
    GCA
     CAT
      ATC
       TCA
```

Каждые две соседние строки длины 3 пересекаются по общей части длины 2

Ищем перестановку

- Наша цель — найти строку, все подстроки которой — это AGC, ATC, CAG, CAT, CCA, GCA, TCA, TCC

Ищем перестановку

- Наша цель — найти строку, все подстроки которой — это AGC, ATC, CAG, CAT, CCA, GCA, TCA, TCC
- Другими словами, нам нужно найти перестановку этих строк длины 3, такую что каждые две последовательные строки пересекаются по длине 2

Ищем перестановку

- Например, пусть мы ищем последовательность, в которой все подпоследовательности, это ACG, TAC и CGC

Ищем перестановку

- Например, пусть мы ищем последовательность, в которой все подпоследовательности, это ACG, TAC и CGC

Ищем перестановку

- Например, пусть мы ищем последовательность, в которой все подпоследовательности, это ACG, TAC и CGC

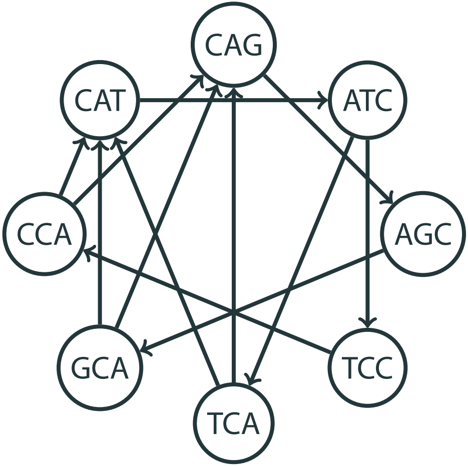
Ищем перестановку

- Например, пусть мы ищем последовательность, в которой все подпоследовательности, это ACG, TAC и CGC
- Переставим их так: TAC, ACG, CGC

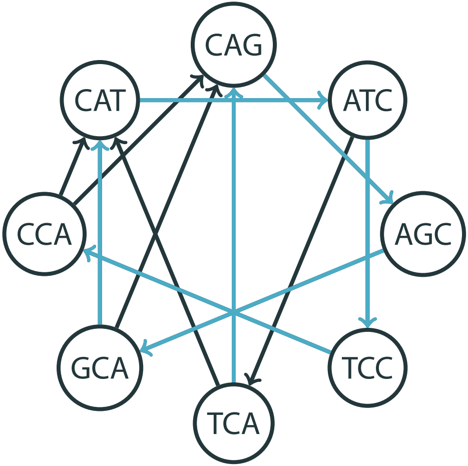
Ищем перестановку

- Например, пусть мы ищем последовательность, в которой все подпоследовательности, это ACG, TAC и CGC
- Переставим их так: TAC, ACG, CGC
- Получаем слово TACGC

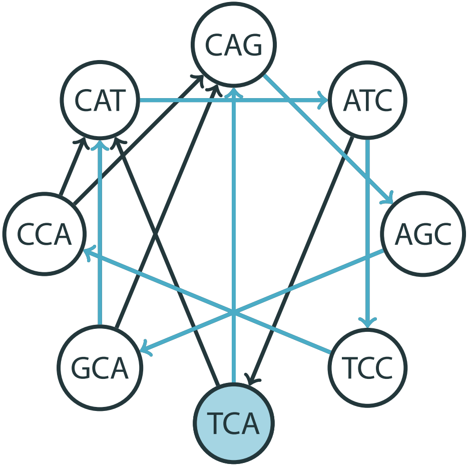
Граф пересечений



Граф пересечений

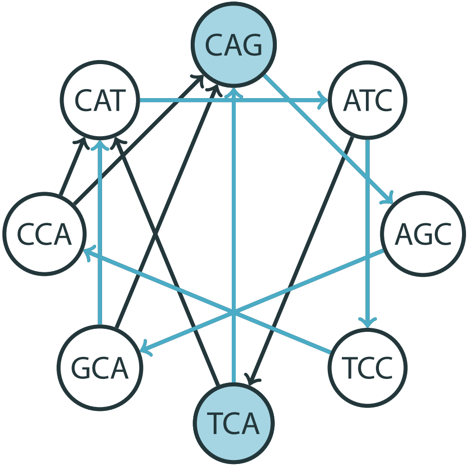


Граф пересечений



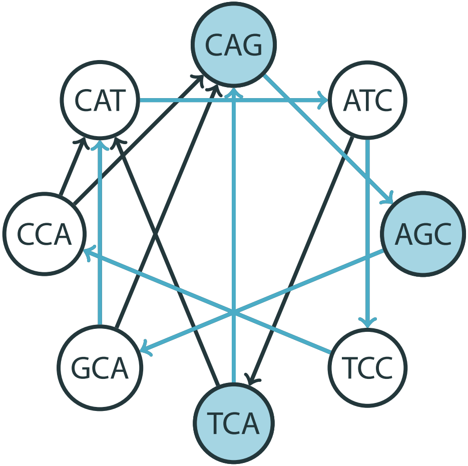
TCA

Граф пересечений



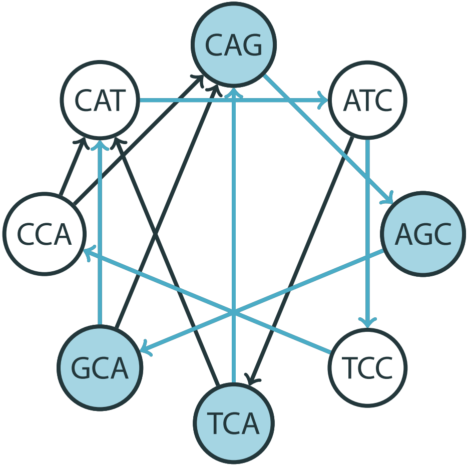
TCAG

Граф пересечений



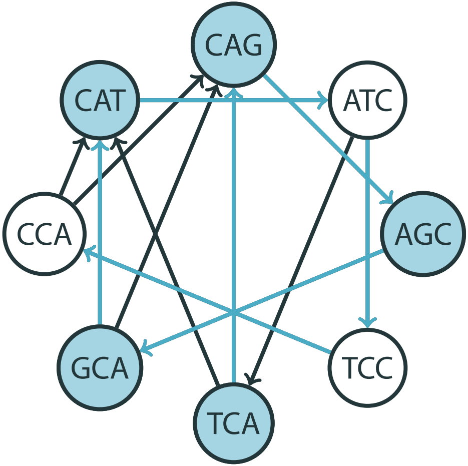
TCAGC

Граф пересечений



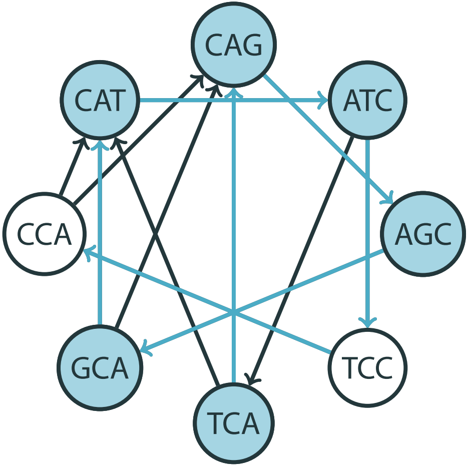
TCAGCA

Граф пересечений



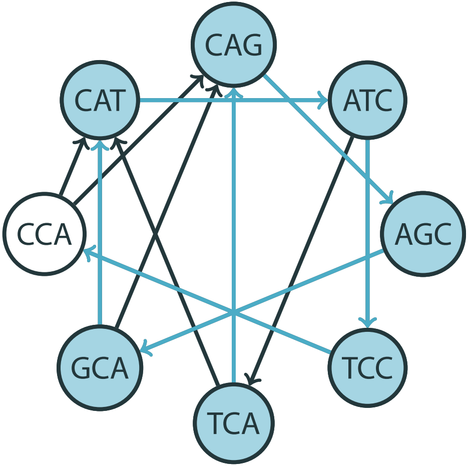
TCAGCAT

Граф пересечений



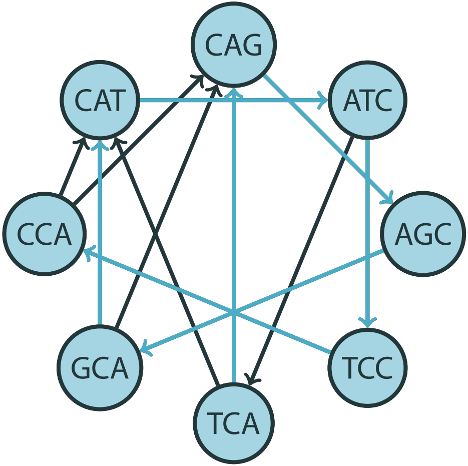
TCAGCATC

Граф пересечений



TCAGCATCC

Граф пересечений



TCAGCATCCA

Справились ли мы с задачей?

- Мы свели задачу к поиску гамильтонова пути!

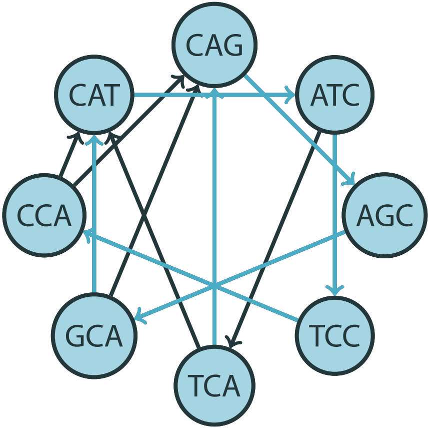
Справились ли мы с задачей?

- Мы свели задачу к поиску гамильтонова пути!
- Но у нас нет эффективного алгоритма для поиска гамильтонова пути

Справились ли мы с задачей?

- Мы свели задачу к поиску гамильтонова пути!
- Но у нас нет эффективного алгоритма для поиска гамильтонова пути
- Так что этот подход бесполезен для последовательностей большого размера

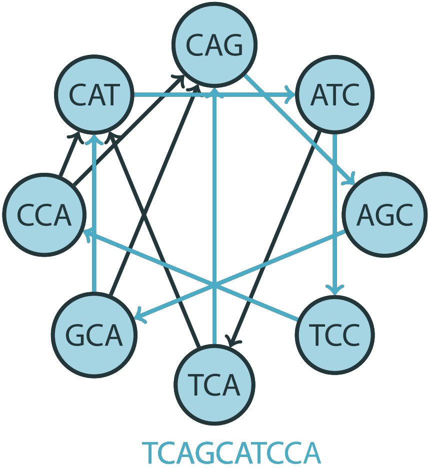
Граф пересечений



TCAGCATCCA

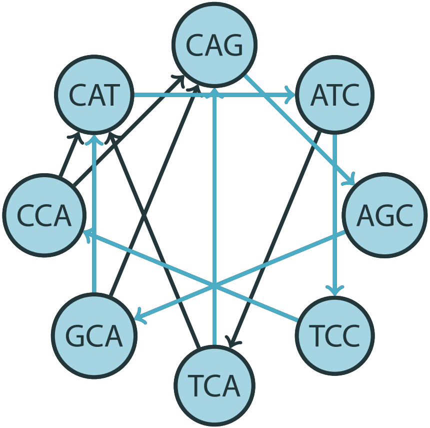
- Чему соответствуют ребра на картинке?

Граф пересечений



- Чему соответствуют ребра на картинке?
- Последовательностям длины 4!

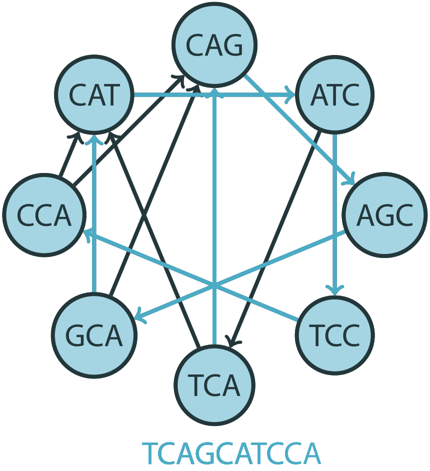
Граф пересечений



TCAGCATCCA

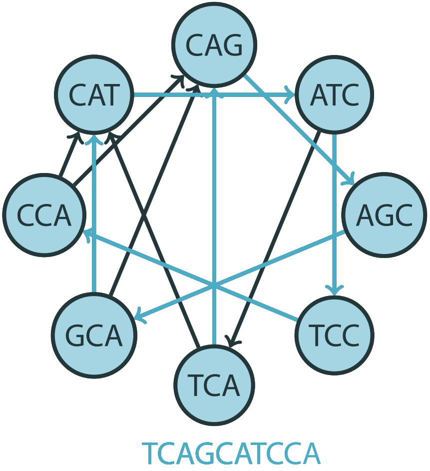
- Синие ребра — все последовательности длины 4 в нашем слове

Граф пересечений



- Сейчас:
  вершины — последовательности длины 3,
  ребра — последовательности длины 4

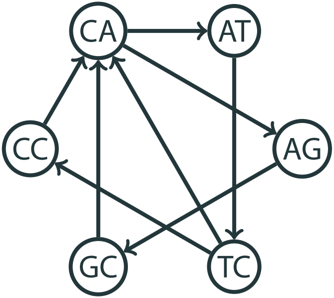
Граф пересечений



- Идея:
  вершины — последовательности длины 2,
  ребра — последовательности длины 3

Граф пересечений

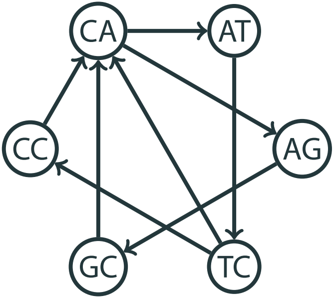
AGC, ATC, CAG, CAT, CCA, GCA, TCA, TCC



Проводим только ребра, соответствующие данным нам последовательностям длины 3

Граф пересечений

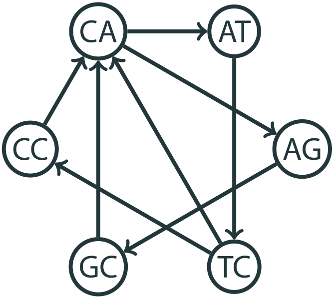
AGC, ATC, CAG, CAT, CCA, GCA, TCA, TCC



Теперь нужно найти перестановку ребер

Граф пересечений

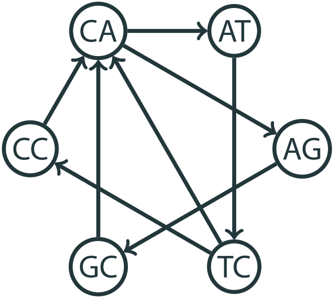
AGC, ATC, CAG, CAT, CCA, GCA, TCA, TCC



А это эйлеров путь!

Граф пересечений

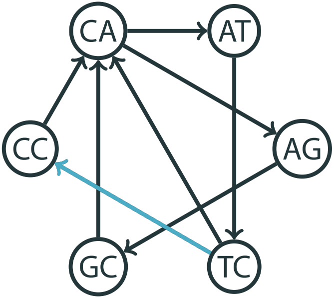
AGC, ATC, CAG, CAT, CCA, GCA, TCA, TCC



Мы знаем как такой путь построить

Граф пересечений

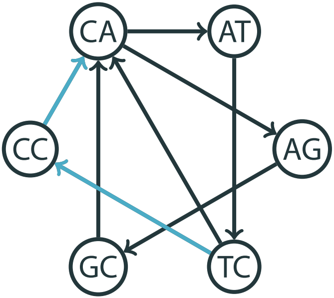
AGC, ATC, CAG, CAT, CCA, GCA, TCA, TCC



TCC

Граф пересечений

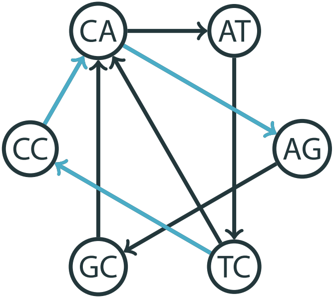
AGC, ATC, CAG, CAT, CCA, GCA, TCA, TCC



TCCA

Граф пересечений

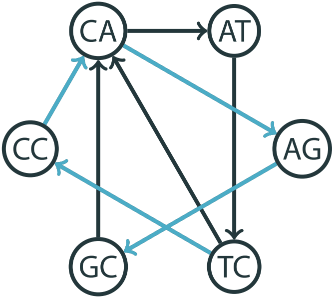
AGC, ATC, CAG, CAT, CCA, GCA, TCA, TCC



TCCAG

Граф пересечений

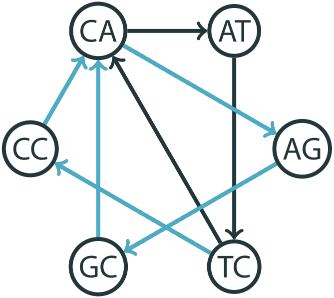
AGC, ATC, CAG, CAT, CCA, GCA, TCA, TCC



TCCAGC

Граф пересечений

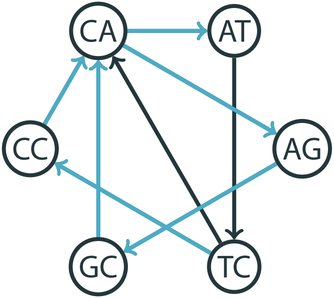
AGC, ATC, CAG, CAT, CCA, GCA, TCA, TCC



TCCAGCA

Граф пересечений

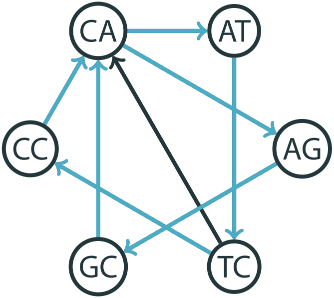
AGC, ATC, CAG, CAT, CCA, GCA, TCA, TCC



TCCAGCAT

Граф пересечений

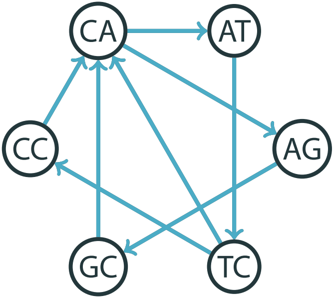
AGC, ATC, CAG, CAT, CCA, GCA, TCA, TCC



TCCAGCATC

Граф пересечений

AGC, ATC, CAG, CAT, CCA, GCA, TCA, TCC



TCCAGCATCA

Что мы узнали

- Эйлеровы пути проходят по всем ребрам ровно один раз

Что мы узнали

- Эйлеровы пути проходят по всем ребрам ровно один раз
- Гамильтоновы пути проходят по всем вершинам ровно один раз

Что мы узнали

- Эйлеровы пути проходят по всем ребрам ровно один раз
- Гамильтоновы пути проходят по всем вершинам ровно один раз
- Выглядят похоже, но принципиально различаются с точки зрения алгоритмической сложности

Что мы узнали

- Эйлеровы пути проходят по всем ребрам ровно один раз
- Гамильтоновы пути проходят по всем вершинам ровно один раз
- Выглядят похоже, но принципиально различаются с точки зрения алгоритмической сложности
- Сборка генома: важно найти правильную модель!

Что мы узнали

- Эйлеровы пути проходят по всем ребрам ровно один раз
- Гамильтоновы пути проходят по всем вершинам ровно один раз
- Выглядят похоже, но принципиально различаются с точки зрения алгоритмической сложности
- Сборка генома: важно найти правильную модель!
- Описанная идея используется в современных сборщиках генома