

Kombinatorika pro bioinformatiky

Aplikace.

Ondřej Pangrác

6.1.2021

Grafové problémy

Eulerovský tah. Projít všechny hrany grafu, každou právě jednou.

Věta

Graf bez izolovaných vrcholů má eulerovský tah právě tehdy, když je souvislý a všechny jeho vrcholy mají sudý stupeň.

Lze řešit v polynomiálním čase.

Hamiltonovská kružnice. Projít všechny vrcholy grafu, každý právě jednou.

Rozhodnout, zda graf obsahuje ham. kružnici je NP–úplný.

Není znám žádný polynomiální algoritmus.

Sekvenování

Sekvenování pomocí DNA microarrays

- ▶ 1988 – návrh metody

Sekvenování

Sekvenování pomocí DNA microarrays

- ▶ 1988 – návrh metody
- ▶ 1994 – vytvořeno první (komerčně použitelné) DNA microarray (16 000)

Sekvenování

Sekvenování pomocí DNA microarrays

- ▶ 1988 – návrh metody
- ▶ 1994 – vytvořeno první (komerčně použitelné) DNA microarray (16 000)
- ▶ 2002 – DNA microarray velikosti 500 000

Sekvenování

Sekvenování pomocí DNA microarrays

- ▶ 1988 – návrh metody
- ▶ 1994 – vytvořeno první (komerčně použitelné) DNA microarray (16 000)
- ▶ 2002 – DNA microarray velikosti 500 000
- ▶ ... – až několik milionů

Sekvenování

Sekvenování pomocí DNA microarrays

- ▶ 1988 – návrh metody
- ▶ 1994 – vytvořeno první (komerčně použitelné) DNA microarray (16 000)
- ▶ 2002 – DNA microarray velikosti 500 000
- ▶ ... – až několik milionů
- ▶ $4^{10} = 1048576$

Metoda

Pro pevně zvolené k uvažujme všechny řetězce DNA délky k .

Metoda

Pro pevně zvolené k uvažujme všechny řetězce DNA délky k .

Neznámý řetězec je nastříhán na úseky délky k a ty jsou připojeny na ty na DNA microarray.

Metoda

Pro pevně zvolené k uvažujme všechny řetězce DNA délky k .

Neznámý řetězec je nastříhán na úseky délky k a ty jsou připojeny na ty na DNA microarray.

Detekovány všechny podřetězce délky k (tzv. spektrum) a z něho sestaven původní řetězec.

Spektrum

Pro s řetězec DNA a k přirozené číslo definujeme spektrum(s,k) jako množinu všech podřetězců délky k.

Spektrum

Pro s řetězec DNA a k přirozené číslo definujeme $\text{spektrum}(s,k)$ jako množinu všech podřetězců délky k.

Příklad:

$$\text{spektrum}(\text{ATAGCT},3) = \{\text{ATA}, \text{TAG}, \text{AGC}, \text{GCT}\}$$

Spektrum

Pro s řetězec DNA a k přirozené číslo definujeme spektrum(s,k) jako množinu všech podřetězců délky k.

Příklad:

$\text{spektrum}(\text{ATAGCT}, 3) = \{\text{ATA}, \text{TAG}, \text{AGC}, \text{GCT}\}$

Problém:

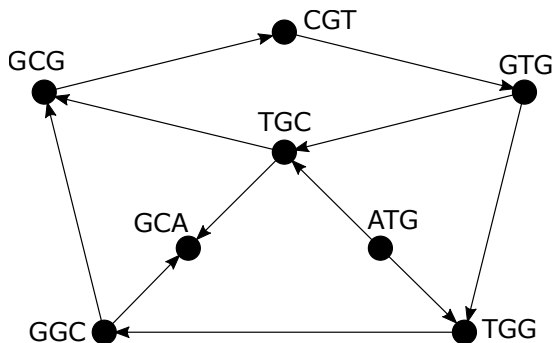
rekonstrukce řetězce DNA z jeho podřetězců délky k

Input: spektrum neznámého řetězce s

Output: řetězec s (s daným spektrem)

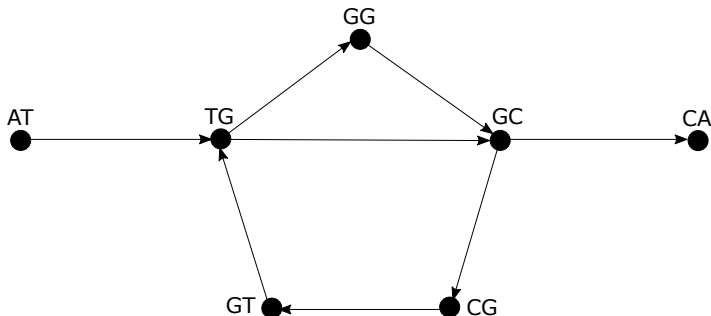
Hamiltonovská kružnice

$\text{spektrum}(s,3) = \{\text{ATG}, \text{CGT}, \text{GCA}, \text{GCG}, \text{GGC}, \text{GTG}, \text{TGC}, \text{TGG}\}$



Eulerovký tah

$$\text{spektrum}(s,3) = \{\text{ATG}, \text{CGT}, \text{GCA}, \text{GCG}, \text{GGC}, \text{GTG}, \text{TGC}, \text{TGG}\}$$



$s_1 = \text{ATGGCGTGCA}$

$s_2 = \text{ATGCGTGGCA}$

Možné komplikace

- ▶ DNA microarrays – velikost a praktická použitelnost

Možné komplikace

- ▶ DNA microarrays – velikost a praktická použitelnost
- ▶ více přípustných řešení – různé řetězce se stejným spektrem (pro dané k)

Možné komplikace

- ▶ DNA microarrays – velikost a praktická použitelnost
- ▶ více přípustných řešení – různé řetězce se stejným spektrem (pro dané k)
- ▶ vícenásobný výskyt k -tic v řetězci