3. Izračunajte zaporedje s, ki vsebuje 4−terice podane v multimnožici

$$S = \{CGAA, ATAG, TAGT, GAAT, TAGA, AGTA, AATA, AGAT, GTAG, GATA, ATAG\}$$

tako, da bo $Spekter(s, 4) = S$, z uporabo

(a) Hamiltonske poti

(b) Eulerjeve sledi

Ali je rešitev enolično določena?

a) Hamiltonske poti

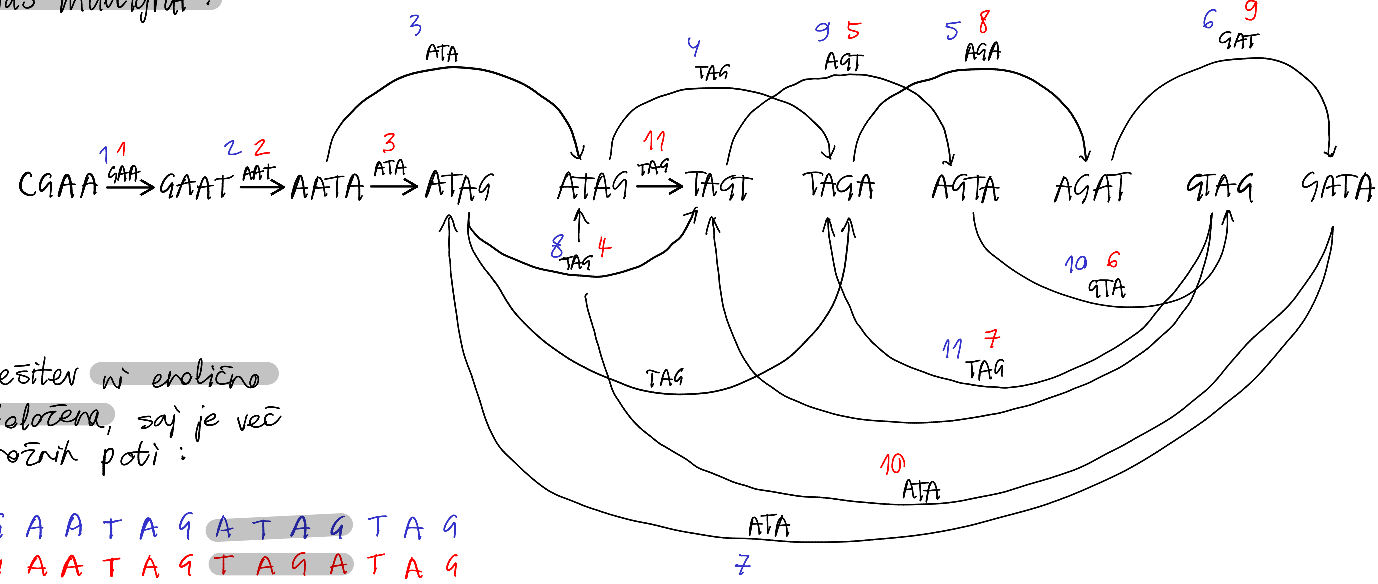
- Hamiltonska pot obišče vsako točko natanko enkrat.
- Hamiltonski cikel je cikel, ki gre skozi vse točke enkrat.
- Ne moremo podati preprostega pogoja potrebnega za obstoj hamiltonskega cikla, saj je problem NP - poln (enako za ham. poti).

- Metode sekvenciranja najprej določijo množico podzaporedij neznanega zaporedja, ki se med seboj prekrivajo. Kako ugotoviti celotno zaporedje?
  - Dano je zap. s nad abecedo Σ in $\ell \in \mathbb{N}$. ℓ-spekter zaporedja je multimnožica vseh podzap. zap. s dolžine ℓ. Označimo ga $Spekter(s, \ell)$.
  - Problem izračuna zaporedja s, za katerega velja $Spekter(s,\ell) = S$, je učinkovito rešljiv.

Naj bo $V = S$. $p, q \in E$ natanko tedaj, ko se p in q prekrivata $\Leftrightarrow$ ko je zadnjih $\ell-1$ znakov p enakih prvih $\ell-1$ znakov q.

p: A C C T G
C C T G G : q

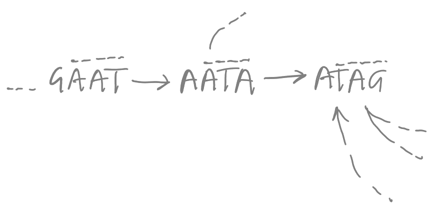
Naš multigraf:



Rešitev ni enolično določena, saj je več možnih poti:

C G A A T A G A T A G T A G
C G A A T A G T A G A T A G

Namesto, da pišeš nad povezave $\ell-1$ besede, smo naredi pikice na točkah:



3. Izračunajte zaporedje s, ki vsebuje 4−terice podane v multimnožici

$$S = \{CGAA, ATAG, TAGT, GAAT, TAGA, AGTA, AATA, AGAT, GTAG, GATA, ATAG\}$$

tako, da bo $Spekter(s, 4) = S$, z uporabo

(a) Hamiltonske poti
(b) Eulerjeve sledi

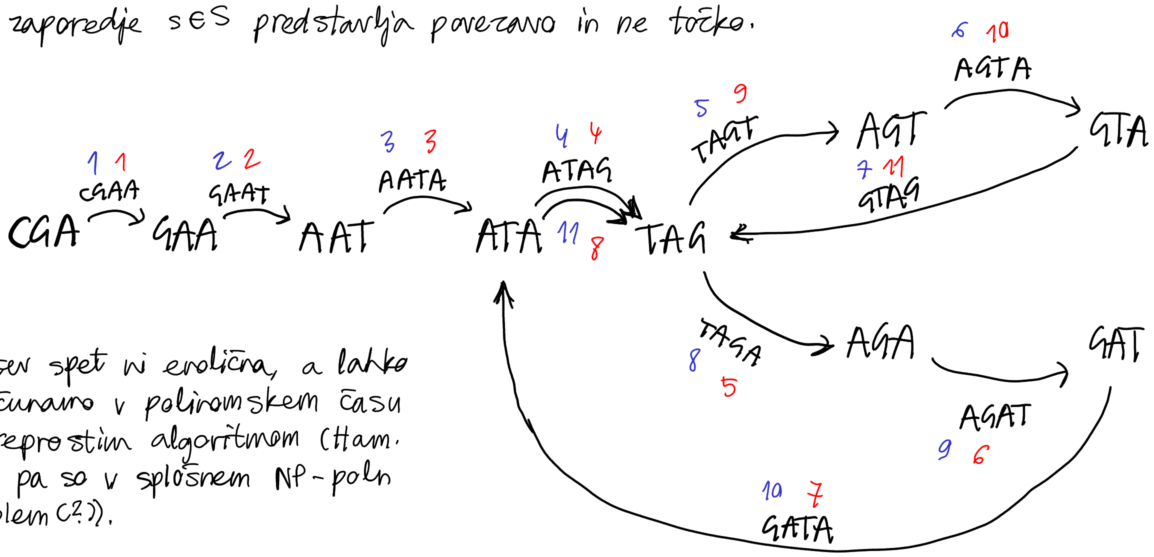
Ali je rešitev enolično določena?

b) Eulerjeve sledi

Za vhodne podatke (ℓ, S) konstruiramo digraf $D = (V, E)$ (v katerem so možne večkratne povezave), kjer je:

- V množica vseh zaporedij dolžine $\ell-1$, ki nastopajo kot podzaporedja v S.
- $(p, q) \in E$ natanko tedaj, ko $\exists$ zaporedje $s \in S$, da je prvih $\ell-1$ znakov s enakih p, zadnjih $\ell-1$ s pa q. Povezavo označimo z ℓ-terico s.

Vsako zaporedje $s \in S$ predstavlja povezavo in ne točko.



Rešitev spet ni enolična, a lahko izračunamo v polinomskem času s preprostim algoritmom (Ham. poti pa so v splošnem NP-poln problem (?)).

Rešitvi:

C G A A T A G T A G A T A G
C G A A T A G A T A G T A G

(enako kot prej :0)