

15.1.2016

# Skóre cesty

- Je třeba nepreferovat příliš krátké cesty (nevyrovnaný počet insercí a delecí)
- Cesta vede z počátečního do koncového vrcholu a tvoří ji postupně hrany  $e_1, \dots, e_n$  (může se stát, že  $e_i = e_j$  pro  $i \neq j$ ).
- Nově použitý vzorec pro výpočet skóre:

- $$\left( \prod_{i=1}^n p_i \right)^a \left( 1 - \left( \frac{I - D}{I + D + 1} \right)^{2b} \right)$$

- $p_1, \dots, p_n$  – pravděpodobnosti přechodu hran  $e_1, \dots, e_n$

- $I$  – celkový počet insercí (jednobázových)

$D$  – celkový počet delecí (jednobázových)

- $a, b$  – parametry

# Důsledky nového vzorce

- Po nalezení nové cesty se vždy provádí SWA
  - ↻ Zjišťování počtu insercí a delecí
  - ↻ SWA se provádí i pro cesty, které nakonec nebudou vybrány
  - ↻ Zdá se býti hlavní příčinou výkonnostních problémů (zatím neměřeno profilerem)
- Algoritmus již nepreferuje krátké cesty
  - ↻ Záleží na hodnotách parametrů  $a$  a  $b$
- Ořezávání nefunguje tak dobře jako předtím
  - ↻ Zejména při velkém rozdílu parametrů  $a$  a  $b$
  - ↻ Rozdíl v počtu insercí a delecí značně ovlivňuje skóre
- Hledání cest trvá podstatně déle
  - ↻ Pro  $a = 1$ ,  $b = 8$  může trvat i 40 minu
  - ↻ To asi není příliš dobrét

# Snahy o optimalizaci

- Odstranění vrcholů se vstupním či výstupním stupněm 0
  - ⌘ Bude třeba odstranit části grafu, ze kterých se nelze dostat do koncového uzlu
- Nejpravděpodobnější cesty (z pohledu pravděpodobností přechodu hran) se prohledávají nejdříve
- Vylepšení datových struktur
  - ⌘ Dříve byly hrany reprezentovány dvojicí k-merů
  - ⌘ Nyní propojení hran a vrcholů přímo ukazateli
  - ⌘ Nemusí se používat (hašovací) tabulky vrcholů a hran při hledání cest
- Zkratky
  - ⌘ Hledání cest se provádí průchodem do hloubky (DFS)
  - ⌘ Přeskakování vrcholů s jednou vstupní a jednou výstupní hranou (pravděpodobnosti přechodu jednotkové)

# Další možnosti optimalizace (eliminace útvarů přidávajících špatné cesty)

