

Primena PQ stabala u bioinformatiki

Tadija Šebez

Matematička gimnazija

16. 05. 2024.

1. PQ stablo

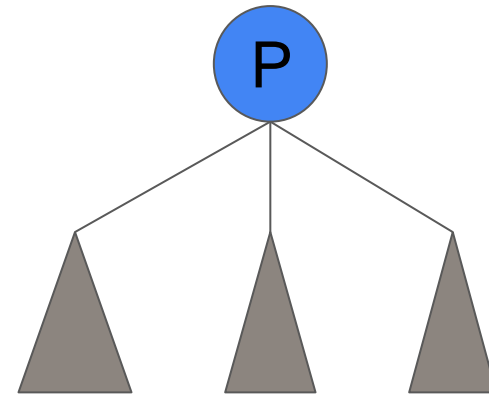
2. Osnovne primene

3. Primene u bioinformatiki

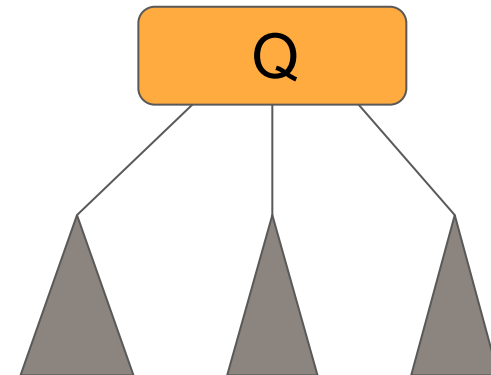
- Da li postoji permutacija brojeva od 1 do N, takva da zadovoljava sledeće uslove?
- Dato je M skupova S_i koji sadrže brojeve od 1 do N.
- Brojevi iz svakog skupa moraju biti uzastopni u permutaciji.
- Primer:
 - $N = 5, M = 2$
 - $S_1 = \{3, 4, 5\}$
 - $S_2 = \{1, 2, 3, 5\}$
 - Validne permutacije:
 - 1 2 3 5 4
 - 2 1 5 3 4
 - 4 5 3 2 1
 - Nevalidne permutacije:
 - 1 3 5 4 2
 - 3 1 2 5 4
 - 1 2 3 4 5

- PQ stablo je struktura podataka koja u svojoj strukturi enkodira sva moguća rešenja prethodnog problema.
- Ova struktura dozvoljava da se inkrementalno dodaju uslovi (skupovi).
- Primeri problema koje PQ stablo direktno rešava:
 - Da li postoji validna permutacija?
 - Koliko postoji validnih permutacija?
 - Naći leksikografski najmanju validnu permutaciju.
 - Naći sve validne permutacije.
- Postoji 3 tipa čvorova:
 - Listovi
 - P čvorovi
 - Q čvorovi

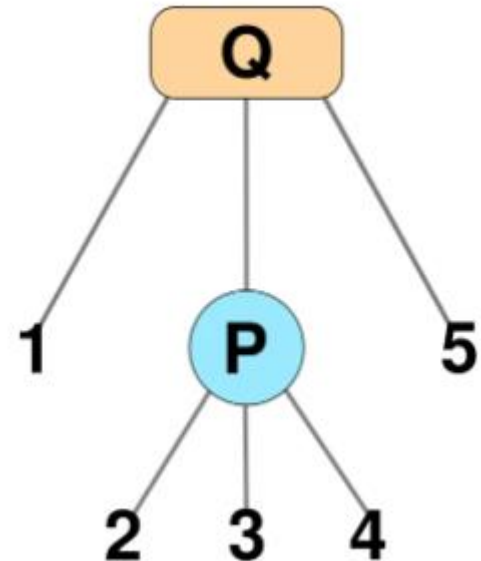
- Ima najmanje 2 deteta.
- Deca mogu da se proizvoljno permutuju.



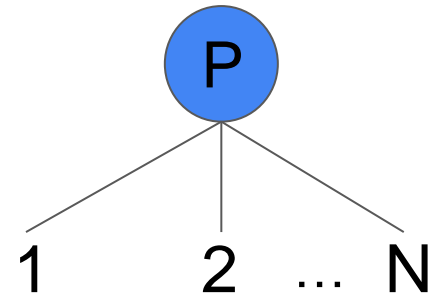
- Ima najmanje 3 deteta.
- Redosled dece može da se obrne.



- DFS obilazak PQ stabla daje validnu permutaciju listova.
- Svaki dozvoljeni redosled dece P i Q čvorova daje validnu permutaciju.
- Skup rešenja za primer sa slike je:
 - 1 2 3 4 5
 - 1 2 4 3 5
 - 1 3 2 4 5
 - 1 3 4 2 5
 - 1 4 2 3 5
 - 1 4 3 2 5
 - 5 2 3 4 1
 - 5 2 4 3 1
 - ...

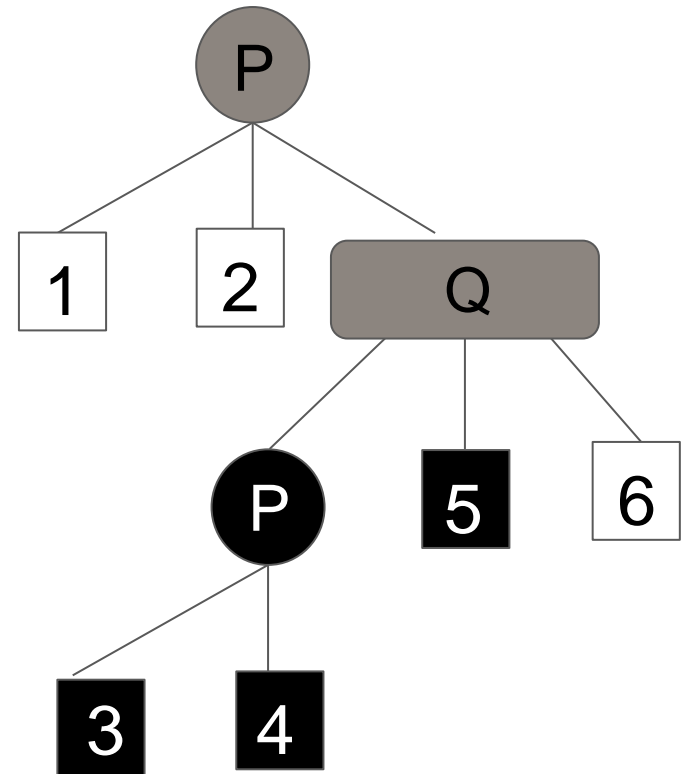


- PQ stablo se inicijalizuje tako da encoduje sve moguće permutacije.
- Koren je P čvor, a njegova deca su listovi od 1 do N.



- Čvorovi se dodaju i izbacuju kako bi se dodao uslov.
- Dve faze:
 - Bojenje čvorova
 - Menjanje stabla

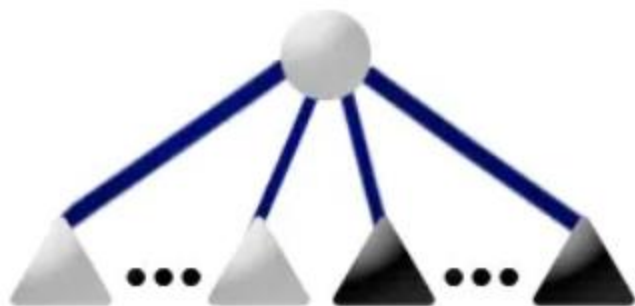
- Listovi koji moraju da budu susedni se boje u crno.
- Ostali listovi se boje u belo.
- Ako su svi listovi u podstablu nekog čvora iste boje, i taj čvor je iste te boje.
- Ostali čvorovi su sivi.



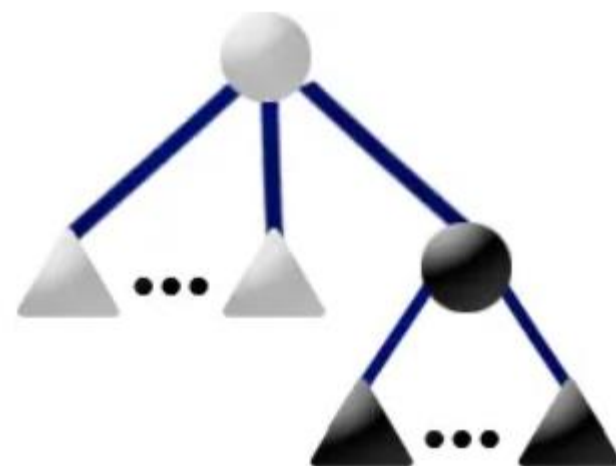
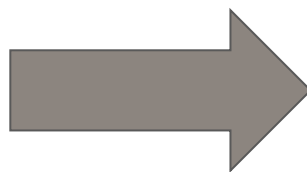
- Crni i sivi čvorovi se procesuiraju od dole na gore.
- Postoji 10 šablona na osnovu kojih se stablo menja.
- Ako se čvor koji procesuiramo ne uklapa ni u jedan od šablona onda ne postoji permutacija koja zadovoljava sve uslove.

- Leaf šablon:
 - Ako je trenutni čvor list, ništa ne menjamo.
- P1 šablon:
 - Ako je trenutni čvor crni P čvor, ništa ne menjamo.
- Q1 šablon:
 - Ako je trenutni čvor crni Q čvor, ništa ne menjamo.

- Trenutni čvor je najmanji zajednički predak za sve crne listove.

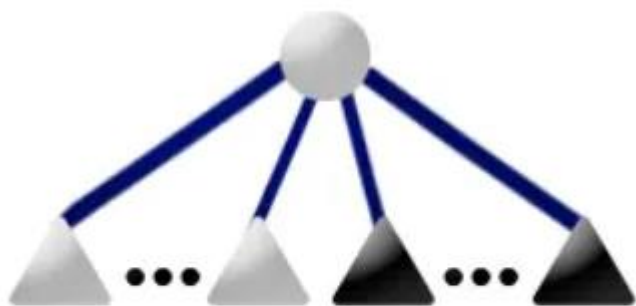


P2 Template: P-Node with both empty and full children

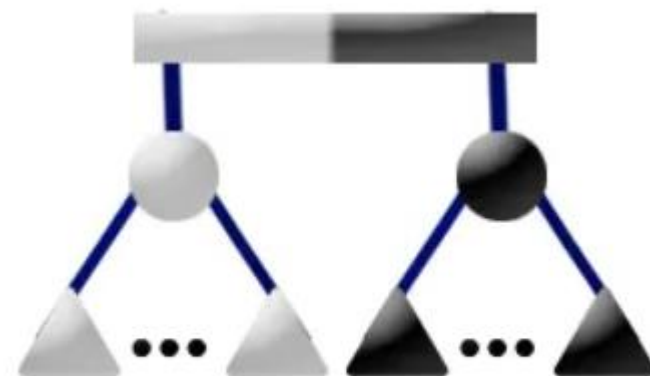
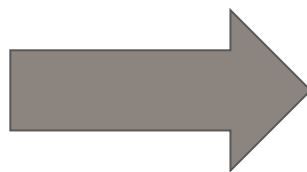


P2 Replacement

- Trenutni čvor **nije** najmanji zajednički predak za sve crne listove.

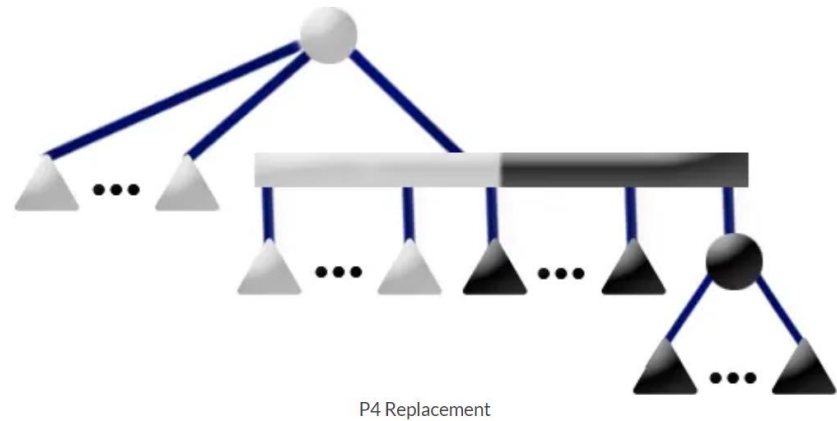
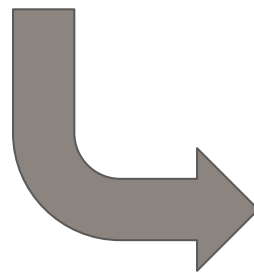
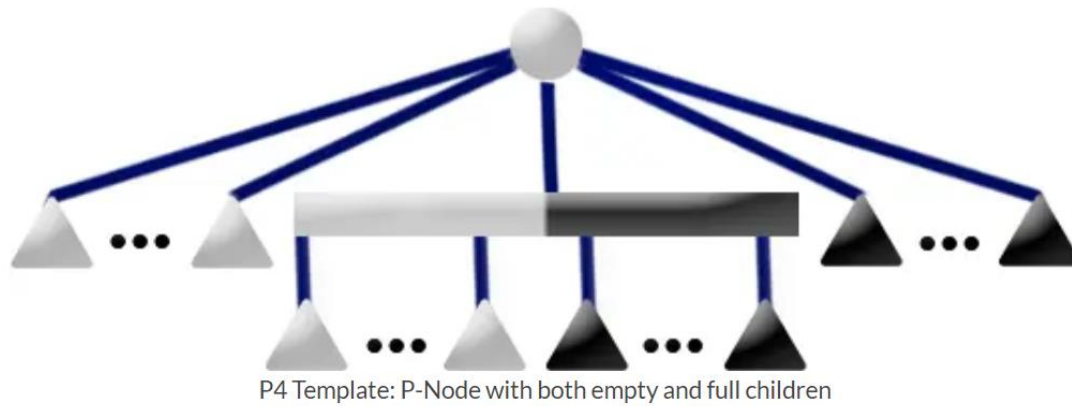


P3 Template: P-Node with both empty and full children

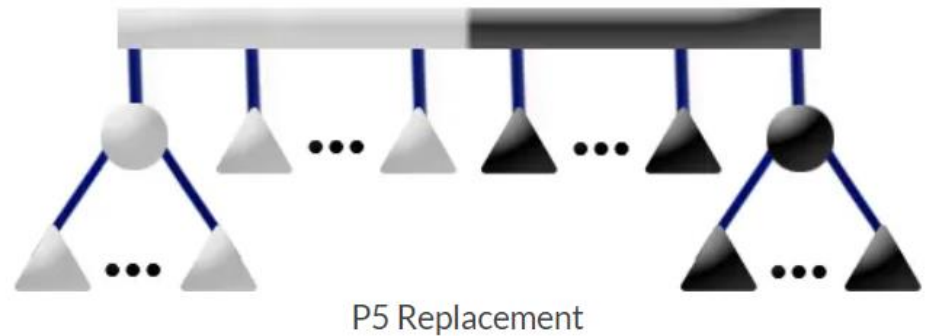
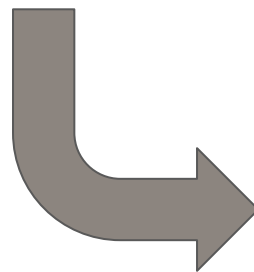
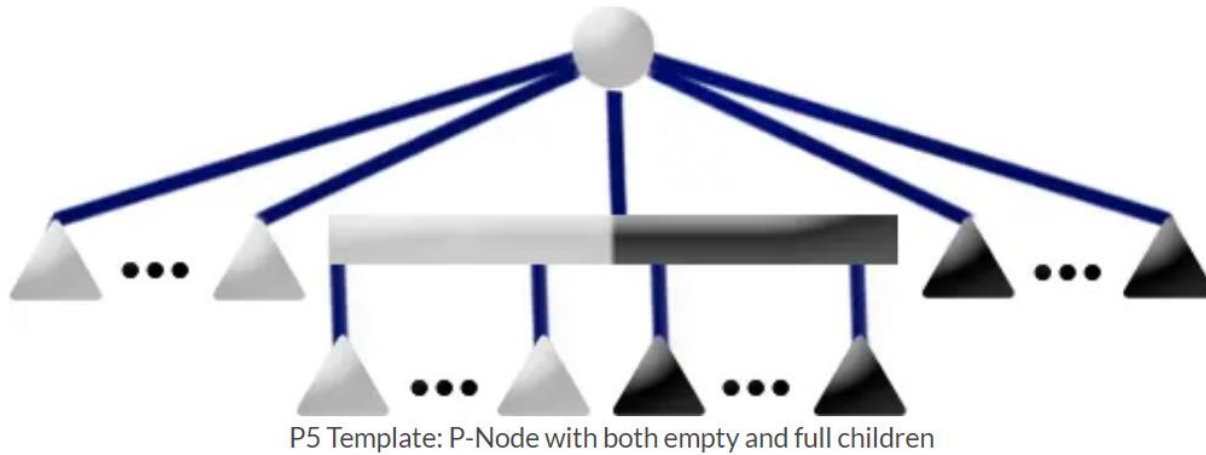


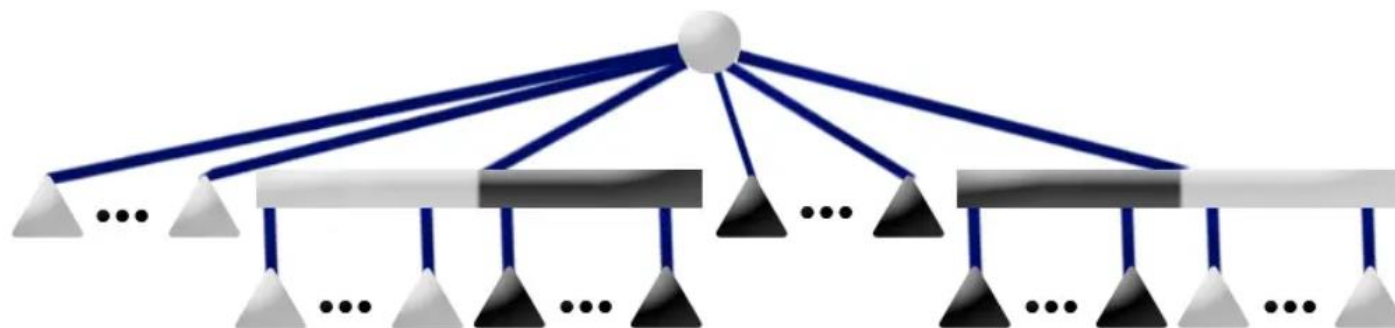
P3 Replacement

- Trenutni čvor je najmanji zajednički predak za sve crne listove.

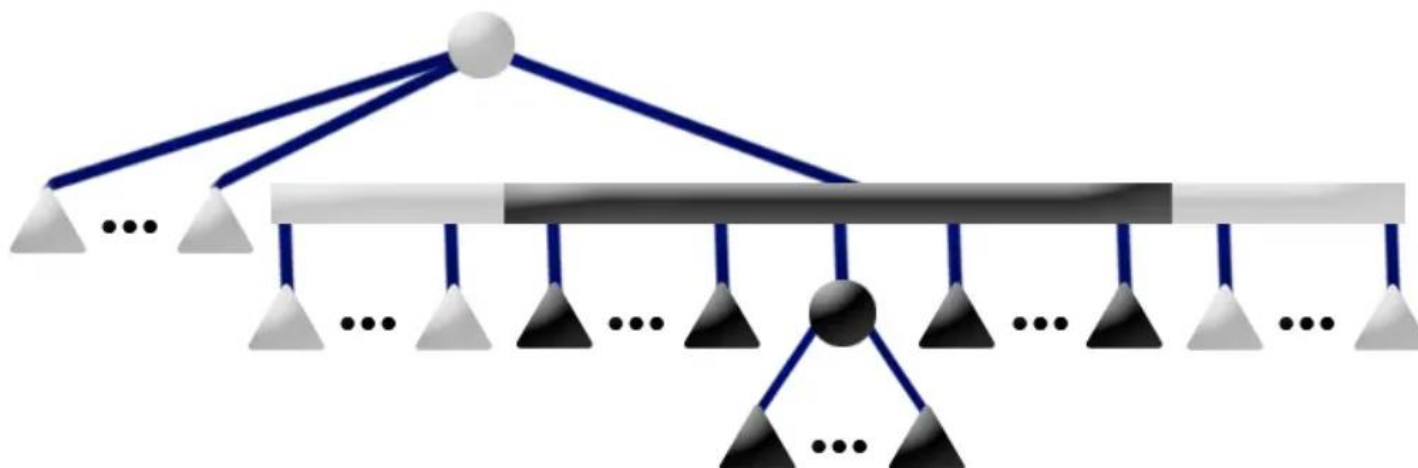


- Trenutni čvor **nije** najmanji zajednički predak za sve crne listove.

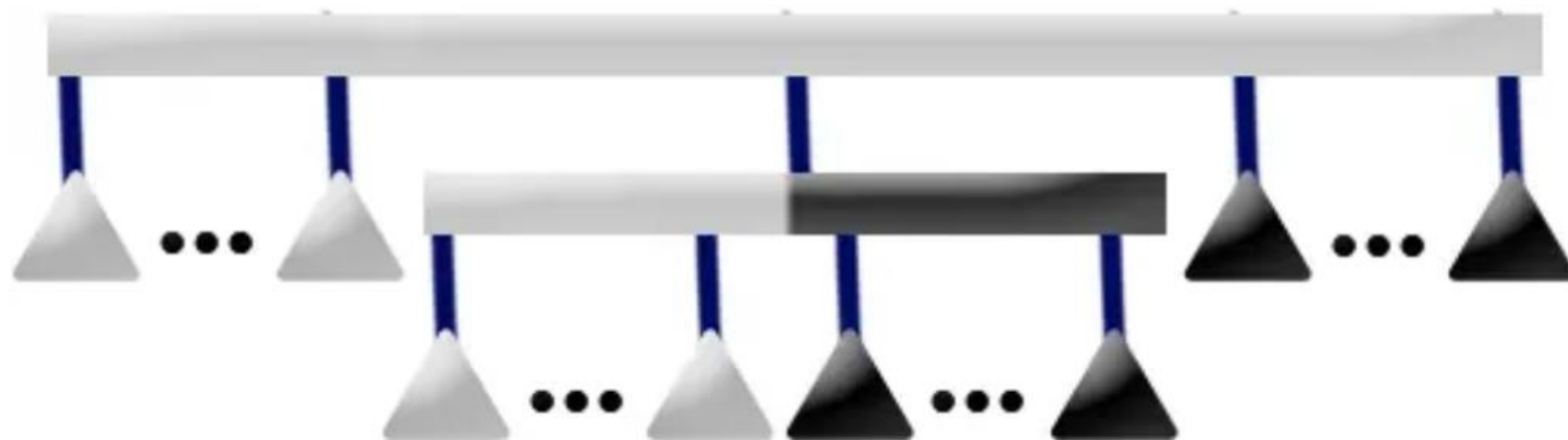




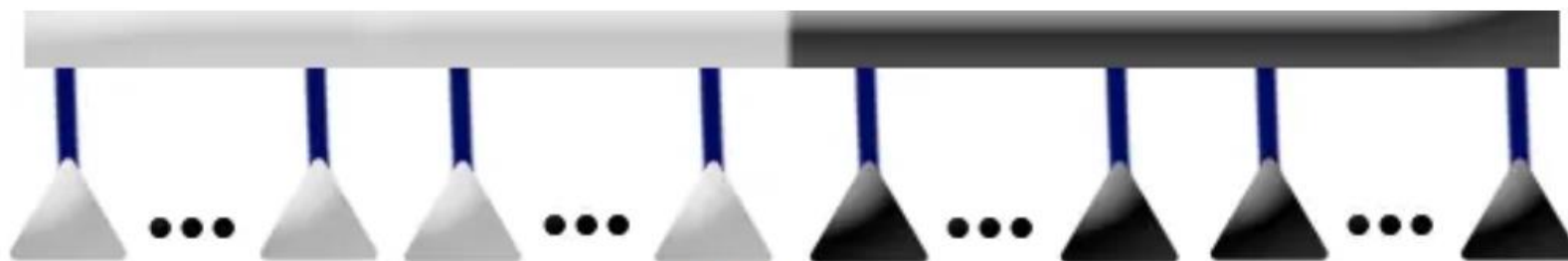
P6 Template: P-Node with both empty and full children



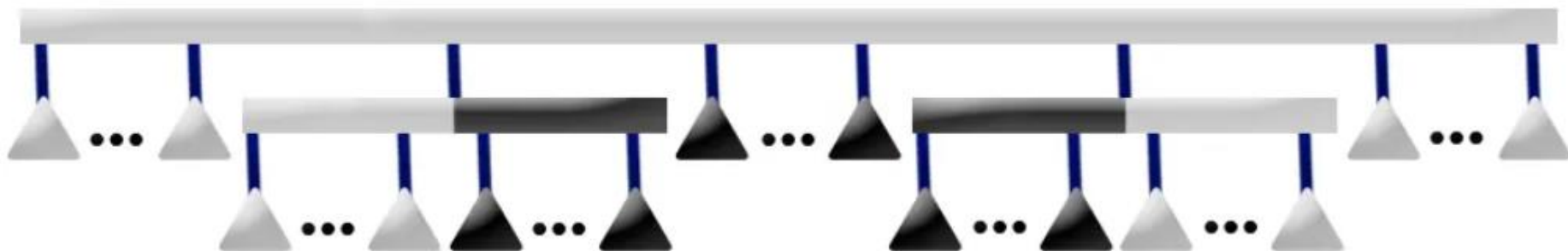
P6 Replacement



Q2 Template: Q-Node with a single partial child



Q2 Replacement



Q3 Template: Q-node with two partial children



Q3 Replacement

- Neka je K zbir veličina skupova svih uslova.
- Broj čvorova u PQ stablu je $O(N)$.
- Dobra implementacija može da dostigne vremensku složenost $O(N+M+K)$.
- Olakšane implementacije uglavnom rade u $O(N*M)$.

- Isti problem sa cikličnim permutacijama.
- Veoma slična struktura podataka.

1. PQ stablo

2. Osnovne primene

3. Primene u bioinformatiki

- Data je matrica jedinica i nula.
- Da li je moguće permutovati kolone tako da su u svakom redu sve jedinice uzastopne?

- Da li se neusmereni graf može nacrtati u ravni bez presecanja grana?

- Da li se svakom čvoru neusmerenog grafa može dodeliti interval tako da su dva čvora povezana ako i samo ako se njihovi intervali seku?

1. PQ stablo
2. Osnovne primene
3. Primene u bioinformatiki



- U idealnom slučaju DNK sekvenca može da se predstavi kao permutacija gena.
- U prirodi moguće je da se isti gen pojavljuje više puta u jednoj DNK sekvenci.

- Pre nego što se geni detektuju više kopija iste DNK sekvence se na nasumičnim mestima razbije na fragmente.
- Za svaki fragment detektuje se skup gena koje fragment nosi.
- Problem: Rekonstruisati DNK sekvencu od fragmenata.
- Idealan slučaj: Nije bilo grešaka pri merenju.
- Realan slučaj: Dešava se da se neki gen ne detektuje ili da se detektuje gen koji nije na fragmentu.
- Rešenje **nije** otporno na greške!

- Poređenje dva genoma.
- Svodi se na računanje sličnosti između dva PQ stabla.
- Breakpoint distance:
 - Koliko je najmanje inverzija podnizova potrebno da se od jednog genoma dobije drugi?

- Minimum-breakpoint-permutation problem
- Naći dve permutacije koje imaju najmanji breakpoint distance.
 - NP-complete
- Ako je jedna permutacija fiksirana?

Hvala na pažnji!

Pitanja?