

# Sammengisleit

Bergur Snorrason

February 12, 2024

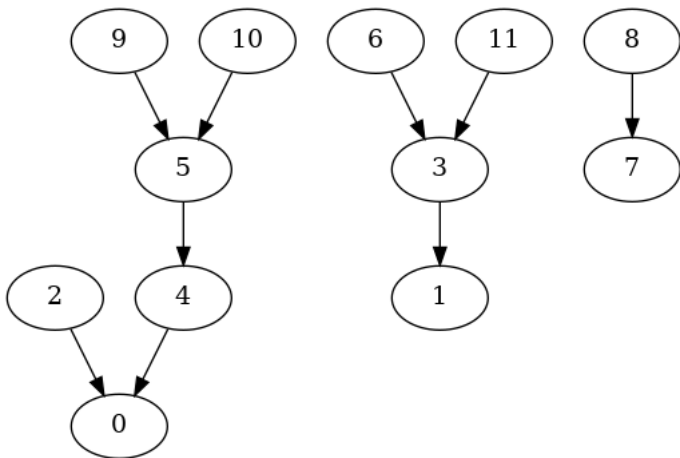
# Sammengisleit

- ▶ *Sammengisleit* (e. *union-find*) er öflug leið til að halda utan um jafngildisflokka tiltekna vensla eða, með öðrum orðum, halda utan um *sundurlæg* mengi.
- ▶ Við viljum:
  - ▶ Bera saman jafngildisflokka mismunandi staka.
  - ▶ Sameina jafngildisflokka.
- ▶ Við tölum um aðgerðirnar `find(x)` og `join(x, y)`.

- ▶ Tökum sem dæmi einstökungasafnið  $\{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}\}$ .
- ▶ `join(1, 3)` gefur okkur  $\{\{1, 3\}, \{2\}, \{4\}, \{5\}\}$ .
- ▶ `join(2, 5)` gefur okkur  $\{\{1, 3\}, \{2, 5\}, \{4\}\}$ .
- ▶ `join(2, 4)` gefur okkur  $\{\{1, 3\}, \{2, 4, 5\}\}$ .
- ▶ `join(1, 4)` gefur okkur  $\{\{1, 2, 3, 4, 5\}\}$ .
- ▶ Á sérhverjum tímapunkti myndi `find(x)` skila einhverju staki sem er í sama mengi og `x`.
- ▶ Mikilvægt er að `find(...)` skilar sama stakinu fyrir sérhvert stak í sérhverjum jafngildisflokki.
- ▶ Til dæmis, í þriðja punktinum myndi `find(1)` og `find(3)` þurfa að skila sama stakinu.
- ▶ Við köllum þetta stak *oddvita* (e. *representative*) jafngildisflokksins.

- ▶ Gerum ráð fyrir að tölurnar sem við munum vinna með séu jákvæðar og minni en  $n$ .
- ▶ Við munum þá gefa okkur  $n$  staka fylki  $p$ , þar sem  $i$ -ta stakið í fylkinu er upphafstillt sem  $i$ .
- ▶ Fylkið  $p$  mun nú geyma *foreldri* sérhvers stak.
- ▶ Foreldrin mynda keðjur, sem svara til jafngildisflokka.
- ▶ Sérhver keðja endar í einhverju staki, sem munu vera oddvitar jafngildisflokka.

- Keðjurnar sem fást með  $\{\{0, 2, 4, 5, 9, 10\}, \{1, 3, 6, 11\}, \{7, 8\}\}$  gætu til dæmis verið gefnar með  $p = [0, 1, 0, 1, 0, 4, 3, 7, 7, 5, 5, 3]$ .



- ▶ Til að fá oddvita flokks tiltekins staks er hægt að fara endurkvæmt upp keðjuna.
- ▶ Til að sameina flokka nægir að breyta foreldri oddvita annars flokksins yfir í eitthvert stak hins flokksins.
- ▶ Báðar þessar aðgerðir er auðvelt að útfæra.

## Frumstæð sammengisleit

```
4 int find(int *p, int x)
5 {
6     if (p[x] == x) return x;
7     return find(p, p[x]);
8 }
9
10 void join(int *p, int x, int y)
11 {
12     p[find(p, x)] = find(p, y);
13 }
14
15 void init(int *p, int n)
16 {
17     for (int i = 0; i < n; i++) p[i] = i;
18 }
```

- Ekki nota þennan kóða. Hann er hægari en það sem á eftir kemur.

## Ekki nota frumstæða sammengisleit

- ▶ Við sjáum nú að tímaflækja `find(...)` er línuleg í lengd keðjunnar, svo þar sem lengd keðjunnar getur verið í versta falli  $n$  þá er `find(...)` með tímaflækjuna  $\mathcal{O}(n)$ .
- ▶ Fallið `join(...)` gerir lítið annað en að kalla tvisvar á `find(...)` svo það er  $\mathcal{O}(n)$ .
- ▶ Við myndum því aðeins ná að svara  $n$  fyrirpsurnum ef  $n \leq 10^4$ .
- ▶ Því er ekki ráðlagt að nota þessa frumstæðu útfærslu.
- ▶ Hana má þó bæta.



## Keðjubjöppuð sammengisleit

- ▶ Lykilatriðið í bætingunni er að stytta keðjurnar.
- ▶ Í hvert sinn sem við köllum á `find(...)` þá fletjum við keðjun sem við heimsækjum.
- ▶ Þetta er gert með því að setja `p[x]` sem oddvita flokks `x`, í hverju skrefi endurkvæmninnar.
- ▶ Þetta köllum við *keðjubjöppun* (e. *path compression*).

- ▶ Gefum okkur  $p = [0, 0, 1, 2, 3, 4, 5, 6, 7]$ .
- ▶ Ljóst er að `find(5)` skilar 0.
- ▶ Ef við notum frumstæða sammengisleit breytist  $p$  ekki neitt þegar kallað er á `find(...)` en með keðjuþjappaðri sammengisleit þjappast keðjan frá og með 5 og því fæst  $p = [0, 0, 0, 0, 0, 0, 5, 6, 7]$ .
- ▶ Takið eftir að nú er líka styttra í oddvitann fyrir stök 6, 7 og 8, þó við heimsóttum þau ekki í endurkvæmninni.

# Keðjubjöppað sammengisleit

```
4 int find(int *p, int x)
5 {
6     if (p[x] == x) return x;
7     return p[x] = find(p, p[x]);
8 }
9
10 void join(int *p, int x, int y)
11 {
12     p[find(p, x)] = find(p, y);
13 }
14
15 void init(int *p, int n)
16 {
17     for (int i = 0; i < n; i++) p[i] = i;
18 }
```

- ▶ Þetta er þó ekki eina bætingin sem er í boði.
- ▶ Við getum einnig sameinað á skilvirkari máta með *stærðarmiðaðri sameiningu*.

## Stærðarmiðuð sameining (e. union by size)

- ▶ Þegar við sameinum keðjur þarf að velja hvor oddvitinn verður ennþá odviti.
- ▶ Við getum valið oddvitann sem hefur fleiri stök í sinni keðju.

```
4 int find(int *p, int x)
5 {
6     return p[x] < 0 ? x : (p[x] = find(p, p[x]));
7 }
8
9 void join(int *p, int x, int y)
10 {
11     int rx = find(p, x), ry = find(p, y);
12     if (rx == ry) return;
13     if (p[rx] > p[ry]) p[ry] += p[rx], p[rx] = ry;
14     else p[rx] += p[ry], p[ry] = rx;
15 }
```

- ▶ Í þessari útfærslu geymir oddvitinn neikvæða tölu, en önnur stök vísa ennþá upp keðjuna.
- ▶ Þessi neikvæða tala svarar til fjölda staka í þeim keðjum sem enda í oddvitanum.
- ▶ Svo `-p[find(x)]` er fjöldi staka í jafngildisflokki `x`.

# Keðjubjöppað og stærðarmiðuð sammengisleit

```
4 int find(int *p, int x)
5 {
6     return p[x] < 0 ? x : (p[x] = find(p, p[x]));
7 }
8
9 void join(int *p, int x, int y)
10 {
11     int rx = find(p, x), ry = find(p, y);
12     if (rx == ry) return;
13     if (p[rx] > p[ry]) p[ry] += p[rx], p[rx] = ry;
14     else p[rx] += p[ry], p[ry] = rx;
15 }
16
17 void init(int *p, int n)
18 {
19     for (int i = 0; i < n; i++) p[i] = -1;
20 }
21
22 int size(int *p, int x)
23 {
24     return -p[find(p, x)];
25 }
```

- ▶ Það er flóknara að lýsa tímaflækju keðjuþjappaðrar sammengisleitar með stærðarmiðaðri sameiningu.
- ▶ Á heildina litið (e. *amortized*) er tímaflækja hvernar aðgerðar  $\mathcal{O}(\alpha(n))$ , þar sem  $\alpha$  er andhverfa Ackermann fallsins.
- ▶ Fyrir þau  $n$  sem við fáumst við er  $\alpha(n)$  nánast fast.
- ▶ Við ímyndum okkur því alltaf að hver aðgerð sammengisleitar sé  $\mathcal{O}(1)$ .

