# Učinkovitost omrežij

Poročilo

Jure Babnik
Zala Stopar Špringer

2020
November

# Kazalo

# 1 Priprava okolja

Pred začetkom simulacij sva si pripravila delovno okolje. Za programerski del naloge sva uporabila *Python* in knjižnico *Graph-Theory*.
Defirirala sva si funkcije, ki so nama ustvarile različne enostavne grafe, kot so mreže, 3-dimenzionalne mreže, popolna binomska drevesa, cikle, itd. Vsi grafi so neusmerjeni. Nato pa sva si še definirala funkcijo *generate_random*, ki sprejme število vozlišč $n$, na katerih funkcija naredi naključen usmerjen graf.
Prav tako sva si napisala funkcije, ki izračunajo učinkovitost omrežja.

Formula za **povprečno učinkovitost** grafa $G$ je definirana kot:

$$E(G) = \frac{1}{n(n-1)} \sum_{i \neq j \in G} \frac{1}{d(i,j)},$$

kjer je $d(i,j)$ dolžina najkrajše poti med $i$-to in $j$-to točko, $n$ pa je število vseh točk v grafu.

**Globalna učinkovitost** je definirana kot:

$$E_{glob}(G) = \frac{E(G)}{E(K_n)},$$

kjer $K_n$, predstavlja poln graf na $n$ točkah.

**Lokalna učinkovitost** je definirana kot:

$$E_{loc}(G) = \frac{1}{n} \sum_{i \in G} E(G_i),$$

kjer $G_i$ predstavlja podgraf grafa $G$, ki je sestavljen le iz sosedov točke $i$ (brez točke $i$).

Vsa koda je zbrana v datoteki *graphs.py*

## 2 Povprečna učinkovitost v preprostih grafih

Ustvarila sva nekaj preprostih grafov različnih oblik in velikosti. Po zgornji formuli sva izračunala povprečno učinkovitost posameznega grafa in rezultate med seboj primerjala. Zanimalo naju je predvsem, kaj se dogaja z učinkovitostjo, ko povečujemo število vozlišč. Poleg tega pa želiva primerjati učinkovitosti grafov z enakim (podobnim) številom točk in različnih oblik.

### 2.1 Mreže $m \times n$

| $n$ | Povprečna učinkovitost | | | | | | |
|---|---|---|---|---|---|---|---|
| | $m=1$ | $m=2$ | $m=3$ | $m=4$ | $m=5$ | $m=10$ | $m=20$ |
| 2 | 1 | 0.83333333 | 0.7111111 | 0.625 | 0.5607404 | 0.3842982 | 0.2500188 |
| 3 | 0.8333333 | 0.7111111 | 0.6157407 | 0.5464646 | 0.4938095 | 0.3453350 | 0.2285143 |
| 4 | 0.7222222 | 0.625 | 0.5464646 | 0.4883333 | 0.4436090 | 0.3150960 | 0.2114040 |
| 5 | 0.6416667 | 0.5607407 | 0.4938095 | 0.4436090 | 0.4046429 | 0.2910526 | 0.1975453 |
| 10 | 0.4286596 | 0.3842982 | 0.3453350 | 0.3150960 | 0.2910526 | 0.2176605 | 0.1539108 |
| 20 | 0.2734463 | 0.2500188 | 0.2285143 | 0.2114040 | 0.1975453 | 0.1539108 | 0.1133842 |

Tabela 1: Povprečna učinkovitost $m \times n$ mrež

### 2.2 3-dimenzionalne mreže

| $n$ | $r=2$ | | | | $r=3$ | | | |
|---|---|---|---|---|---|---|---|---|
| | $m=2$ | $m=3$ | $m=5$ | $m=10$ | $m=2$ | $m=3$ | $m=5$ | $m=10$ |
| 2 | 0.6904761 | 0.5959596 | 0.4782456 | 0.3357206 | 0.5959596 | 0.5193800 | 0.4223864 | 0.3018993 |
| 3 | 0.5959596 | 0.5193800 | 0.4223864 | 0.3018993 | 0.5193800 | 0.4562206 | 0.3753992 | 0.2728478 |
| 5 | 0.4782456 | 0.4223864 | 0.3498866 | 0.2565895 | 0.4223864 | 0.3753992 | 0.3141630 | 0.2339359 |
| 10 | 0.3357206 | 0.3018993 | 0.2565895 | 0.1954146 | 0.3018993 | 0.2728478 | 0.2339359 | 0.1805703 |

Tabela 2: Povprečna učinkovitost $m \times n \times r$ mrež

| $n$ | $r=5$ | | | | $r=10$ | | | |
|---|---|---|---|---|---|---|---|---|
| | $m=2$ | $m=3$ | $m=5$ | $m=10$ | $m=2$ | $m=3$ | $m=5$ | $m=10$ |
| 2 | 0.4782456 | 0.4223864 | 0.3498866 | 0.2565895 | 0.3357206 | 0.3018993 | 0.2565895 | 0.1954146 |
| 3 | 0.4223864 | 0.3753992 | 0.3141630 | 0.2339359 | 0.3018993 | 0.2728478 | 0.2339359 | 0.1805793 |
| 5 | 0.3498866 | 0.3141630 | 0.2669560 | 0.2032711 | 0.2565895 | 0.2339359 | 0.2032711 | 0.1600518 |
| 10 | 0.2565895 | 0.2339359 | 0.2032711 | 0.1600518 | 0.1954146 | 0.1805703 | 0.1600518 | 0.1298527 |

Tabela 3: Povprečna učinkovitost $m \times n \times r$ omrežij

## 2.3 Cikli

| n | Povprečna učinkovitost |
|---|---|
| 3 | 1 |
| 4 | 0.8333333 |
| 5 | 0.75 |
| 10 | 0.4851852 |
| 20 | 0.3030493 |
| 50 | 0.1549371 |
| 100 | 0.0906910 |

Tabela 4: Povprečna učinkovitost ciklov z $n$ točkami

## 2.4 Binomska drevesa

| n | Povprečna učinkovitost |
|---|---|
| 2 | 0.8333333 |
| 3 | 0.5634921 |
| 4 | 0.3907937 |
| 5 | 0.2776549 |
| 6 | 0.2028584 |
| 7 | 0.1530067 |
| 8 | 0.1193636 |
| 9 | 0.0962377 |

Tabela 5: Povprečna učinkovitost popolnih binomskih dreves globine $n$

## 2.5 Primerjava

Najprej primerjamo mreže oblike $1 \times n$ s cikli dolžine $n$. So namreč zelo podobne oblike, razlikujejo pa se v eni sami povezavi, ki je prisotna le v ciklu. Do razlike v učinkovitosti pride ravno zaradi te povezave. Če si za primer izberemo $n = 3$ opazimo, da je v ciklu poljuben par vozlišč med seboj oddaljen 1 povezavo, medtem ko se v mreži dimenzije $3 \times 1$ pojavi par, ki je oddaljen 2 povezavi. Ker se v formuli razdalja med parom vozlišč pojavlja v imenovalcu, večja razdalja pomeni manjšo učinkovitost. Poleg tega se v primeru cikla zaradi dodatne povezave pot med marsikaterim parom točk skrajša. Tudi najini izračuni povejo enako zgodbo. Če si spet pogledmo kot primer $n = 3$, ima cikel povprečno učinkovitost 1, mreža pa le 0.833. Tudi za večje grafe, npr. $n = 20$, opazimo razliko; cikel ima poveprečno učinkovitost 0.303, mreža pa le 0.273.

Podobna pričakovanja sva imela glede primerjave med mrežami oblike $1 \times n$ ter $m \times n$ (na enakem številu točk). Mreže z več povezavami so zaradi podobnega premisleka kot v prejšnjem primeru bolj učinkovite. Če si primerjamo npr. mrežo $10 \times 1$ z mrežo $5 \times 2$ opazimo, da ima kljub enakemu številu točk slednja učinkovitost 0.561, kar je več kot 0.273 (učinkovitost mreže $10 \times 1$).

Podobno, če za primer vzamemo mreži dimenzij $10 \times 2$ in $5 \times 4$, lahko iz tabele razberemo, da ima tabela dimenzije $5 \times 4$ večjo učinkovitost, kljub enakemu številu točk.

Kot nadaljevanje prejšnje točke sva primerjala tudi 2-dimenzionalne mreže s 3-dimenzionalnimi. Pričakovala sva podoben rezultat; da so 3-dimenzionalne bolj učinkovite od 2-dimenzionalnih. Predvidevanje se je izkazalo za resnično. 2-dimenzionalne mreže dimenzije $m \times n$ lahko razumemo kot 3-dimenzionalno dimenzij $1 \times m \times n$. Po podobnem premisleku kot v primerjavi mrež $1 \times n$ z $m \times n$ rezultat ni nobeno presenečenje.

Nazadnje sva primerjala še mreže z binomskimi drevesi. Zaradi povsem različne strukture obeh grafov nisva bila povsem prepričana, kakšne rezultate naj pričakujeva. Vseeno pa sva se rahlo nagibala v prid mrež, saj imajo več povezav. Rezultati so najino hipotezo podprli. Za primer vzamemo mrežo dimenzije $3 \times 5$ ter binomsko drevo višine 4 (saj imata oba 15 vozlišč). Mreža ima učinkovtost 0.494, drevo pa le 0.390. Podobno ima mreža dimenzij $5 \times 5 \times 5$ 125 točk in povprečno učinkovitost 0.267, drevo višine 7 pa 127 točk in povprečno učinkovitost 0.153 (torej bistveno manj).

Po zgornjih ugotovitvah sva vsakič prišla do sklepa, da je graf z več povezavami bolj učinkovit. Zato sva se odločila, da bova binomska drevesa in mreže primerjala tako, da med seboj primerjava grafa z enakim (podobnim) številom povezav. Za začetek si poglejmo manjši primer; mreža $2 \times 3$ in binomsko drevo višine 3. Drevo ima 6 povezav in povprečno učinkovitost 0.563, mreža pa ima povezavo več in učinkovitost 0.711. Če nato primerjamo mrežo dimenzij $2 \times 3 \times 5$ z binomskim drevesom višine 6, dobimo podoben rezultat; mreža ima 59 povezav in povprečno učinkovitost 0.422, drevo pa 3 povezave več in učinkovitost 0.202. Je pa treba omeniti, da ima mreža precej manj točk kot drevo.

# 3 Povprečna učinkovitost izračunana s približkom

Ker je formula za računanje povprečne učinkovitosti zelo zahtevna, sva se odločila, da bova preverila, kako velika je napaka, če povprečno učinkovitost izračunamo s približki. Zato sva napisala funkcijo *average_sim*, tako da sva ustrezno spremenila formulo za povprečno učinkovitost. Funkciji podamo odstotek, kolikšen delež parov točk naj vzame. Nato sešteje obratne vrednosti dolžine povezav med izbranimi pari točk, na koncu pa vsoto deli s številom parov. Računamo torej po enakem postopku, le da vzamemo samo naključno podmnožico parov točk.
Za različne grafe in različno število vozlišč sva računala približke za 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 25% in za 50% parov točk. Za vsak graf in vsak odstotek sva postopek ponovila 10 krat, 100 krat in 1000 krat ter nato izračunala povprečje dobljenih rezultatov.

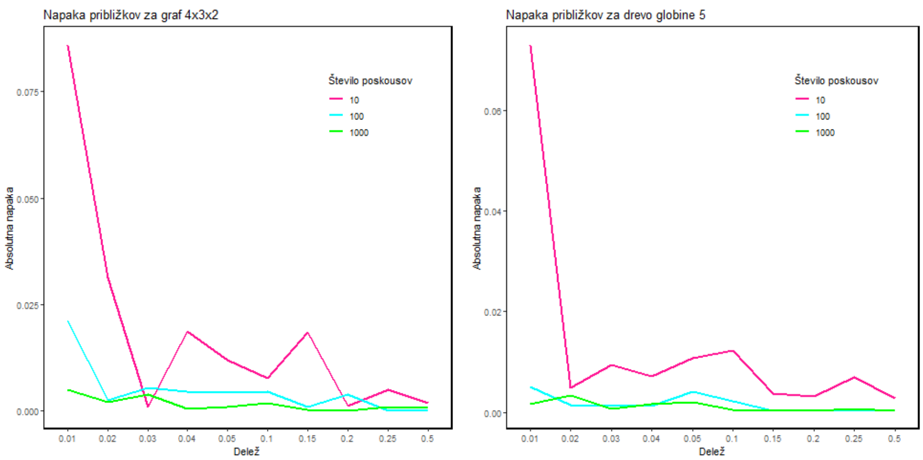
## 3.1 Mreža dimenzije $3 \times 1$

| Delež | Povprečna učinkovitost | Približek 10 | Približek 100 | Približek 1000 |
|---|---|---|---|---|
| 0.01 | 0.8333 | 0.8 | 0.79 | 0.815 |
| 0.02 | 0.8333 | 0.85 | 0.825 | 0.8405 |
| 0.03 | 0.8333 | 0.75 | 0.81 | 0.8365 |
| 0.04 | 0.8333 | 0.85 | 0.86 | 0.8405 |
| 0.05 | 0.8333 | 0.9 | 0.855 | 0.828 |
| 0.1 | 0.8333 | 0.75 | 0.815 | 0.8315 |
| 0.15 | 0.8333 | 1.0 | 0.87 | 0.8335 |
| 0.2 | 0.8333 | 0.85 | 0.835 | 0.8298 |
| 0.25 | 0.8333 | 0.8 | 0.845 | 0.8315 |
| 0.5 | 0.8333 | 0.85 | 0.8383 | 0.8322 |

Tabela 6: Povprečna učinkovitost in približki za mrežo dimenzije $3 \times 1$

Najprej sva se lotila majhnega primera in rešitve zapisala v preglednico. Opazimo, da se z večanjem števila ponovitev približek bliža točni vrednosti. Poleg tega pa lahko vidimo tudi, da delež ne vpliva prav dosti na približek. To je seveda logično, saj ima opazovani graf le tri pare točk. V vsaki ponovitvi izberemo navzgor zaokroženo število parov točk. V primeru, ko izberemo delež manjši od $\frac{1}{3}$, bomo izbrali le en par točk. V najinem primeru se to zgodi vsakič, razen ko izberemo 50% parov točk.

## 3.2 Mreža dimenzije $4 \times 3 \times 2$ in drevo z globino 5




Slika 1: Prikaz absolutne napake za mrežo velikosti $4 \times 3 \times 2$ in drevo globine 5

Iz zgornjega grafa lahko razberemo, da več kot imamo ponovitev, manjša je napaka. Mreža (levo) ima 552 različnih parov točk, drevo (desno) pa 930. Vidimo, da za razliko od prejšnjega primera, povprečna absolutna napaka pada z večanjem deleža parov točk.

# 4 Globalna učinkovitost

Zanimala naju je tudi globalna učinkovitost. Odločila sva se, da jo bova izračunala samo za 3-dimenzionalne mreže in za drevesa. Natančneje, izbrala sva mreži dimenzij $5 \times 5 \times 5$ in $7 \times 7 \times 7$ ter drevesi globin 7 in 9. To je zato, ker so $n \times m$ mreže pravzaprav 3-dimenzionalne, s tem, da je ena dimenzija enaka 1, cikli pa so podobni mrežam $m \times 1 \times 1$.
V vsaki ponovitvi sva grafu odstranila določen delež povezav, za idealen graf pa sva vzela prvotni graf. Z odstranjevanjem povezav sva simulirala nedelujoče povezave. Odstranjevala sva jih na dva načina. Pri prvem sva povezave odstranjevala povsem naključno, pri drugem pa tako, da so bile "nedelujoče"povezave zbrane okoli ene točke oziroma enega območja točk. Pri tem sva epicenter izbrala naključno.

| Delež | Mreža $5 \times 5 \times 5$ | | Mreža $7 \times 7 \times 7$ | | Drevo globine 7 | | Drevo globine 9 | |
|---|---|---|---|---|---|---|---|---|
| | Naključno | Območje | Naključno | Območje | Naključno | Območje | Naključno | Območje |
| 0.01 | 0.997986 | 0.996002 | 0.998649 | 0.992360 | 0.973268 | 0.959905 | 0.919040 | 0.947637 |
| 0.02 | 0.996436 | 0.983139 | 0.996750 | 0.979270 | 0.841759 | 0.959088 | 0.915984 | 0.895127 |
| 0.03 | 0.993971 | 0.976350 | 0.994922 | 0.962184 | 0.852714 | 0.908842 | 0.891352 | 0.799904 |
| 0.04 | 0.991503 | 0.961187 | 0.992949 | 0.956771 | 0.563543 | 0.818094 | 0.775182 | 0.799239 |
| 0.05 | 0.988266 | 0.949432 | 0.991327 | 0.930986 | 0.761260 | 0.815371 | 0.760988 | 0.642016 |
| 0.10 | 0.976811 | 0.863428 | 0.981988 | 0.867520 | 0.694765 | 0.660604 | 0.378180 | 0.388729 |
| 0.15 | 0.955474 | 0.816746 | 0.966606 | 0.791326 | 0.435229 | 0.461789 | 0.282195 | 0.245252 |
| 0.20 | 0.931926 | 0.761820 | 0.944243 | 0.728858 | 0.456664 | 0.404411 | 0.137504 | 0.160098 |
| 0.25 | 0.902518 | 0.707743 | 0.921387 | 0.660562 | 0.245534 | 0.271947 | 0.102830 | 0.110347 |
| 0.50 | 0.627812 | 0.393352 | 0.657113 | 0.362266 | 0.114200 | 0.092441 | 0.033627 | 0.036456 |

Tabela 7: Globalna učinkovitost mrež in dreves

Iz tabele kkkk lahko razberemo, da je učinkovitost veliko manjša, če so nedelujoče povezave skoncentrirane na določenem območju. Odstranitev povezav je veliko manj vplivala na globalno učinkovitost kot sva pričakovala. Še posebej pri naključno odstranjenih povezavah. Dobljen rezultat je smiselen, saj se v primeru zgoščenega nedelujočega območja dolžina poti med določenimi pari točk drastično poveča.
Do podobnih ugotovitev pridemo tudi pri binomskih drevesih, z dvema pomembnima razlikama. Globalna učinkovitost je tokrat slabša, če povezave odstranimo naključno. Vzrok za to je dejstvo, da med poljubnim parom točk obstaja natanko ena pot. Če je katera koli povezava na poti nedelujoča, dolžina poti postane enaka $\infty$ (torej pot ne obstaja). Če pa je nedelujoče območje zgoščeno, pa dodatne nedelujoče povezave "pretrgajo"večinoma le take poti, ki so že "pretrgane". Zaradi podobnega vzroka so globalne učinkovitosti izračuanne pri drevesih precej nižje kot pri mrežah. Tudi to lahko pripišemo sami strukturi obeh omrežij. Odstranitev povezav ima pri drevesih torej većji vpliv kot pri mrežah.

# 5 Sklep

V najini nalogi sva opazovala, kako se različne učinkovitsti spreminjajo z večajnem števila vozlišč grafa ter spreminjanjem njegove oblike. Izračunala sva točno vrednosti povprečne učinkovitosti grafov ter približek, prav tako pa sva izračunala gobalno učinkovitost grafov, ki imajo nekaj povezav nedelujočih.

Po pričakovanjih povprečna učinkovitost pada z večanjem števila točk. To se zdi smiselno, saj so v večjih omrežjih posamezni pari točk med seboj bolj oddaljeni. Slednje velja za grafe vseh zgoraj omenjenih oblik. Ugotovila sva tudi, da so izmed preizkovanih omrežij najbolj učinkovite 3-dimenzionalne mreže, najmanj pa mreže dimenzije $n \times 1$. To je zato, ker imajo 3-dimenzionalne mreže več povezav, čeprav primerjamo grafa z enakim številom vozlišč. S tem, ko sva primerjala binomska drevesa in mreže, sva ugotovila, da večje število povezav ne pomeni nujno večje učinkovitosti. Iz najinih izračunov sklepava, da so najbolj učinkoviti tisti grafi, pri katerih je posamezno vozlišče povezano s čim več ostalimi. Za primer lahko uzamemo poln graf, ki ima povprečno učinkovitost vedno enako 1.

Približki za računanje povprečne učinkovitosti se nama zdijo precej natančni. Po najinem mnenju je smiselno, da bi povprečno učinkovitost računali na manjši podmnožici parov točk npr. na 1% vseh parov. Rezultati bi bili bolj natančni, če bi postopek ponovili večkrat in nato izračunali povprečje, vendar več ponovitev pomeni večjo zahtevnost. Je pa treba omeniti, da sva zaradi velike računske zahtevnosti vse poskuse izvajala na relativno preprostih grafih. Smiselno bi bilo raziskave nadaljevati tudi na večjih oz. bolj obsežnih grafih preden delava kakršne koli končne zaključke.

Ugotovila sva, da so v primeru nedelujočih povezav mreže veliko bolj učinkovite kot pa binomska drevesa, kar je posledica tega, da med naključnim parom točk v binomskem drevesu obstaja le ena pot. Seveda pa se učinkovitost manjša z večanjem števila odstranjenih povezav.