

Število podpoti pri grafih z danim ciklomatskim številom

*Poročilo projektne naloge
Skupina 18: Maja Košmrlj, Jan Maradin*

Uvod

V tej projektni nalogi sva za dano ciklomatsko število in število vozlišč poiskala grafe z najmanjšim in največjim številom podpoti ter geodetskim številom podpoti, nato pa sva analizirala njihove oblike in lastnosti. Zanimalo naju je predvsem, kako so videti takšni grafi in kako se njihove strukture postopoma spreminjajo z večanjem števila vozlišč ter ciklomatskega števila.

Projektne naloge sva se v bistvu lotila na dva med seboj povezana načina. V prvem delu sva se osredotočila na analizo različnih metod simuliranega ohlajanja (SA) ter primerjavo rezultatov pri spremembah parametrov. V tem delu sva obravnavala grafe z manjšim številom vozlišč in vsemi možnimi vrednostmi ciklomatskega števila. V drugem delu naloge pa sva se usmerila na grafe z manjšim ciklomatskim številom ter podrobneje analizirala njihove oblike in značilnosti.

Teoretično ozadje

- **Graf:** Graf $G = (V, E)$ je sestavljen iz množice vozlišč V in množice povezav E , kjer je vsaka povezava element oblike $\{u, v\}$ za $u, v \in V$.
- **Povezan graf:** Graf G je povezan, če med vsakim parom vozlišč obstaja pot.
- **Pot:** Pot v grafu G dolžine ℓ je zaporedje različnih vozlišč $(v_0, v_1, \dots, v_\ell)$, kjer za vsak $1 \leq i \leq \ell$ velja, da je $v_{i-1}v_i \in E$.
- **Ciklomatsko število:** Je število neodvisnih ciklov, ki jih graf vsebuje, oziroma nam pove koliko povezav bi morali odstraniti, da bi dobili gozd. Ter je definirano kot:

$$\mu(G) = |E| - |V| + 1$$

- **Podpot** Naj bo $G = (V, E)$ povezan graf in $P = v_1v_2 \dots v_k$ pot v grafu G . Podpot grafa G je vsaka zaporedna podzaporedje oglišč te poti, torej vsaka pot oblike

$$v_i v_{i+1} \dots v_j, \quad \text{kjer } 1 \leq i < j \leq k.$$

- **Število podpoti:** Je skupno število vseh enostavnih poti v grafu, vključno s trivialnimi potmi in ga označimo s $p_n(G)$.

- **Geodetsko število** Za $G = (V, E)$ povezan graf označimo geodetsko število podpoti z $gp_n(G)$ in definiramo kot skupno število vseh geodetskih enostavnih poti v grafu. Za vsaka dva vozlišča $u, v \in V(G)$ označimo z $d(u, v)$ razdaljo med njima, z $\sigma^*(u, v)$ pa število vseh enostavnih $u-v$ poti dolžine $d(u, v)$. Tedaj je

$$gp_n(G) = \sum_{u, v \in V(G)} \sigma^*(u, v).$$

Vračunane so tudi poti dolžine 0, torej $u = v$. Geodetsko število podpoti tako meri, koliko najkrajših enostavnih poti vsebuje graf.

- **Kaktusni graf** je graf, v katerem se poljubna dva cikla stikata v največ enem vozlišču. To pomeni, da se cikli lahko dotikajo, vendar se ne prekrivajo.
- **Graf PTC(n,k)** (angl. *Pseudo Triangle Chain*), oziroma psevdotrikotna veriga, je poseben primer kaktusnega grafa z n vozlišči in k cikli. Vsi cikli so trikotniki, povezani v verigo in zato graf dosega največje možno število podpoti med vsemi takšnimi grafi.
- **Izrek** (Knor, 2025) Graf $PTC(n, k)$ enolično maksimizira število podpoti med vsemi kaktusnimi grafi z n vozlišči in $k \geq 2$ cikli.

Simulated annealing- simulirano ohlajanje

Za velike grafe sva tako v 1. delu kot v 2. delu uporabljala Simulated annealing. To je optimizacijski algoritem, ki se uporablja za iskanje približno najboljših rešitev na grafih, zlasti kadar je iskalni prostor zelo velik in kompleksen. Metoda posnema proces fizičnega ohlajanja kovin: na začetku dovoljuje tudi „slabše“ premike, da lahko pobegne iz lokalnih ekstremov, nato pa postopno zmanjšuje verjetnost takšnih premikov, ko se temperatura znižuje. Prednost metode je, da lahko najde dobre rešitve tam, kjer bi druge metode hitro obtičale, saj s svojo naključnostjo bolj učinkovito raziskuje celoten prostor možnih rešitev. Pri vsakem koraku algoritem iz trenutne rešitve x generira novo kandidatno rešitev x' . Če je nova rešitev boljša, jo sprejmemo. V nasprotnem primeru jo sprejmemo z verjetnostjo

$$P = e^{-\frac{\delta}{T}},$$

kjer je δ razlika med novo vrednostjo ciljne funkcije in trenutno najboljšo vrednostjo. Parameter T predstavlja temperaturo, ki se skozi iteracije postopoma zmanjšuje. Temperatura se ohlaja po pravilu

$$T_{k+1} = \alpha T_k,$$

pri čemer sva v tej projektni nalogi izbrala $\alpha = 0.995$. Začetno temperaturo pa sva po potrebi prilagajala.

1. del projektne naloge

V prvem delu projektne naloge sva želela preveriti, kateri način simuliranega ohlajanja (SA) je najučinkovitejši. Za dano število vozlišč in vsa pripadajoča ciklomatska števila sva poiskala zgolj grafe z najmanjšim oziroma največjim številom podpoti, pri čemer se nisva osredotočala na geodetske podpoti. Ker sva grafe iskala za vsa možna ciklomatska števila, je bil postopek zelo časovno zahteven, zato sva analizo omejila na grafe z največ 9 vozlišči.

Izračun točnih ekstremnih vrednosti

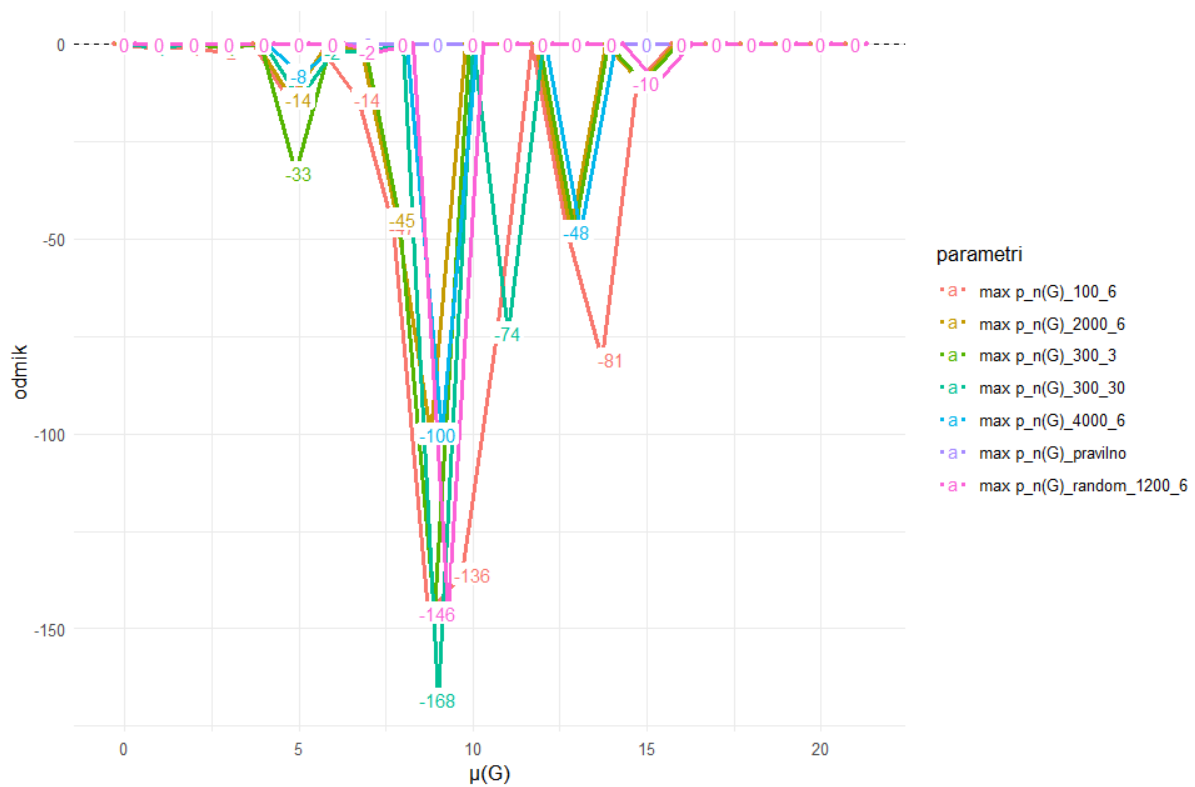
- Vse datoteke za ta del so shranjene v mapi `1_del`; koda je zapisana v datoteki `1del_mali_grafi.ipynb`, rezultati pa v `1del_mali.csv`.
- Najprej sva definirala funkcijo, ki v podanem grafu poišče vse podpoti. To funkcijo sva nadgradila tako, da za dano ciklomatsko število in število vozlišč najprej generira vse možne grafe za ta par, nato pa izbere in shrani tistega z najmanjšim oziroma največjim številom podpoti. To funkcijo sva nato pognala na vseh grafih z največ 9 vozlišči in vsemi možnimi ciklomatskimi števili (to število je navzgor omejeno s polnostjo grafa; na primer pri $|V| = 9$ je maksimalni $\mu = 28$).
- Celotno izvajanje kode je trajalo približno 80 minut.

1. del z algoritmom SA

V tem delu sva proučevala delovanje algoritma simuliranega ohlajanja (SA) pri različnih nastavitvah parametrov. Posebej naju je zanimal vpliv različnih začetnih približkov na kakovost rezultatov pri grafih z $n = 9$, medtem ko sva pri grafih z $n = 8$ analizirala, kako spremembe parametrov algoritma vplivajo na njegovo uspešnost. Dobljene rezultate sva primerjala s točnimi vrednostmi, izračunanimi v prejšnjem delu.

$$n = 8$$

- Nato sva izvajala SA na grafih z osmimi vozlišči. Tu sva pregledovala kako spreminjanje temperature in števila korakov vpliva na rezultate in čas trajanja algoritma. Primerjala sva kaj je razlika ko je začetna temperatura 30 ali 3, ter kakšne so razlike v rezultatih ko je število korakov 100, 2000 ali 4000. Za konec pa sva pognala kodo brez dobrih začetnih približkov, torej sva kot začetni graf dala v SA naključen graf z izbranim ciklomatskim številom in osmimi vozlišči.
- **Opomba:** Za maksimum pri 4000 korakih je zagon kode trajal 8 minut za maksimum pri začetni temperaturi 30 in številu korakov 300 pa 3,5 minut.
- **UGOTOVITVE**

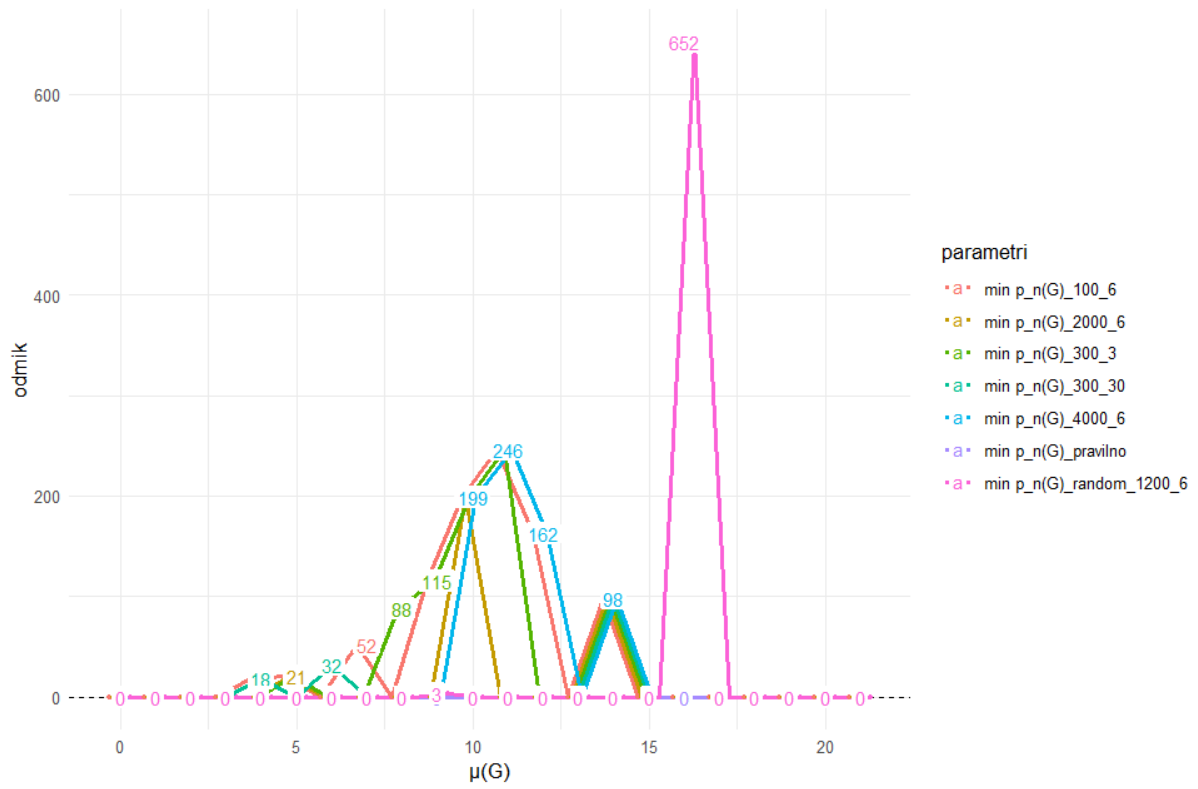


Slika 1: Graf odmikov pri različnih parametrih SA za maksimalno število podpoti v grafih z $n = 8$

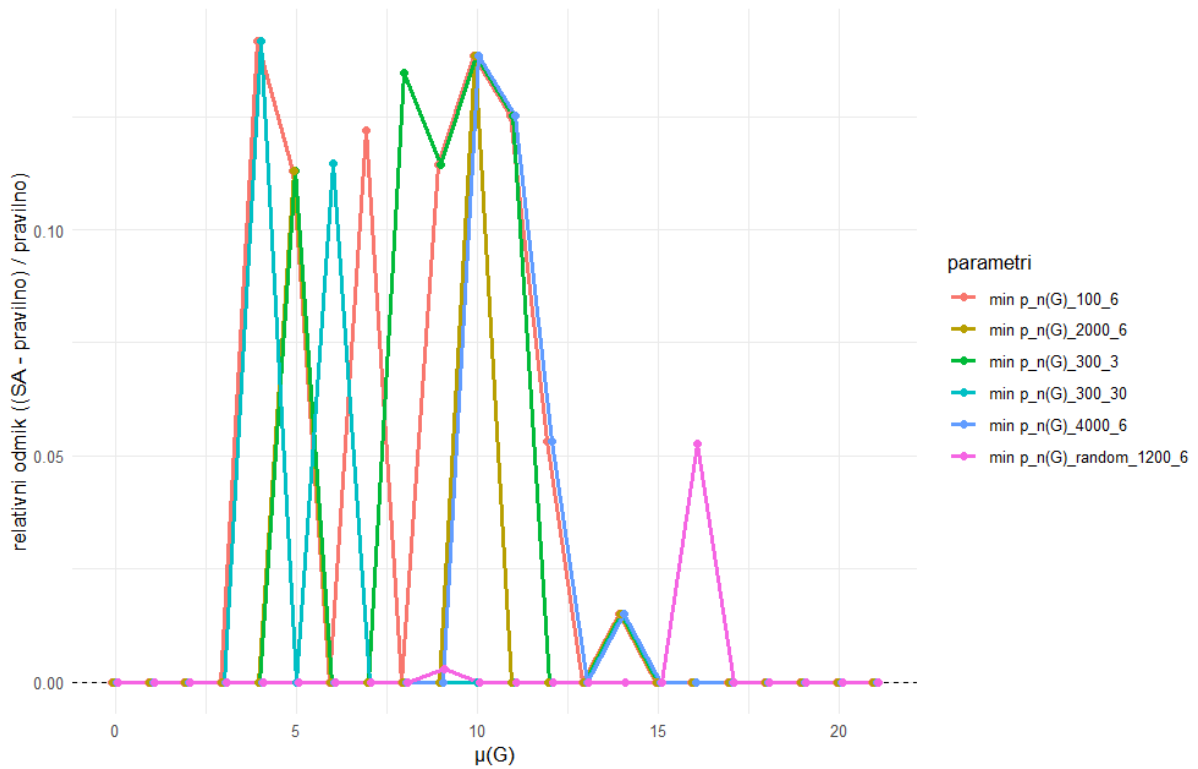
	parametri	vsota_odmikov	stevilo_odmikov
1	max p_n(G)_100_6	-576	13
2	max p_n(G)_2000_6	-219	6
3	max p_n(G)_300_3	-283	6
4	max p_n(G)_300_30	-261	6
5	max p_n(G)_4000_6	-156	3
6	max p_n(G)_pravilno	0	0
7	max p_n(G)_random_1200_6	-158	3

Slika 2: Povzetek za maksimalno število podpoti za $n = 8$

Iz zgornjih grafov za maksimalne število podpoti lahko opazimo, da sta najmanjša odstopanja dosegla algoritem SA, inicializiran z naključnimi grafi, ter algoritem SA s 4000 koraki. Pričakovano je, da večje število korakov vodi do boljšega približka, vendar pa je bilo presenetljivo, da je tudi naključna inicializacija pogosto pripeljala do pravilne vrednosti. Kot sva pričakovala, je algoritem SA z najmanjšim številom korakov vrnil najslabše približke, kar je razvidno tudi iz prikazanih grafov. Poleg tega sva ugotovila, da povišanje začetne temperature ni bistveno izboljšalo iskanja grafa z največjim številom podpoti.



Slika 3: Graf odmikov pri različnih parametrih SA za minimalno število podpoti v grafih z $n = 8$



Slika 4: Graf relativnih odmikov pri različnih parametrih za minimalno število podpoti v grafih z $n = 8$

	parametri	vsota_odmikov	stevilo_odmikov
1	min p_n(G)_100_6	911	8
2	min p_n(G)_2000_6	318	3
3	min p_n(G)_300_3	767	6
4	min p_n(G)_300_30	148	3
5	min p_n(G)_4000_6	705	4
6	min p_n(G)_pravilno	0	0
7	min p_n(G)_random_1200_6	655	2

Slika 5: Tabela odmikov za minimum

Rezultati odmikov pri iskanju grafov z minimalnim številom podpoti so bili nepričakovani. Kot je razvidno s slike 5, so se najmanjša odstopanja pojavila pri višji začetni temperaturi, medtem ko večje število korakov ni nujno prispevalo k boljšim približkom. Najmanjše število odmikov je bilo doseženo pri metodi z naključnimi začetnimi grafi, pri čemer je bil tudi relativni odmik pri tej metodi najmanjši. Največje absolutno odstopanje te metode (652) se je pojavilo pri večjem ciklomatskem številu, kjer je relativni odmik manjši, ker ima graf večje število podpoti. Najslabše približke sva dobila pri začetni temperaturi 6 in stotih korakih.

$$n = 9$$

- Za grafe z devetimi vozlišči sva definirala tri različne načine SA-ja, vse pripadajoče datoteke za so shranjene v mapi 1_del: SA_n9, koda je zapisana v datoteki 1del_SA_n_9.ipynb.
- V prvi način SA sva začetne začetne približke sva skonstruirala iz točnih vrednosti prejšnjega dela, tako da sva vzela graf z 8 vozlišč in istim ciklomatskim številom, ter mu dodala vozlišče in eno povezavo tako, da je graf ostal povezan. Ko pa sva želela obravnavati graf z ciklomatskim številom, ki je večji od 21 (to je maksimum ciklomatsko število za grafe z 8 vozlišči) sva določila da SA vzame približek ki ga je izračunal prej in doda eno "naključno" povezavo tako da se ciklomatsko število za 1 poveča. Za SA sva uporabila naslednje nastavitve: max_steps = 300, T_start = 3, T_end = 0.001, cooling = 0.99. Rezultati tega načina so zapisani v 1del_SA_n_9.csv.
- Drugi način je bil definiran na podoben način, le da sva namesto točnih približkov za grafe z osmimi vozlišči vzela približke, ki sva jih dobila ko sva SA izvedla na grafih velikost $n = 8$ pri temperaturi $T = 3$ in $N = 300$, torej sva iskala približke za grafe z največjim/najmanjšim številom podpoti iz približkov. Rezultati so zapisani v datoteki 1del_SA_n9_priblizki8.csv.
- V tretjem načinu sva namesto že izračunanih približkov, kot začetni graf za SA definirala, da koda izbere samo random graf z izvranim ciklomatskim številom in devetimi vozlišči, rezultate pa sva zapisala v 1del_SA_n9_priblizki8.csv.
- Za vsako od smeri (torej za max ali min) se je za vsako od teh treh načinov koda izvajala med 30 in 40 minut
- **UGOTOVITVE:**

2. del projektne naloge

V drugem delu projektne naloge sva se osredotočila na opazovanje oblik grafov. Fiksirala sva ciklomatsko število do $\mu = 8$ in izvajala kodo na majhnih grafih do $n = 10$, nato pa še simulirano ohlajanje (SA) na grafih z $n = 11$ do $n = 30$. Nato sva uporabila podani izrek Knorra in preverila, ali drži. Za konec sva tako za majhne grafe kot tudi za velike grafe z metodo SA izvedla analizo oblik tudi za grafe z najmanjšim oziroma največjim številom geodetskih podpoti.

Vsa koda za drugi del je zapisana v mapi `2_del`. Za geodetske podpoti pa je znotraj te mape mapa `Geodesic_subpath_number`, kjer so zapisane vse datoteke, povezane z iskanjem grafov z največjim oziroma najmanjšim številom geodetskih podpoti.

Najmanjše/največje število podpoti

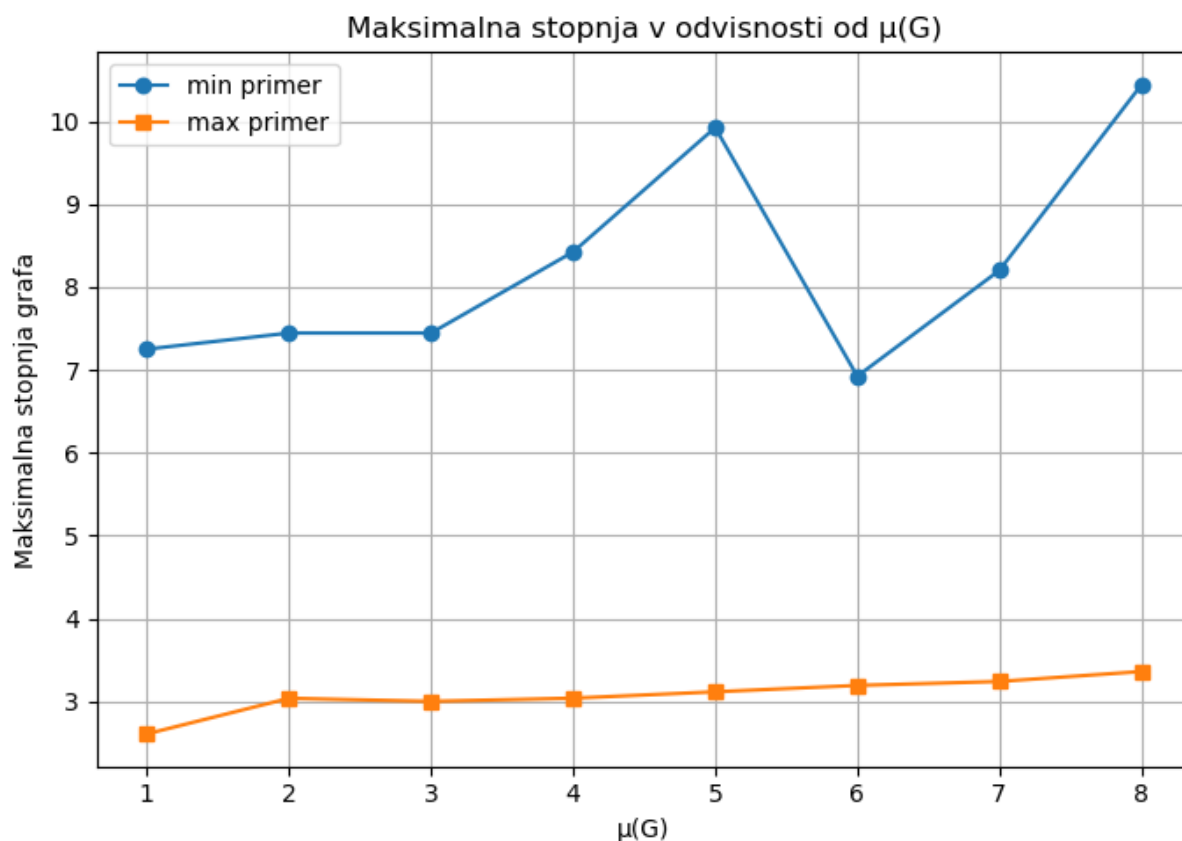
Mali grafi

- Tako kot v prvem delu sva najprej napisala kodo za majhne grafe in zanje poiskala točne rešitve. Za ciklomatska števila od $\mu = 1$ do $\mu = 8$ sva za grafe z največ 10 vozlišči generirala vse grafe ter izmed njih izbrala tiste z najmanjšim oziroma največjim številom podpoti.

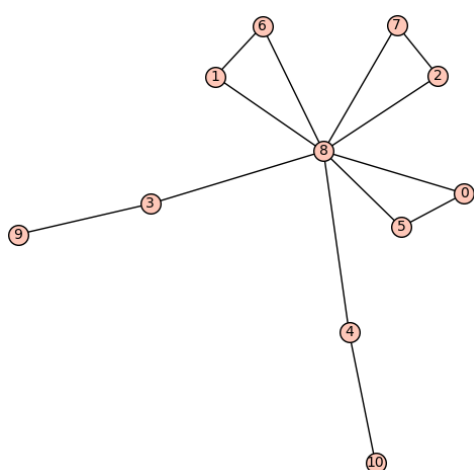
Koda je zapisana v datoteki `2del_subpath_mali_grafi.ipynb`, rezultati pa v `2del_subpath_mali_grafi.csv`.

Za lažjo analizo oblik grafov sva si vse slike v drugem delu shranjevala v mapi `slike_max` in `slike_min`, kjer ime slike vsebuje par (μ, n) , na katerega se nanaša.

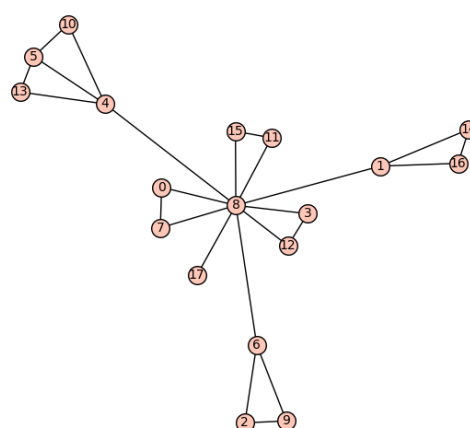
- Da se je koda iztekla hitreje, sva jo prilagodila za uporabo večih jeder.
- **Ugotovitev:**



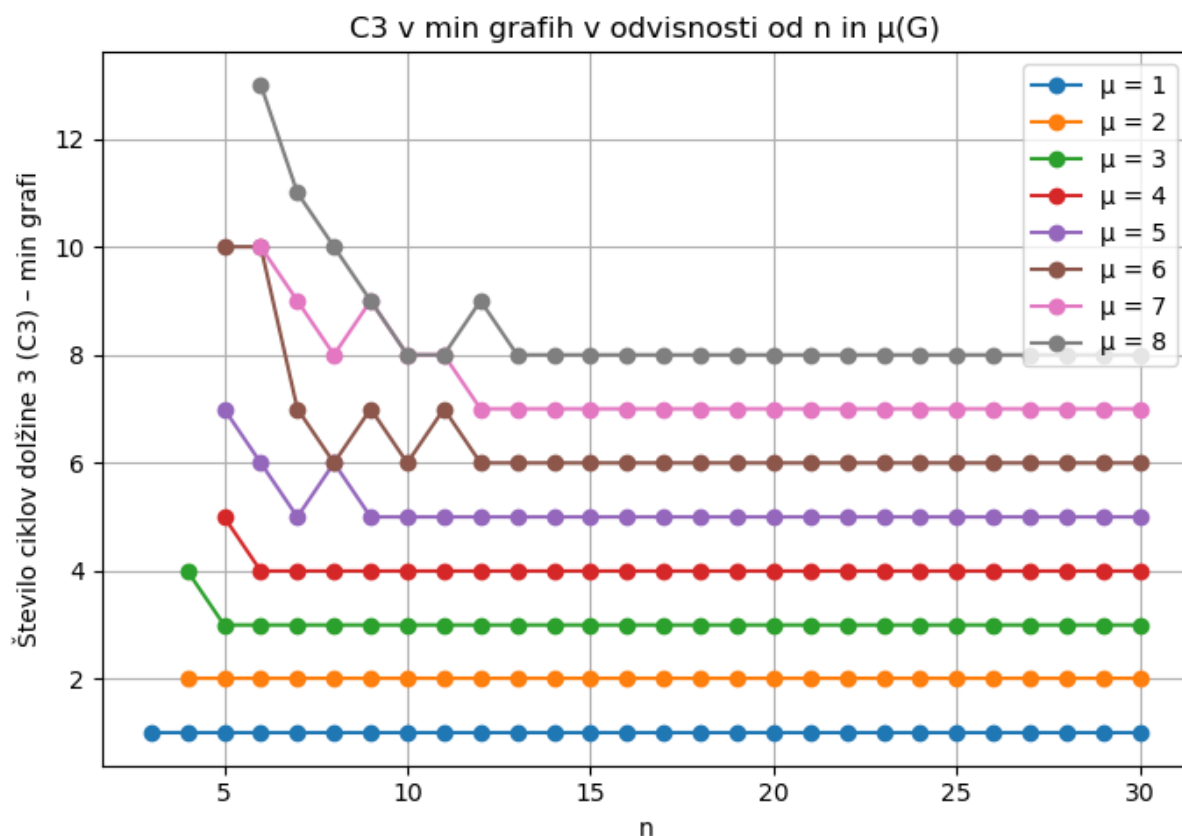
Pri grafih z majhnim številom vozlišč in majhnim ciklomatskim številom sva ugotovila, da so se pri minimiziranju števila podpoti grafi nagibali k strukturi, sestavljeni iz ciklov dolžine 3, ki so imeli skupno eno vozlišče, kar se je pojavilo v primerih, ko je bilo glede na dano ciklomatsko število na voljo dovolj vozlišč. Tako se je ob povečanju ciklomatskega števila dodal še en nov cikel dolžine 3, ko je število vozlišč dovolj veliko vrednost.



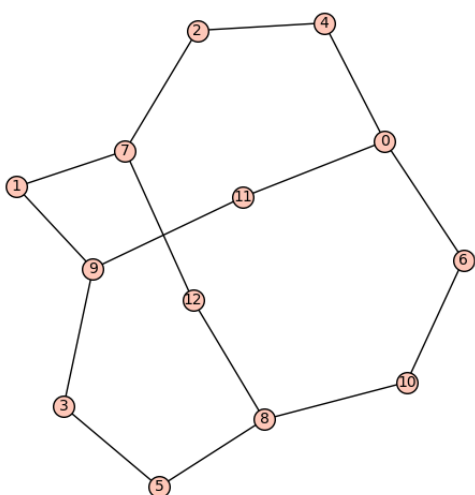
Slika 6: MIN Graf za $\mu = 3$



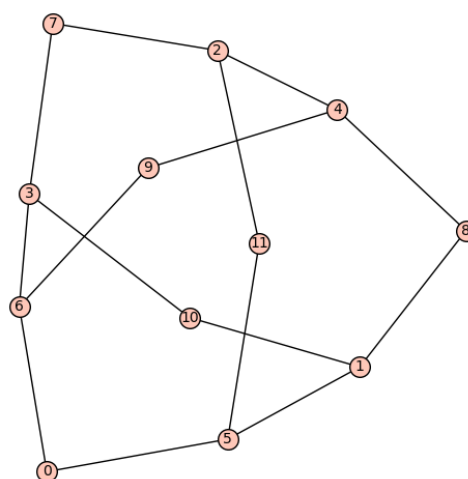
Slika 7: MIN Graf za $\mu = 7$



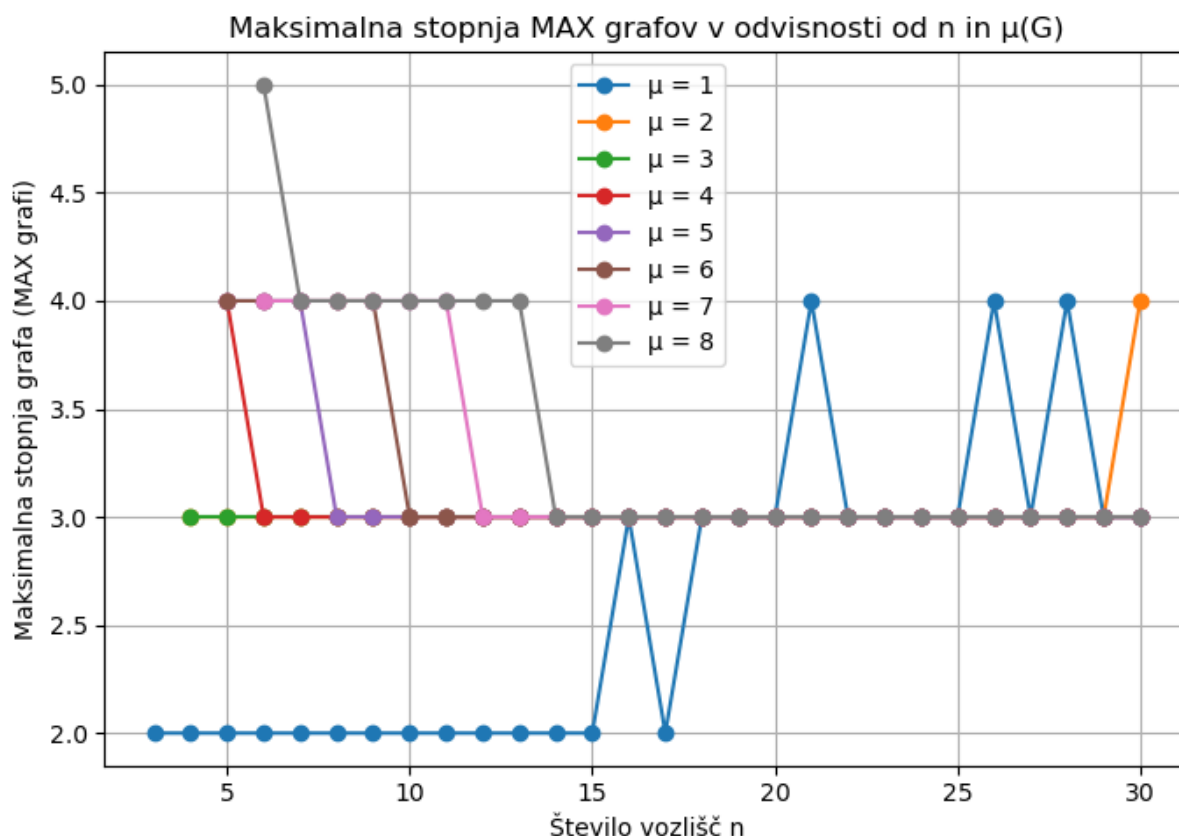
Ostala vozlišča so tvorila vsaj eno steblo, ki je v večini primerov izhajala iz istega središčnega vozlišča. Pri maksimiziranju števila podpoti sva opazila, da je imel pri dobro skonstruiranih grafih vsakemu vozlišču prirejeno vsaj stopnjo 2. To pomeni, da je bilo vsako vozlišče del nekega cikla. Poleg tega se je pri fiksnem ciklomatskem številu (izjema je primer, ko je ciklomatsko število 1) z dodajanjem novih vozlišč maksimalna stopnja grafa zmanjševala proti 3.



Slika 8: MAX Graf za $\mu = 3$



Slika 9: MAX Graf za $\mu = 4$



Veliki grafi

- Tudi v drugem delu sva za »večje grafe« uporabila metodo simulated annealing (SA). Za vsa μ od 1 do 8 sva izvedla SA za grafe z $n = 10$ do $n = 30$.
- Začetne približke sva izbrala podobno kot v prvem delu: vzela sva graf z enakim ciklomatskim številom, vendar z enim vozliščem manj, nato pa sva v ta graf dodala še eno vozlišče in eno povezavo, tako da je graf ostal povezan.
- Povezavo sva dodajala na dva različna načina. Najprej sva jo v graf dodala naključno, nato pa sva jo dodala v tisto vozlišče, ki ima največjo stopnjo.
- Koda za oba načina je zapisana v datoteki `2del_subpath_SA.ipynb`. Rezultati za naključno dodajanje povezave so zapisani v `2del_subpath_SA.csv`, za dodajanje v vozlišče z največjo stopnjo pa v `2del_subpath_SA_v_sredisce.csv`.

- **Ugotovitve za oba načina dodajanja povezave:**

Za minimum oba pristopa vrneti enako vrednost najmanjšega števila podpoti, razlikujeta pa se obliki grafov. To je smiselno, saj je število podpoti enako ne glede na to, ali so vsa vozlišča pripeta na eno »sredinsko« vozlišče ali pa so razporejena v verigo; tudi število ciklov ostane enako v obeh primerih. Na primer, pri $\mu = 3$ ima graf z minimalnim številom podpoti tri trikotne cikle, preostala vozlišča pa so razporejena v verigo.

Tudi za maksimum sta pristopa primerljiva: v nekaterih primerih ena metoda vrne nekoliko večje število podpoti, v drugih pa druga. Večjo vrednost navadno vrne pristop, ki najde graf, kjer so nova vozlišča vključena v cikel in ne kot listi, pripeti

na eno vozlišče. To obnašanje je posledica stohastične narave metode SA in ga ne moreva povsem odpraviti.

Lastnosti: V drugem delu projektne naloge sva prav tako želela preveriti ali Izrek Knor-a zares trdi. V drugem smislu nisva našla smiselne ideje kako bi imelementirata ta izrek v njihino projektno nalogo, saj sva midva iskala grafe z največjim in najmanšim številom podpoti na vseh grafi in ne samo na kakusnih. ZA vse grafe v datoteki `2del_subpath_SA.csv`, sva v datoteki `pregledovanje_lastnosti_subpath.ipynb` Zapisovala funkcije, ki za vsak `max` in vsak `min` graf ali je kaktusni in ali je PTC graf. Rezulati so zapisani v datoteki `pregledovanje_lastnosti_subpath.csv` Izrek Knor-a trdi, da v primeru da je graf z največjim številom podpoti kaktusen mora biti tudi PTC. Da sva preverila če ta implikacija res trdi sva zapisala funkcijo, ki nam glede narezultate lastnosti preveri če drži in rezultat je bil da je izrek res

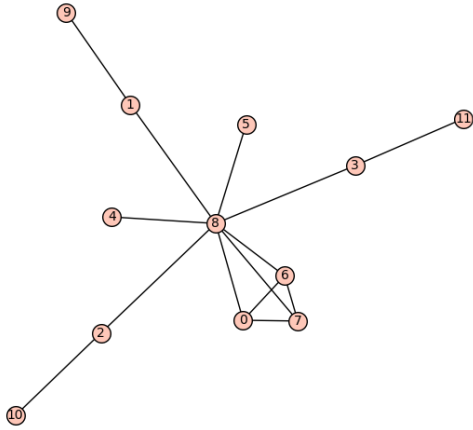
Najmanjše/največje število geodetskih podpoti

Mali grafi

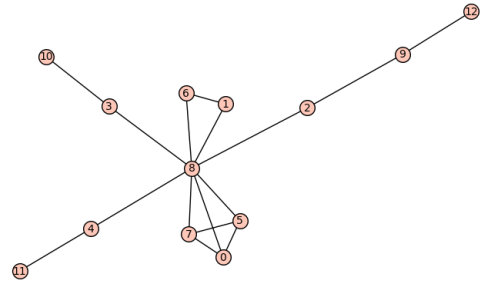
- Za geodetske podpoti sva uporabila isto kodo kot sva jo zapisala v 2.delu za majhne grafe, torej kot je zapisana v `2del_subpath_mali_grafi.ipynb`, na novo sva definirala samo funkcijo, ki išče geodetske podpoti
- koda je zapisana v datoteki `2del_geodesic_mali_garfi.ipynb.ipynb`. Rezultati so zapisani v `2del_geodesic_mali_garfi.csv` Slike pa so shranjene v mapah `slike_geodesic_max` in `slike_geodesic_min`
- kodo sva poglajala za ciklomatska števila od $\mu = 1$ do $\mu = 8$ sva za grafe z največ 10 vozlišči

Veliki grafi

- za velike grafe pri iskanju geodetskih poti sva prav tako uporabila isto kodo kot v 2.delu za isanje podpoti, le da sedaj koda išče najmanjše število geodetskih podpoti
- koda je zapisana v datoteki `2del_geodesic_SA.ipynb.ipynb`. Rezultati so zapisani v `2del_geodesic_SA.csv` Slike pa so shranjene v mapah `slike_geodesic_max` in `slike_geodesic_min`
- za geodetske podpoti sva začetnim približkom povezala samo randomy in ne v vozlišče z največjo stopnjo.
- **ugotovitve**
Pri minimiziranju geodetskega števila poti v grafih se je pri $\mu \geq 3$ začelo pojavljati, da grafi niso več vsebovali zgolj ciklov dolžine 3, temveč so začeli tvoriti strukture, ki jih lahko opišemo kot nekakšne piramide v prostoru. Pri $\mu = 3$ ima graf eno tristrano piramido in eno centralno vozlišče. Pri $\mu = 4$ se na to centralno vozlišče doda še en cikel dolžine 3, pri $\mu = 5$ pa še en tak cikel, dokler se pri $\mu = 6$ ne ustvari ena štiristrana piramida.

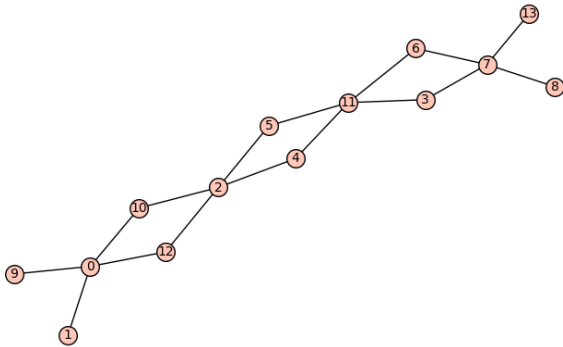


Slika 10: Graf za $\mu = 3$

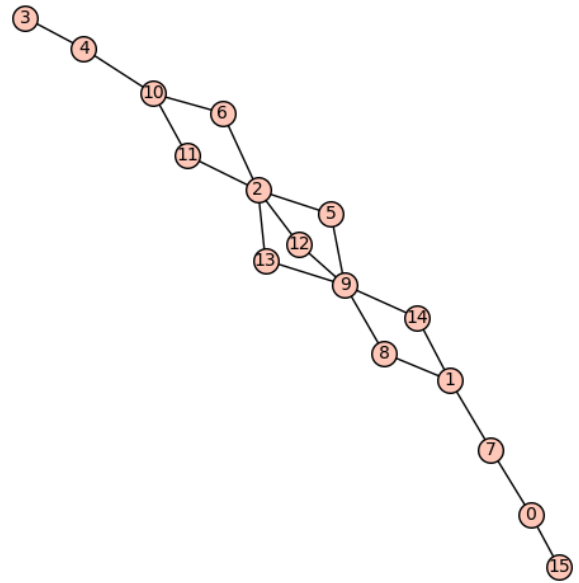


Slika 11: Graf za $\mu = 4$

Pri nadaljnjem povečevanju ciklomatskega števila se vzorec nadaljuje podobno: najprej nastajajo novi cikli dolžine 3, dokler se ponovno ne oblikuje nova piramidna struktura. Pri maksimiziranju sva ugotovila, da se v odvisnosti od geodetskega števila oblikujejo grafi, v katerih so cikli povezani zaporedno in skupaj tvorijo verigo.



Slika 12: MAX graf za $\mu = 3$



Slika 13: MAX graf za $\mu = 4$