

Število podpoti pri grafih z danim ciklomatskim številom

Poročilo projektne naloge pri predmetu finančni praktikum
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 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 podrobnejše 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

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

- **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 enim 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 ohljanje

Za velike grafe sva tako v 1. delu kot v 2. delu uporabljala Simulated annealing (SA). To je optimizacijski algoritem, ki se uporablja za iskanje približkov 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 algoritom 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 želeta preveriti, kateri način simuliranega ohlajanja je najučinkovitejši. Za dano število vozlišč in vsa pripadajoča ciklomatska števila sva poiskala 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

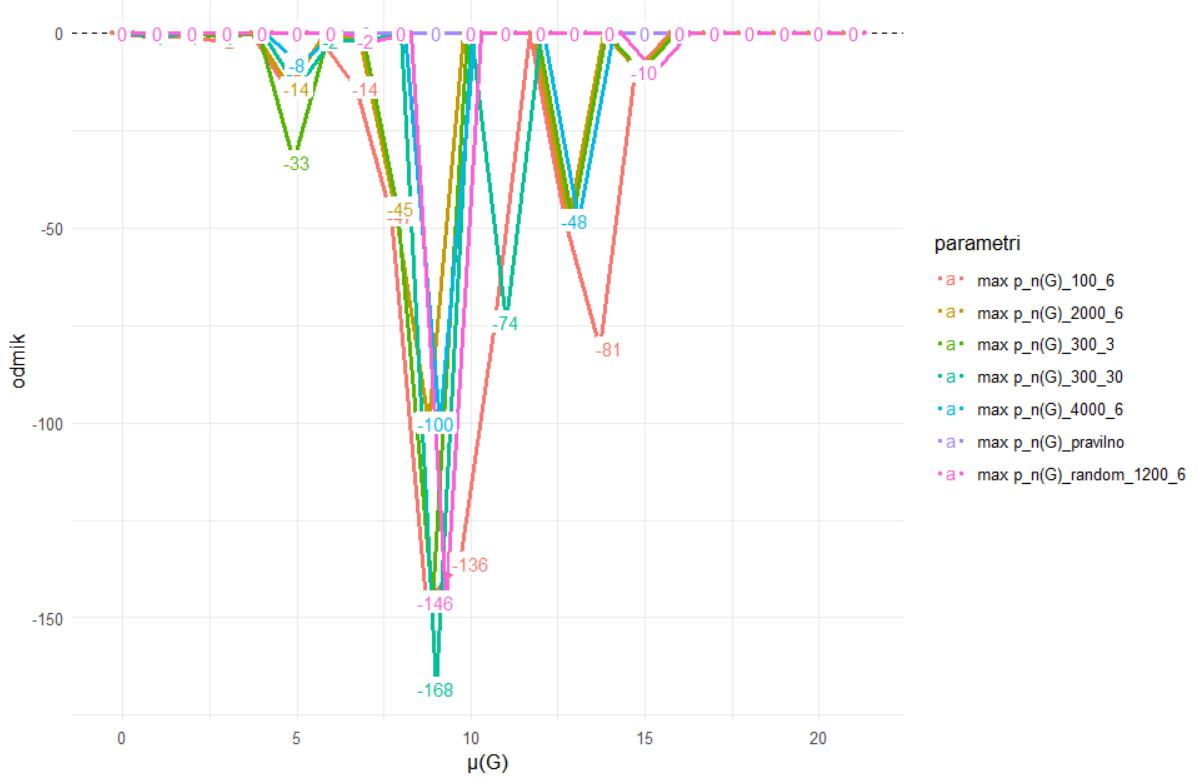
- Vse datoteke za ta del so shranjene v mapi `1_del`; koda je zapisana v datoteki `1del_mali_grafi.ipynb`, rezultati pa v `1del_mali.csv`.
- Najprej sva definirala funkcijo, ki v podanem grafu poišče vse podpoti. To funkcijo sva nadgradila tako, da za dano ciklomatsko število in število vozlišč najprej 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 nastavitevah 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.

$$n = 8$$

- Na grafih z osmimi vozlišči 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 z naključnimi začetnimi grafi, torej sva kot začetni približek simuliranega ohlajanja vzela naključen graf z izbranim ciklomatskim številom in osmimi vozlišči.
- 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:**

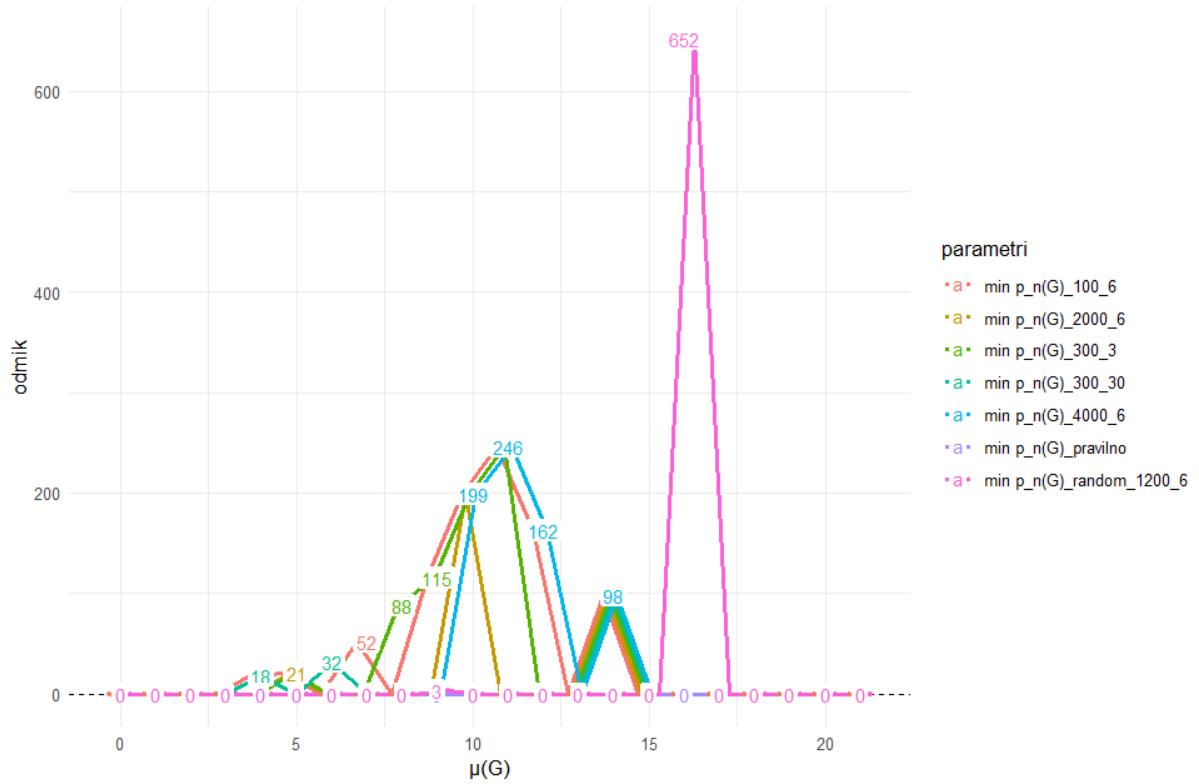


Slika 1: Graf odmikov pri različnih parametrih SA za maksimalno število podpoti v grafih z $n = 8$

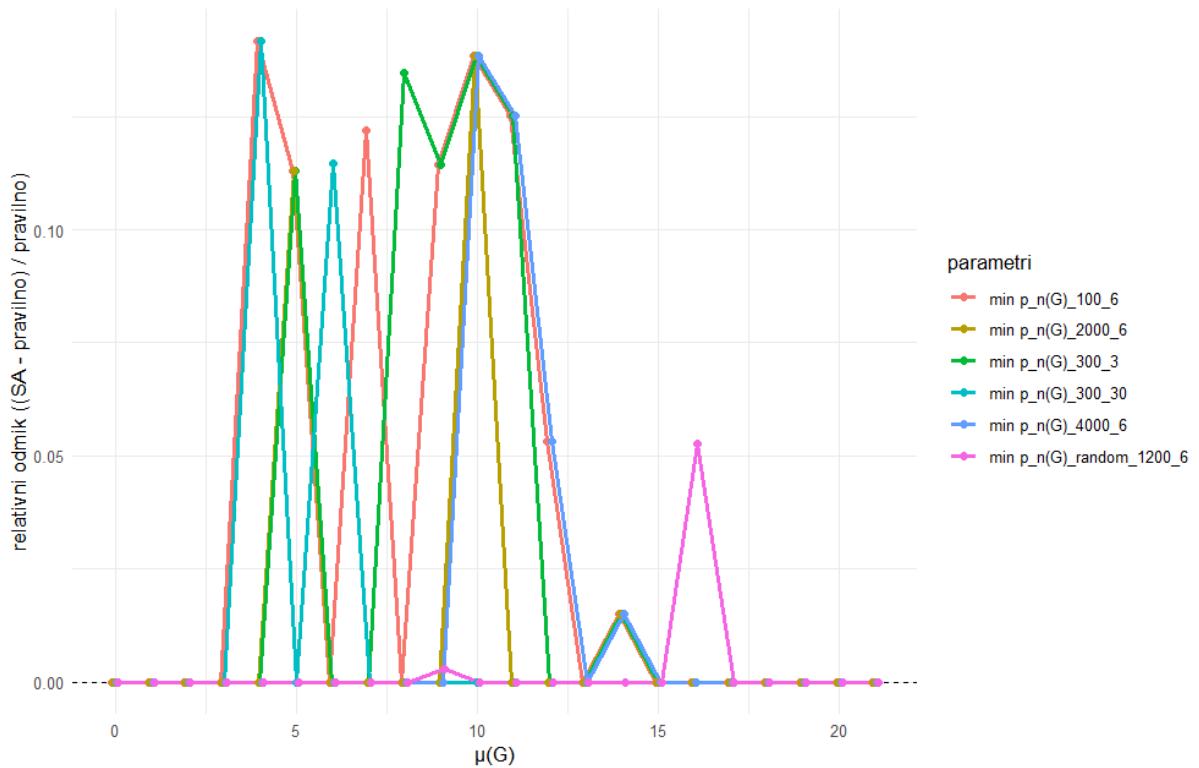
Parametri	Vsota odmikov	Število odmikov
max p_n(G)_100_6	-576	13
max p_n(G)_2000_6	-219	6
max p_n(G)_300_3	-283	6
max p_n(G)_300_30	-261	6
max p_n(G)_4000_6	-156	3
max p_n(G)_pravilno	0	0
max p_n(G)_random_1200_6	-158	3

Tabela 1: Povzetek odmikov za maksimalno število podpoti

Iz zgornjih grafov za maksimalno število podpoti lahko opazimo, da sta najmanjša odstopanja dosegla algoritom SA, inicializiran z naključnimi grafi, ter algoritom 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 pridelala do pravilne vrednosti. Kot sva pričakovala, je algoritom SA z najmanjšim številom korakov vrnil najslabše približke. Poleg tega sva ugotovila, da povišanje začetne temperature ni bistveno izboljšalo iskanja grafa z največjim številom podpoti.



Slika 2: Graf odmikov pri različnih parametrih SA za minimalno število podpoti v grafih z $n = 8$



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

Parametri	Vsota odmikov	Število odmikov
min p_n(G)_100_6	911	8
min p_n(G)_2000_6	318	3
min p_n(G)_300_3	767	6
min p_n(G)_300_30	148	3
min p_n(G)_4000_6	705	4
min p_n(G)_pravilno	0	0
min p_n(G)_random_1200_6	655	2

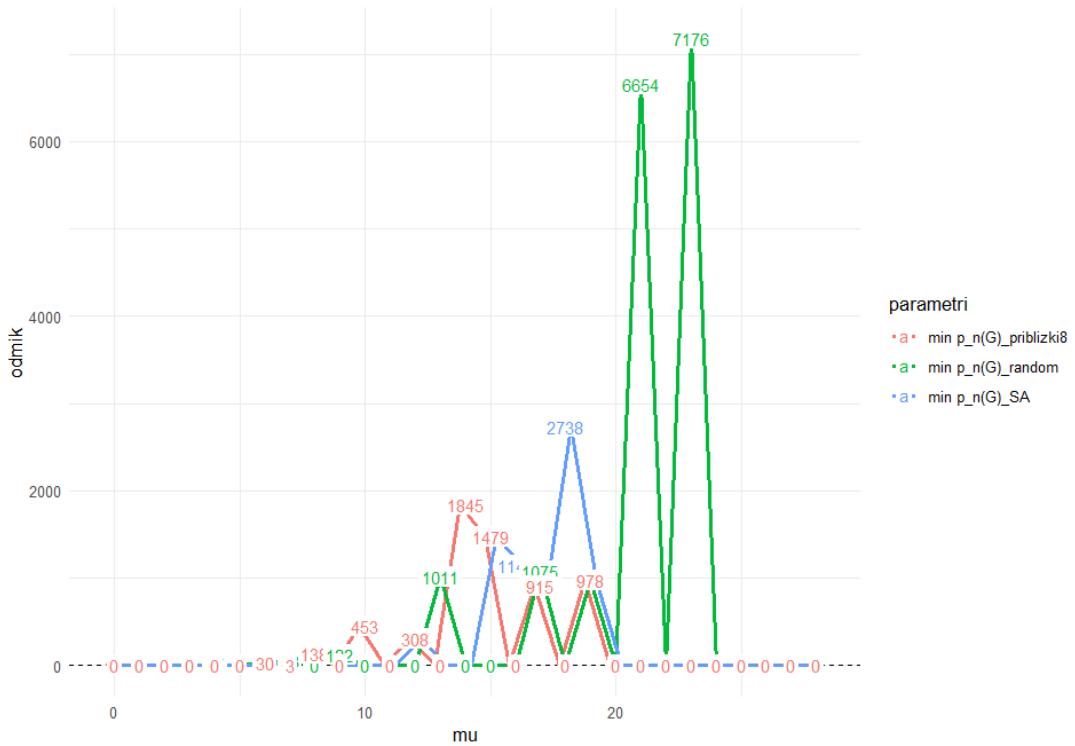
Tabela 2: Povzetek odmikov za minimalno število podpoti

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.

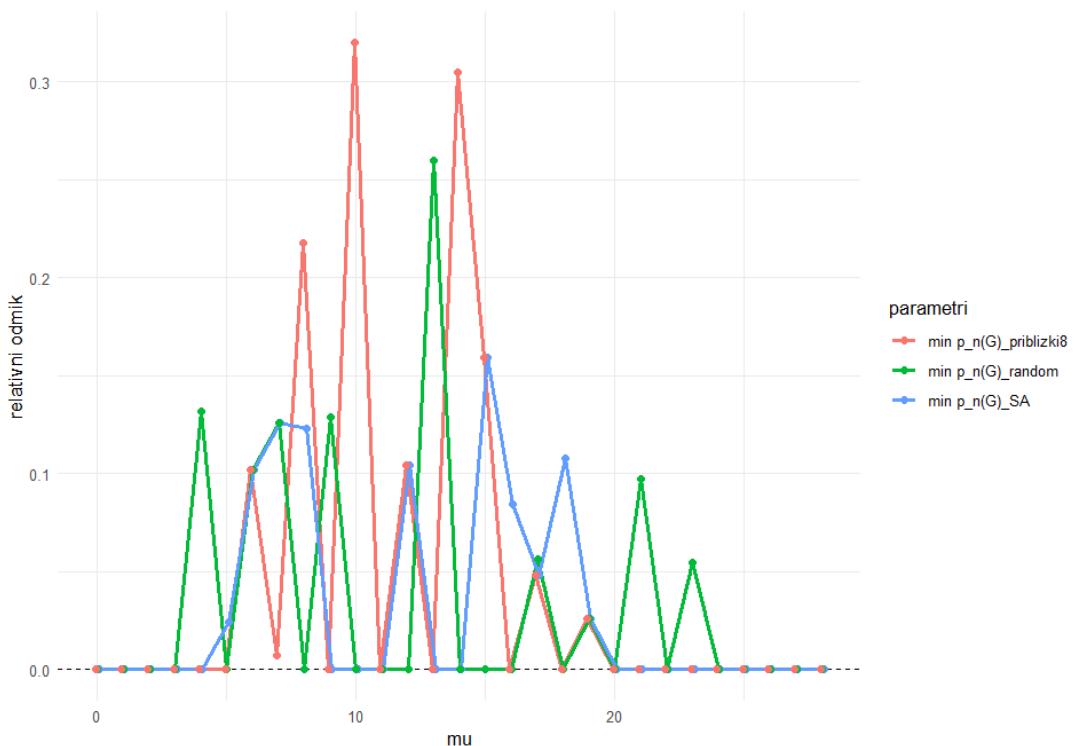
$n = 9$

- Za grafe z devetimi vozlišči sva definirala tri različne načine algoritma SA. Vse pripadajoče datoteke so shranjene v mapi `1_de1/SA_n9`, koda pa je zapisana v datoteki `1de1_SA_n_9.ipynb`.
- V prvem načinu algoritma SA na grafih z devetimi vozlišči sva začetne približke skonstruirala iz točnih vrednosti. Vzeli sva graf z osmimi vozlišči in enakim ciklomatskim številom ter mu dodali vozlišče in eno povezavo, tako da je graf ostal povezan. Kadar pa sva obravnavali graf s ciklomatskim številom, večjim od 21 (to je maksimalno ciklomatsko število za grafe z osmimi vozlišči), sva določili, da SA kot začetni približek vzame rešitev, izračunano v prejšnjem koraku, in ji doda eno naključno povezavo, s čimer se ciklomatsko število poveča za 1. Za algoritmom SA sva uporabili naslednje nastavitev: `max_steps = 300, T_start = 3, T_end = 0.001, cooling = 0.99`. Rezultati tega načina so zapisani v datoteki `1de1_SA_n_9.csv`.
- Drugi način je bil definiran podobno kot prvi, razlika pa je bila v tem, da sva namesto točnih približkov za grafe z osmimi vozlišči uporabili približke, ki sva jih dobili z algoritmom SA v prejšnjem delu, pri temperaturi $T = 3$ in $N = 300$. Tako sva iskala približke za grafe z največjim oziroma najmanjšim številom podpoti na podlagi že izračunanih približkov. Rezultati so zapisani v datoteki `1de1_SA_n9_priblizki8.csv`.
- V tretjem načinu pa sva namesto "dobrih" približkov, kot v prejšnjih dveh primerih, uporabili zgolj naključni graf z izbranim ciklomatskim številom in devetimi vozlišči. Rezultate sva zapisali v datoteko `1de1_SA_n9_priblizki8.csv`.
- Za vsako izmed smeri (torej za maksimum ali minimum) se je koda za vsakega izmed treh načinov izvajala med 30 in 40 minut.

- UGOTOVITVE:



Slika 4: Graf odmikov pri različnih začetih grafih SA za minimalno število podpoti v grafih z devetimi vozlišči

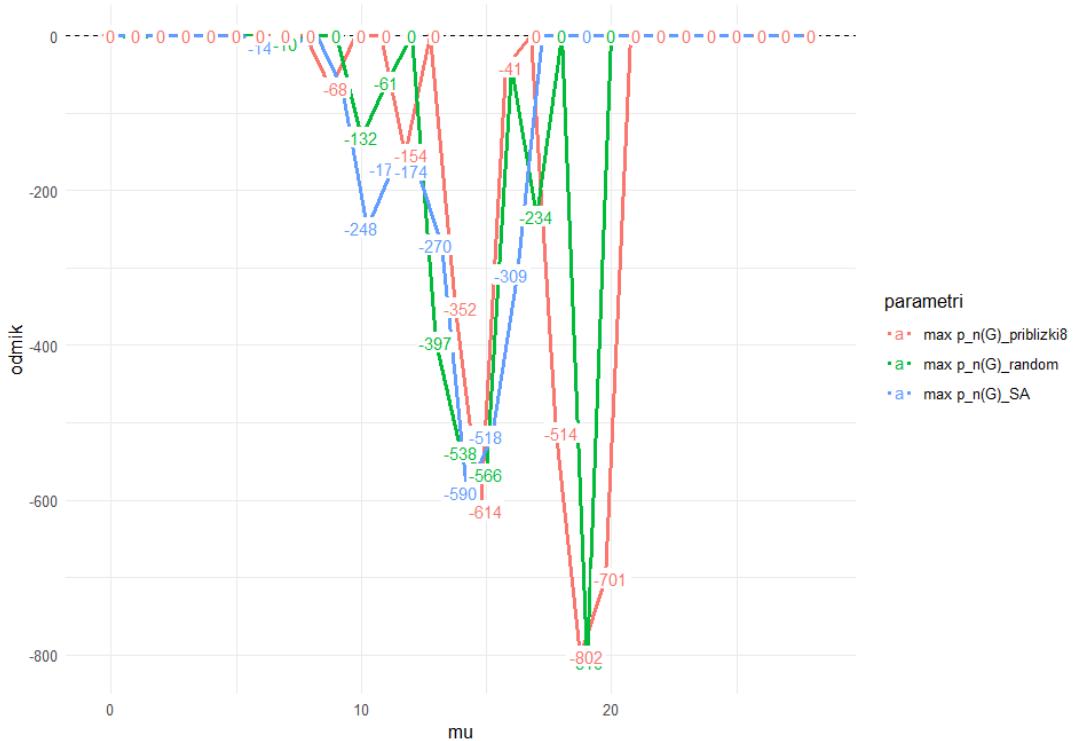


Slika 5: Graf relativnih odmikov pri različnih začetih grafih SA za minimalno število podpoti v grafih z devetimi vozlišči

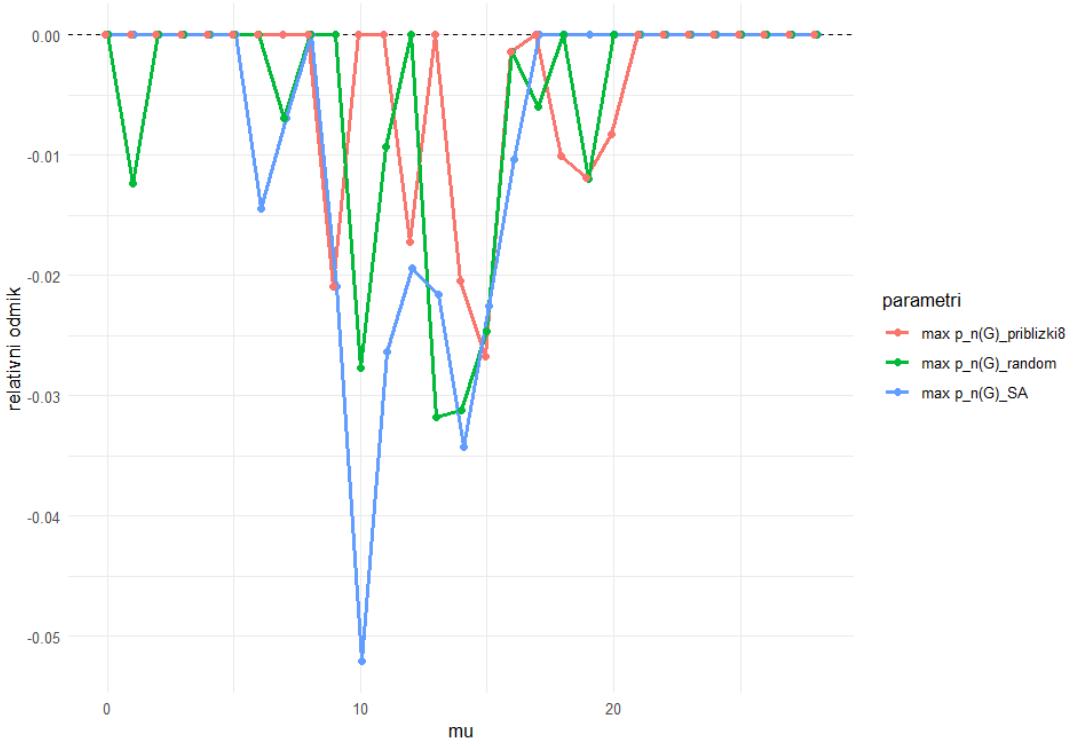
Parametri	Vsota odmikov	Vsota relativnih odmikov	Število odmikov
min p_n(G)_priblizki8	6149	1.2887322	9
min p_n(G)_random	17116	0.9820783	9
min p_n(G)_SA	7730	0.9046324	10

Tabela 3: Povzetek odmikov za minimalno število podpoti

Kot pričakovano je algoritem z naključnimi začetnimi grafi dosegel najslabše približke pri iskanju grafov z minimalnim številom podpoti. Večja odstopanja so se pojavila, zlasti ko je bilo ciklomatsko število večje od 25. Nepričakovano je imel najmanjše absolutne odmike drugi način (priblžki8). Kljub temu pa je imel ta način največje relativne odmike, saj so se ti pojavljali predvsem pri manjših ciklomatskih številih. Na podlagi izračunanih odmikov, se je zato prvi način izkazal kot najustreznejši, saj ima tako v absolutnem kot tudi v relativnem smislu najmanjše napake.



Slika 6: Graf odmikov pri različnih začetih grafih SA za maksimalno število podpoti v grafih z devetimi vozlišči



Slika 7: Graf relativnih odmikov pri različnih začetih grafih SA za maksimalno število podpoti v grafihi z devetimi vozlišči

Parametri	Vsota odmikov	Vsota relativnih odmikov	Število odmikov
max p_n(G)_priblizki8	-3246	-0.1169827	8
max p_n(G)_random	-2790	-0.1635572	10
max p_n(G)_SA	-2373	-0.2292007	10

Tabela 4: Povzetek odmikov za maksimalno število podpoti

Vsi trije načini so pri iskanju maksimuma dosegli bistveno manjše relativne odmike kot pri iskanju minimuma. Pri maksimumu je bil največji relativni odmik približno 5 %, medtem ko so se pri minimumu pojavljali relativni odmiki, večji od 30 %. Poleg tega so si bile uporabljene metode tudi med seboj bližje po uspešnosti. Najmanjše absolutne odmike je imel prvi način, vendar pa je imel hkrati tudi največje relativne odmike, saj so se napake pojavljale predvsem pri manjših ciklomatskih številah.

2. del projektne naloge

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

Vsa koda za drugi del je shranjena v mapi `2_del`. Dokumenti, povezani z geodetskim številom, pa se nahajajo v podmapi `Geodesic_subpath_number`.

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.

Veliki grafi

- Tudi v drugem delu sva za »večje grafe« uporabila metodo simuliranega ohlajanja. 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:**

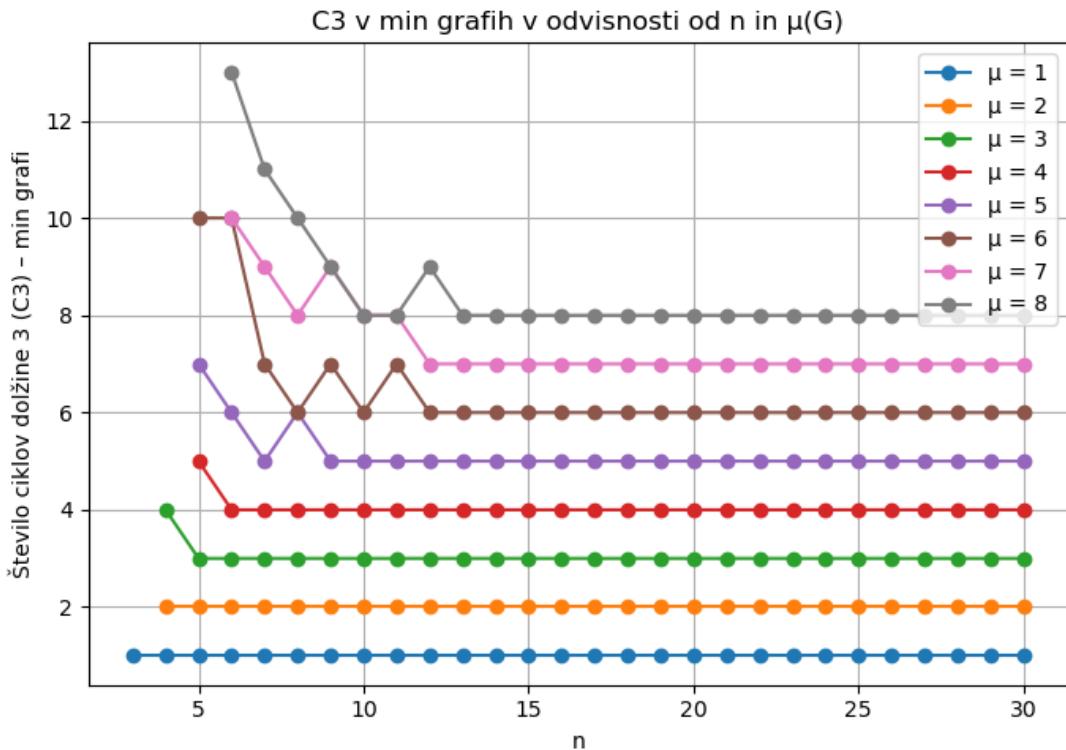
Za minimum oba pristopa vrneta enako vrednost najmanjšega števila podpoti, razlikujeta pa se obliki grafov. To je smiselno, saj je število podpoti enako ne glede na to, ali so vsa vozlišča pripeta na eno »sredinsko« vozlišče ali pa so razporejena v verigo; tudi število ciklov ostane enako v obeh primerih.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

razporejena v verigo.

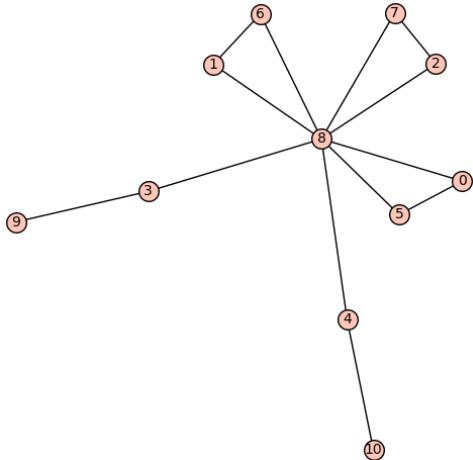
Tudi za maksimum sta pristopa primerljiva: v nekaterih primerih ena metoda vrne nekoliko večje število podpoti, v drugih pa druga. Večjo vrednost navadno vrne pristop, ki najde graf, kjer so nova vozlišča vključena v cikel in ne kot listi, pripeti na eno vozlišče. To obnašanje je posledica stohastične narave metode SA in ga ne moreva povsem odpraviti.

Ugotovitve:

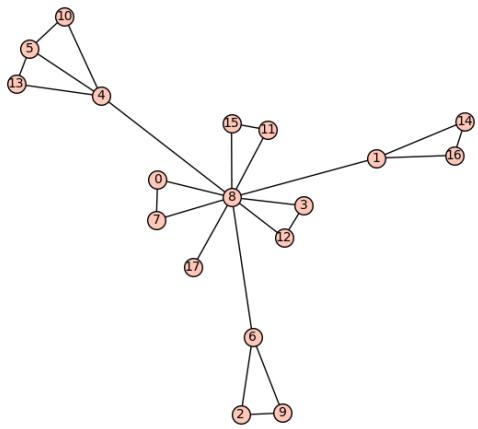
Pri grafih z majhnim številom vozlišč in majhnim ciklomatskim številom sva ugotovila, da so se pri minimiziranju števila podpoti grafi nagibali k strukturi, sestavljeni iz ciklov dolžine 3, kar je razvidno tudi iz slik 8 in 9. Cikli dolžine tri so večinoma imeli eno skupno vozlišče. To se je pojavilo v primerih, ko je bil $\mu \leq 5$ in je bilo na voljo dovolj vozlišč, da so se taki cikli sploh lahko formirali. Ob povečanju ciklomatskega števila sva opazila, da se je (ob dovolj velikem številu vozlišč) dodal samo še en nov cikel dolžine 3. Ostala vozlišča so tvorila eno ali več stebl, ki so v večini primerov izhajala iz istega središčnega vozlišča. Pri $\mu \geq 6$ pa so se začele tvoriti nekoliko drgačne strukture, kjer pa niso več vsi cikli dolžine 3 povezani z skupnim vozliščem. Nastalo je več skupkov ciklov dolžine 3, ki so bili povezani na centralno vozlišče.



Slika 8: Števila ciklov dolžine 3 za minimum glede na μ in n

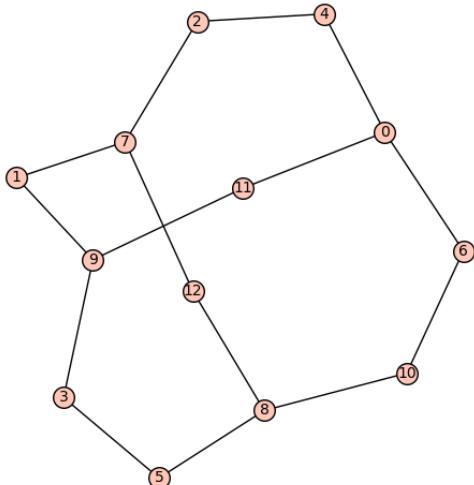


Slika 9: MIN Graf za $\mu = 3$ in $n = 11$

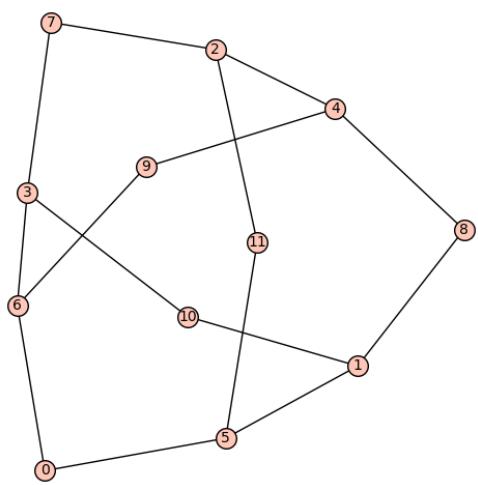


Slika 10: MIN Graf za $\mu = 7$ in $n = 18$

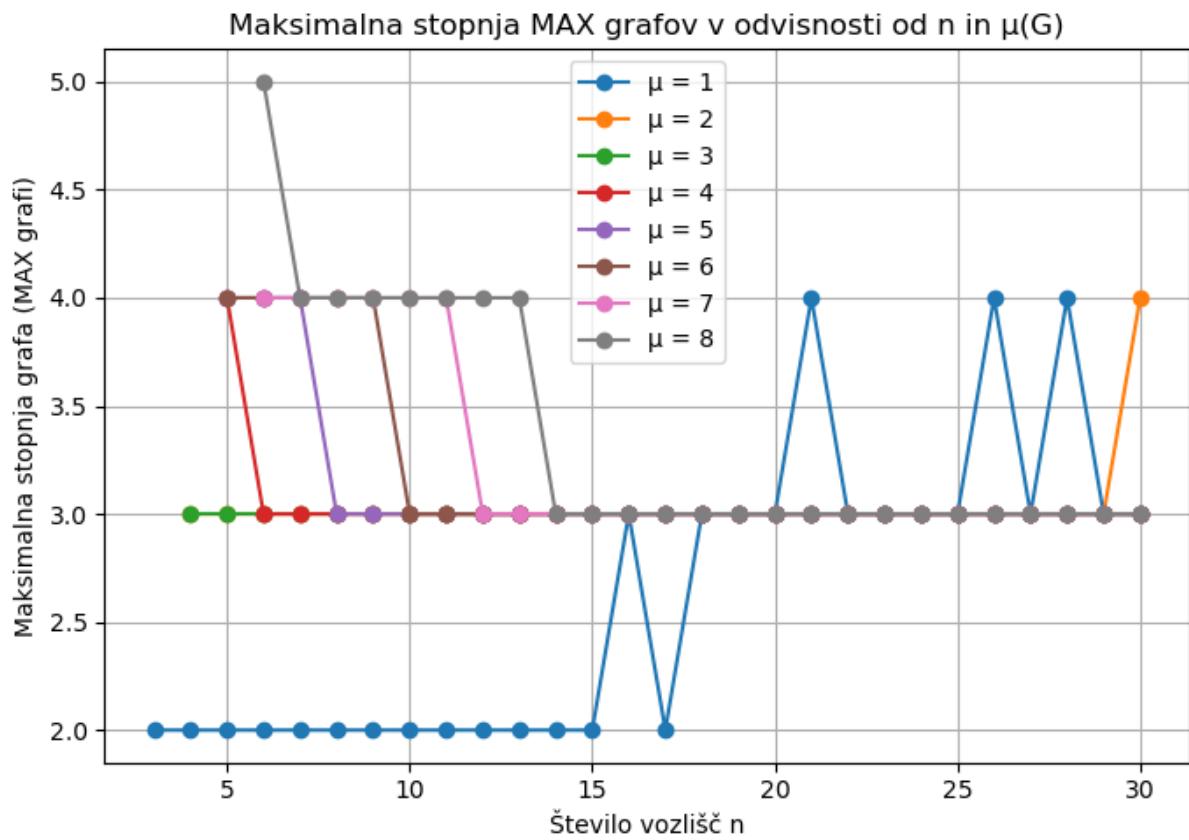
Pri maksimiziranju števila podpoti sva opazila, da je imel pri dobro skonstruiranih grafih vsako vozlišče vsaj stopnjo 2. To pomeni, da je bilo vsako vozlišče del nekega cikla. Poleg tega se je pri fiksном ciklomatskem številu (izjema je primer, ko je ciklomatsko število 1) z dodajanjem novih vozlišč maksimalna stopnja grafa zmanjševala proti 3, kot je razvidno iz slike 13. Razlog za izstopanje pri $\mu = 1$ se skriva v slabših optimumih, ki nama jih je zgradil SA. Dokaj očitno je, da bi optimalnih grafi morali tvoriti natanko en cikel maksimalne dolžine.



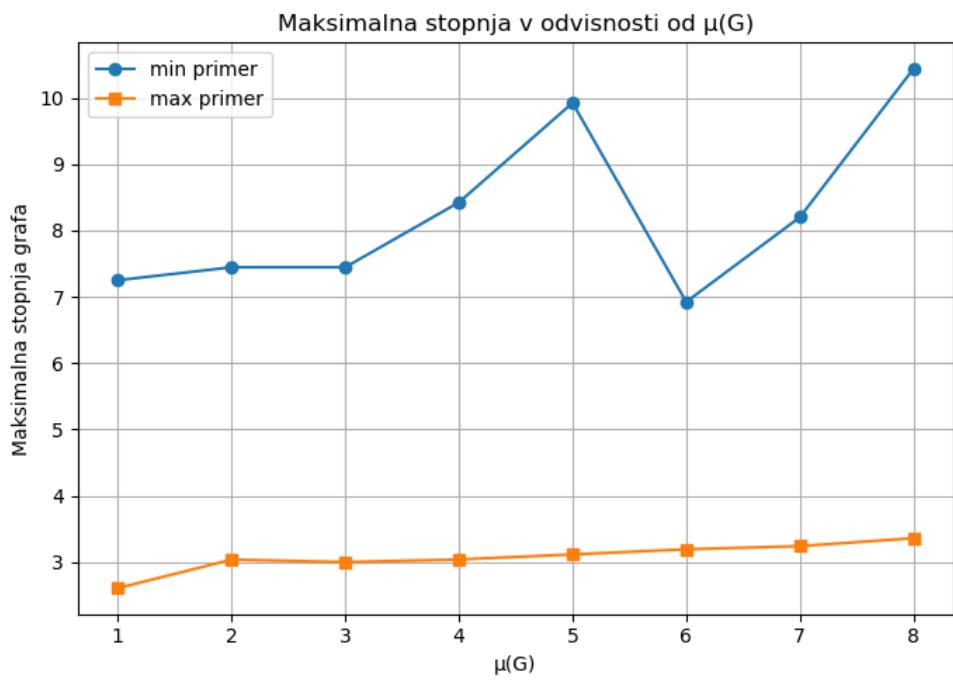
Slika 11: MAX Graf za $\mu = 3$



Slika 12: MAX Graf za $\mu = 4$



Slika 13: Maksimalna stopnja grafov v odvisnosti od n



Slika 14: Povprečna maksimalna stopnja grafov pri fiksnem μ

Izrek Knor-a:

V drugem delu projektne naloge sva želeta preveriti veljavnost izreka, saj je to predstavljalo edini smiseln način, kako ta izrek vključiti v najino projektno naložo. V datoteki `pregledovanje_lastnosti_subpath.ipynb` sva zapisala kodo, ki za vse obravnavane grafe preveri, ali so kaktusni ali PTC-grafi. Dobljene rezultate sva shranila v datoteko `pregledovanje_lastnosti_subpath.csv`. Izrek Knor-a trdi, da mora biti graf z največjim številom podpoti PTC, če je hkrati kaktusen. Da bi preverila, ali ta implikacija dejansko velja, sva zapisala funkcijo, ki na podlagi izračunanih lastnosti preveri izpolnjnost pogoja. Rezultati so pokazali, da izrek v obravnavanih primerih drži.

Najmanjše/največje število geodetskih podpoti

Mali grafi

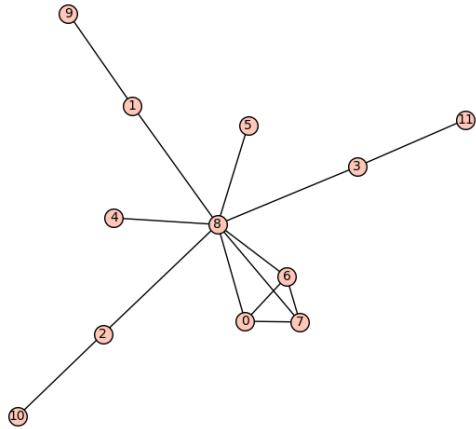
- Za geodetske podpoti sva uporabila isto kodo, kot je zapisana v datoteki `2del_subpath_mali_grafi.ipynb`, pri čemer sva na novo definirala le funkcijo za iskanje geodetskih podpoti. Kot sva že omenila sva obravnala grafe za ciklomatska števila od $\mu = 1$ do $\mu = 8$ in z največ desetimi vozlišči.
- Koda je zapisana v datoteki `2del_geodesic_mali_grafi.ipynb`, rezultati so shranjeni v datoteki `2del_geodesic_mali_grafi.csv`, slike pa so shranjene v mapah `slike_geodesic_max` in `slike_geodesic_min`.

Veliki grafi

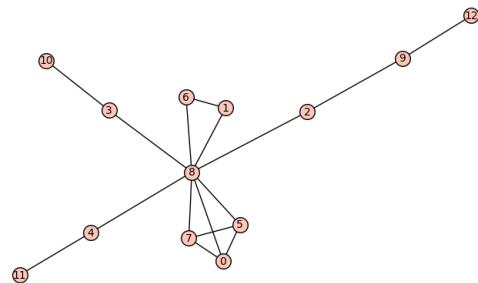
- Pri velikih grafih sva za geodetske podpoti prav tako uporabila isto kodo kot v prejšnjem primeru za velike grafe, le da sedaj koda išče najmanjše število geodetskih podpoti.
- Koda je zapisana v datoteki `2del_geodesic_SA.ipynb`, rezultati so shranjeni v datoteki `2del_geodesic_SA.csv`, slike pa so shranjene v mapah `slike_geodesic_max` in `slike_geodesic_min`.

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. Pri nadalnjem povečevanju ciklomatskega števila se vzorec nadaljuje podobno: najprej nastajajo novi cikli dolžine 3, dokler se ponovno ne oblikuje nova piramidna struktura.

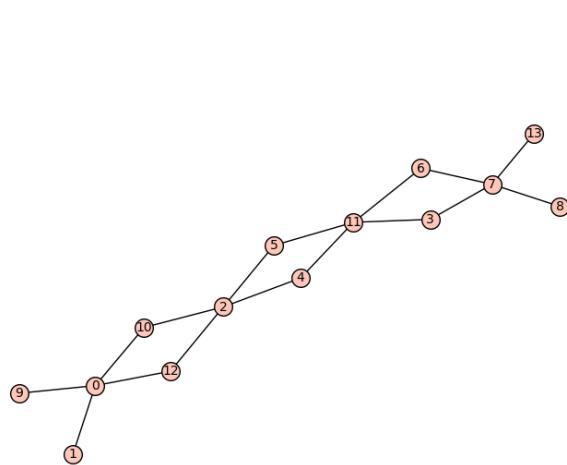


Slika 15: MIN graf za $\mu = 3$ in $n = 12$

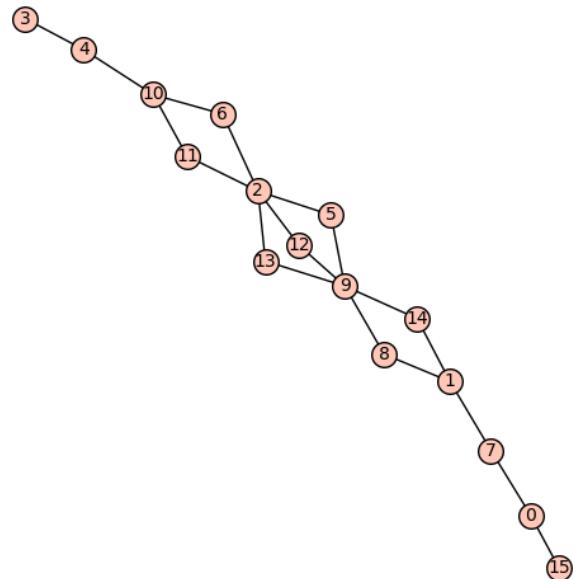


Slika 16: MIN graf za $\mu = 4$ in $n = 13$

Pri maksimiziranju sva ugotovila, da se v veliko primerih oblikejo grafi, v katerih so cikli dolžine štiri povezani zaporedno in skupaj tvorijo verigo, kar je razvidno tudi iz slik 17 in 18.



Slika 17: MAX graf za $\mu = 3$ in $n = 14$



Slika 18: MAX graf za $\mu = 4$ in $n = 16$