Problem največjega polnega podgrafa

Kratko poročilo
Vida Maver, Anamarija Mijatović

1. Opis problema

V projektu se bova ukvarjali z optimizacijskim problemom iskanja največjega polnega podgrafa oziroma z drugimi besedami, s problemom iskanja maksimalne klike v danem grafu. Polni graf reda n je graf z n vozlišči, v katerem obstaja povezava med vsemi paroma različnimi vozlišči. Označimo ga s K_n. V polnem grafu je $\frac{n(n-1)}{2}$ povezav. Ker gre za problem iskanja največjega polnega podgrafa, je najprej podana definicija polnega podgrafa oziroma klike, nato pa še definicija največje klike:

Definicija 1.1. *Naj bo V množica vozlišč grafa G(V, E). Polni podgraf grafa G je taka množica G' $\subset$ G, da za vsak par vozlišč v, w $\in$ V, kjer v $\neq$ w, velja, da sta na grafu G sosednja.*

Definicija 1.2. *Klika C na grafu G je največja (ang. maximal), če ne obstaja klika C' na grafu G, tako da C $\subseteq$ C' in C $\neq$ C'.*

Največja klika je torej polni podgraf, ki ni vsebovan v nobenem drugem polnem podgrafu.

Poleg pojma največje klike poznamo tudi pojem maksimalne klike:

Definicija 1.3. *Kliki C_{max} rečemo maksimalna (ang. maximum) klika, če ne obstaja klika na grafu G, ki bi vsebovala več vozlišč. Število vozlišč v maksimalni kliki označimo z $\omega(G)$ in ga imenujemo klično število.*

Velja, da je vsaka maksimalna klika $C_{max} \subseteq V$ na grafu G največja klika.

Definicija maksimalne klike je tesno povezana z definicijo maksimalne neodvisne množice. Neodvisna množica grafa G je ekvivalentna maksimalni kliki v komplementu grafa G^C. Spomnimo se definicije komplementa grafa:

Definicija 1.4. *Komplement grafa $G = (V, E)$ je graf $\overline{G} = (V, \overline{E})$, kjer je $\overline{E} = \{(i, j) | i, j \in V, i \neq j \text{ in } (i, j) \notin E\}$.*

Optimizacijski problem največjega polnega podgrafa spada v razred NP - težkih problemov. Problem je NP - težek, če zanj ne obstajajo učinkoviti polinomski algoritmi. Maksimalno kliko bi lahko iskali s sistematičnim pregledom vseh podmnožic, vendar takšno iskanje vzame preveč časa in ni uporabno za grafe z večjim številom vozlišč. Na srečo poznamo algoritme, ki potrebujejo manj časa za reševanje problema. Ker je problem maksimalne klike zelo težak, bova na izbranem grafu iskali maksimalno kliko neke omejene velikosti.

Pri problemu iskanja maksimalne klike si lahko pomagamo z več formulacijami v celoštevilskem linearnem programiranju. Ena izmed teh je na primer formulacija za iskanje maksimalne neodvisne množice grafa G. Ker pa je le-ta ekvivalentna iskanju maksimalne klike v komplementu grafa G^C, lahko to formulacijo uporabimo na komplementu grafa G.

Reševanja se lahko lotimo tudi z reševanjem linearne relaksacije, vendar pa ima v večini primerov zelo malo spremenljivk celoštevilske vrednosti v optimalni rešitvi relaksacije in bi reševanje na ta način lahko vodilo v napačne rezultate.

Eden izmed algoritmov za iskanje največjega polnega podgrafa v neusmerjenem grafu je MaxClique algoritem, ki najde maksimalno kliko grafa, ki je neke omejene velikosti. Za uporabo tega algoritma je koristno poznati čimbolj natančne meje, zato je ideja algoritma ta, da bi bile tesne zgornje meje že izračunane v začetku algoritma, saj se s tem skrajša čas potreben za rešitev problema.

Za primer kako je iskanje klik uporabno, si zamislimo socialno omrežje, kjer vozlišča grafa predstavljajo ljudi, povezave grafa pa njihova skupna prijateljstva. V tem primeru klika predstavlja podmnožico ljudi, ki se med seboj vsi poznajo, algoritem za iskanje klik pa lahko uporabimo pri iskanju teh skupin skupnih prijateljstev. Problem iskanja maksimalnih klik se veliko uporablja tudi v bioinformatiki in računalniški kemiji.

Reševanja problema iskanja maksimalne klike se bova lotili v programskem jeziku Sage in sicer na konkretnem grafu. Ukvarjali se bova z iskanjem največjega polnega podgrafa v neusmerjenih grafih. Za konkretni primer sva si izbrali glasovanje znotraj Wikipedijinga omrežja. Če želi uporabnik postati Wikipedijin administrator, ga mora izvoliti zadostno število uporabnikov. Na predlagani spletni strani s strani profesorja sva pridobili podatke za 2794 različnih tako imenovanih volitev, s 103.663 glasovi in 7066 uporabniki, ki so sodelovali v glasovanju (ali so glasovali, ali pa kandidirali). Vozlišča v omrežju predstavljajo uporabnike, usmerjene povezave od vozlišča i do vozlišča j pa predstavljajo, da je uporabnik i glasoval za uporabnika j. Ker se bova ukvarjali z iskanjem maksimalne klike na neusmerjenem grafu, bova podatke interpretirali kot neusmerjen graf - torej, da povezava med vozliščema obstaja, če je podana vsaj v eno smer).

V nadaljnem delu, bova poskusili graf uvesti v programski jezik Sage in ga zapisati v takšni obliki, da bova na njem lahko izvedli celoštevilski linearni program. Kot sva že omenili, je najin problem težko rešljiv, posebej za velike grafe. Da bova lahko izvedli program, bova poskusili primerno zmanjšati število vozlišč in povezav. Poiskali bova tudi tako imenovane kvazi - klike, za katere bo veljalo, da je vsaj 90% povezav, ki bi v kliki moralo biti, že v njej.