

Analiza velikih socialnih omrežij

POVZETEK: V vsakdanjem življenju se vedno bolj pogosto srečujemo z velikimi omrežji, kjer gre število točk/povezav v (deset ali sto) tisoče: socialna omrežja (omrežja prijateljstev, sodelovanja, komuniciranja med posamezniki, organizacijami, državami ...), rodovniki, računalniška omrežja (lokalna računalniška omrežja, omrežja povezav med predstavitenimi stranmi, Internet, ...), referenčna omrežja (omrežja citiranj), diagrami poteka programskih sistemov, organske molekule, transportna omrežja (cestna, vodovodna, električna omrežja ...), itd. Obvladovanje velikih omrežij predstavlja tako časovno, kot tudi prostorsko zahteven problem. Večina standardnih algoritmov za analizo omrežij ima visoke časovne zahtevnosti in so zato neprimerni za analizo velikih omrežij. V sestavku so predstavljeni pristopi k analizi in predstavitev tovrstnih omrežij. Pristopi so podprti s posebnim programom za analizo velikih omrežij. V sestavku je prikazanih nekaj primerov uporabe tega programa pri analizi in prikazu rodovnikov.

KLJUČNE BESEDE: velika omrežja, socialna omrežja, časovna zahtevnost, rodovniki, poročna prepletenost, Pajek.

1. Uvod

V vsakdanjem življenju se okoli nas nahajajo velika omrežja, pa če jih opazimo ali ne. Pripadajo predvsem našemu mikro okolju in makro okolju, če za mero velikosti okolja vzamemo običajni čutni svet ljudi. Velika omrežja nimajo vselej vidne in čvrste oblike, kot jo imajo na primer predmeti, ljudje ali klasične organizacije (stavbe). Kljub temu pa lahko iz ozadja predstavljajo pomemben faktor vsakdanjega dogajanja. S svojo lastno strukturo in dinamiko in s sebi lastnim vplivom namreč stanovitno učinkujejo na naše življenje. Odkar se je obseg našega zavedanja in tudi znanja o njih povečal, kar je predvsem posledica razvoja znanosti in nenehne rasti velikih infrastrukturnih omrežij okoli nas, in odkar so na razpolago še učinkovita orodja za njihovo analizo, postajajo velika omrežja vse bolj zanimiva za sistematično proučevanje.

Med velika omrežja štejemo omrežja, pri katerih gre število enot (točk/povezav) v (deset ali sto) tisoče. Ta dokaj ohlapna značilnost velikih omrežij je v bistvu raziskovalna prednost, ker nas vsebinsko ne omejuje in nam dopušča, da lahko med velika omrežja uvrstimo zelo različne pojave. Kot primer velikih omrežij naj navedemo nekatera bolj znana:

- socialna omrežja: omrežja prijateljstev, mentorstev [14], soavtorstev (Batagelj in Mrvar, 2000), sodelovanja, komuniciranja med posamezniki [3], organizacijami, državami,...

- rodovniki - sorodstvena drevesa ljudi (White et al., 1999; Dremelj et al., 1999), plemenskih živali;
- računalniška omrežja (lokalna računalniška omrežja, debatne skupine, omrežja povezav med predstavitvenimi stranmi, Internet,...);
- diagrami poteka programskih sistemov;
- referenčna omrežja, omrežja citiranj (Hummon in Doreian, 1989; Hummon et al., 1990);
- organske molekule [10];
- transportna omrežja (cestna, vodovodna, električna omrežja, letalska omrežja [15]), itn.

Eno od zanimivih področij analize socialnih omrežij, kjer je na voljo večje število empirično zajetih obsežnih omrežij, so rodovniki (genealogije), s katerimi se ukvarja (predvsem zbira podatke) vse več ljudi. Tako tudi v Sloveniji obstaja Slovensko rodoslovno društvo [12]. V večjih slovenskih rodovnikih najdemo podatke o več 10.000 osebah. Na omrežju Internet pa lahko najdemo še precej večje rodovnike, ki so povezani v poizvedovalni sistem GENDEX [2]. Analitska vprašanja, povezana s temi zbirkami podatkov, so tudi družboslovno izredno vznemirljiva, saj rodovniki vsebujejo velika sorodstvena omrežja, ki se raztezajo skozi daljši čas in lahko - ni pa to nujno - močno vplivajo na socialno vedenje ljudi.

Do nedavna je bilo na številna analitska vprašanja iz rodovnikov še težko odgovoriti, zlasti zato, ker so bili (ročni) načini iskanja odgovorov precej okorni, povezani z velikimi stroški in dolgotrajnimi proučevanji. Danes je položaj že precej drugačen. V tem prispevku bomo na primeru rodovnikov pokazali, da je danes, z metodičnim pristopom k velikim omrežjem in z uporabo ustreznih programov, možno že precej lažje odgovoriti na številna vprašanja, ki so še včeraj izgledala analitsko prezahtevna.

2. Zanimiva vprašanja okrog rodovnikov

Rodoslovni podatki se zbirajo za različne namene:

- rodovniki, zanimivi iz zgodovinskih in drugih vzrokov (npr. poroke med florentinskimi rodbinami (Padgett in Ansell, 1989), rodovnik dubrovniških plemiških družin (Dremelj et al., 1999), rodovnik ameriških predsednikov [1]);
- rodovniki posameznih družin in/ali teritorialnih območij za osebno ali religiozno uporabo, npr. rodovniki Mormonov [11];
- posebni rodovniki;
- študenti in njihovi mentorji pri izdelavi doktorskih disertacij: Theoretical Computer Science Genealogy [14];
- božanstva (antika).

V velikih rodovnikih si lahko postavimo veliko zanimivih vprašanj, na katera lahko odgovorimo samo na osnovi sorodstvenih povezav, ki jih predstavimo z ustreznim omrežjem. Na primer:

- Ali sta poljubni dve osebi v krvnem sorodstvu ali vsaj svaštvu (povezanost preko

ženinih oz. moževih sorodnikov). Vprašanje lahko razširimo na iskanje najkrajše sorodstvene zveze med dvema oseba.

- Iskanje števila vseh prednikov oziroma vseh potomcev dane osebe, oziroma širše - določanje vseh prednikov ali vseh potomcev izbrane osebe.
- Določanje oseb v določeni okolici izbrane osebe (nekaj generacij pred in nekaj za izbrano osebo).
- Ugotavljanje prepletenosti med družinami (večkratnih porok med istimi družinami / rodbinami, ki so imele tudi ekonomski namen - ohranjanje bogastva znotraj družin).
- Iskanje zanimivih vzorcev v rodovniku - porok med sorodniki, otrok z več starši, velikokrat poročenih oseb. Na osnovi teh statistik lahko primerjamo različne rodovnike - oziroma širše - kulture med sabo.
- Iskanje najdaljših nepretrganih verig v rodovniku in še posebej najdaljših nepretrganih moških verig v rodovniku, ki so imele v preteklosti poseben pomen (dedovanje, ohranjanje rodovnega imena,...).
- Statistike: povprečno število otrok, največje število otrok,...

3. Metodologija

Pogoj za sistematično analizo velikih omrežij, kamor spadajo tudi rodovniki, je konceptualnokategorialni aparat na dovolj splošni ravni. Če torej zanemarimo vsebinsko različnost velikih omrežij, nam ravno abstraktne skupne lastnosti velikih omrežij omogočijo, da se jih lahko metodično lotimo na soroden način. Za uspešno analitsko reševanje problemov velikih omrežij bomo omrežja najprej opredelili. Nato bomo nakazali, s katerimi algoritmi je velika omrežja mogoče učinkovito (dovolj hitro) analizirati, našteji podatkovne strukture, ki nam to omogočajo in predstavili program za analizo in prikaz velikih omrežij. Omenjene pristope bomo prikazali na primeru analize izbranega rodovnika.

3.1 Definicija omrežja

Naj bo dana končna množica enot $\mathbf{E} = (X_1, X_2, \dots, X_n)$. Odnosi med enotami naj bodo opisani z eno ali več dvomestnimi *relacijami* $R_t \subseteq \mathbf{E} \times \mathbf{E}$, $t = 1, \dots, r$, ki določajo *omrežje* $N = (\mathbf{E}, R_1, R_2, \dots, R_r)$. Zapis $X_i R X_j$ preberemo: enota X_i je v relaciji R z enoto X_j .

Relacija R lahko predstavlja prijateljstvo, sovraštvo, sorodstvene odnose (X_i je otrok od X_j ali X_i je sestra od X_j ali X_i je poročen z X_j), citiranje (avtor X_i citira avtorja X_j ali v članku X_i se sklicujemo na članek X_j)... V nadaljevanju se bomo v večini primerov omejili na eno samo relacijo R in sicer v primeru rodovnikov na relacijo X_i je otrok od X_j .

Omrežje, v katerem nastopa relacija R , lahko predstavimo na več načinov:

- Predstavitev s pripadajočo **dvojiško matriko**

$$\mathbf{R} = [r_{ij}]_{n \times n}, \text{ kjer je } r_{ij} = \begin{cases} 1 & X_i R X_j \\ 0 & \text{sicer} \end{cases}$$

Včasih je lahko r_{ij} tudi realno število, ki izraža moč relacije R med enotama X_i in X_j .

- opis z **grafom** $G=(V, L)$, kjer je V množica točk (*Vertices*) in L množica (usmerjenih ali neusmerjenih) povezav (*Lines*). Točke grafa predstavljajo enote v omrežju, povezave pa relacijo. Graf predstavimo s sliko, kjer točke grafa predstavimo s krožci, usmerjene povezave s puščicami, neusmerjene povezave pa z daljicami, ki povezujejo ustrezne točke. Zapis $X_i R X_j$ v primeru predstavitve z grafom pomeni, da v grafu obstaja usmerjena povezava (R), ki vodi od točke X_i do točke X_j . Točka X_i se imenuje začetna, točka X_j pa končna točka povezave.

3.2 Učinkoviti algoritmi za analizo velikih omrežij

Časovna $T(n)$ in prostorska $S(n)$ zahtevnost algoritma nam povesta, koliko časa in prostora potrebujemo za njegovo izvedbo na nalogah velikosti n (v našem primeru - število točk ali povezav v omrežju). V večini velikih omrežij je število povezav m istega velikostnega reda kot število točk - $O(n)$ ali največ $O(n \log n)$. Taka omrežja imenujemo *redka omrežja*. V nadaljevanju bomo zato predpostavili, da analiziramo velika, a redka omrežja.

Glede na vse večje pomnilniške zmogljivosti današnjih računalnikov prostorska zahtevnost za redka omrežja ni več kritična. Problem rešimo z ustreznimi podatkovnimi strukturami za notranjo predstavitev omrežja.

Časovna zahtevnost pa ostaja še vedno velik problem, saj tudi veliko hitrejši računalniki ne pomagajo veliko. V teoriji algoritmov veljajo problemi z algoritmi s polinomsko časovno zahtevnostjo za pohlevne - lahko rešljive. Toda, v primeru zelo velikih n , so lahko dejansko prezahtevni že algoritmi s časovno zahtevnostjo $O(n^2)$, kar je razvidno iz tabele 1.

Tabela 1: Časovne zahtevnosti algoritmov (Pentium/64M/90MHz)

T(n)	1.000	10.000	100.000	1.000.000
$O(n)$	0,00 s	0,015 s	0,17 s	2,22 s
$O(n \log n)$	0,00 s	0,00 s	0,40 s	5,14 s
$O(n^2)$	0,07 s	7,50 s	12,5 min	20,83 h
$O(n^3)$	0,10 s	1,67 min	1,16 dni	3,17 let

Iz tabele je razvidno, da so za analizo velikih omrežij primerni le algoritmi, ki imajo *podkvadratične* časovne zahtevnosti. Naj dodamo, da je vsebinska zahtevnost analize velikih omrežij na tej ravni splošnosti neodvisna od predlagane metodologije raziskovalca in je zato kakovost vsebinske interpretacije povsem odvisna od disciplinarne ekspertnosti raziskovalca.

3.3 Podatkovne strukture

Sistem za analizo velikih omrežij smo oprli na naslednje podatkovne strukture:

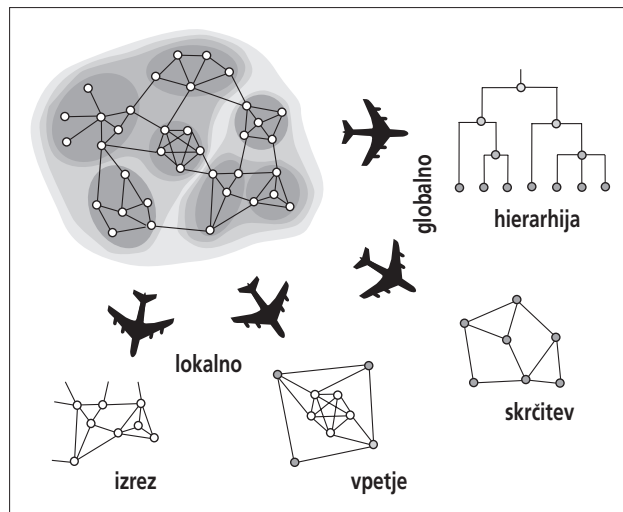
- *omrežje* - osnovna struktura (npr. točke predstavljajo osebe, povezave pa neko sorodstveno razmerje). Potrebujemo torej program, ki razpozna, oziroma predela podatke, zapisane na nek standarden način na vhodni datoteki, v ustrezno omrežje.

- *permutacija* - preureditev točk, ki jo predstavimo s tabelo (npr. preureditev oseb v rodovniku po naraščajoči letnici rojstva);
- *razbitje* - omrežju prirejena tabela, ki predstavlja posebno razvrstitev točk po skupinah (razredih), npr. razvrstitev oseb po spolu v primeru rodovnikov;
- *vektor* - lastnost točk omrežja (realno število, ki predstavlja neko kvantitativno lastnost točke, npr. letnico rojstva osebe);
- *skupina* - seznam izbranih točk (npr. osebe, ki pripadajo dani družini);
- *hierarhija* - hierarhična razvrstitev skupin (in posledično točk) omrežja, ki je predstavljena s splošno drevesno strukturo. Merilo hierarhične razvrstitve je podobnost / različnost točk oziroma skupin.

3.4 Sistem za analizo velikih omrežij

Velikih omrežij ne moremo učinkovito analizirati z uporabo standardnih programov za analizo omrežij, ki povečini temeljijo na matrični predstavitvi omrežja in so zato omejeni na omrežja z nekaj deset ali sto točkami. Za analizo velikih omrežij so potrebni posebni programi. Predstavili bomo program Pajek, ki je bil napisan ob upoštevanju navedenih opredelitev, algoritmov in podatkovnih struktur. Glavni cilji pri zasnovi programa so:

- podpreti *abstrakcijo* s postopno razčlenitvijo velikega omrežja na več manjših omrežij, ki jih lahko nadalje analiziramo z uporabo običajnih metod;
- ponuditi uporabniku močna orodja za *prikaz* omrežij;
- vgraditi večje število *učinkovitih* algoritmov za analizo velikih omrežij.



Slika 1: Cilji pri zasnovi programa Pajek

Na sliki 1 je abstrakcija prikazana z različnimi pogledi na omrežje. Na omrežje lahko pogledamo kot na celoto (*globalen pogled*) ali pa se osredotočimo na del omrežja (*lokalen pogled*).

Obstajata dva različna globalna pogleda: *skrčitev* in *hierarhija*. Pri skrčitvi obravnavamo točke v skupini kot neko novo točko - točke v skupinah stisnemo v skupno točko. Še splošnejši globalni pogled pa je hierarhija - točke v omrežju hierarhično združujemo v skupine in na ta način dobimo drevo. Kasneje se sprehajamo po drevesu in glede na položaj v drevesu 'vidimo' omrežje podrobneje (če se nahajamo v bližini listov drevesa) ali bolj na grobo (če smo v bližini korena drevesa) - točke, ki so pod izbranim vozliščem v drevesu, namreč stisnemo v novo točko.

Obstajata tudi dva lokalna pogleda na omrežje. Prvega imenujemo *izrez* - iz omrežja izrežemo manjši del in ga v nadaljevanju obravnavamo ločeno. Drugi lokalni pogled se imenuje *vpetje* - iz omrežja izrežemo manjši del, hkrati pa prikažemo, kako je ta del povezan z ostankom omrežja.

Poseben poudarek je v programu dan avtomatičnemu določanju prikazov omrežij [13]. V sistem so vključeni številni tovrstni algoritmi: energijska risanja (Kamada in Kawai, 1989; Fruchterman in Reingold, 1991), risanja z uporabo lastnih vektorjev (Datta, 1995; Golub in val Loan, 1996), nivojska risanja (rodovnikov in drugih acikličnih struktur). Sistem nudi tudi orodja za *ročno risanje omrežij*.

Dobljene prikaze lahko pretvorimo v številne izhodne formate, ki si jih lahko nadalje ogledujemo s posebnimi pregledovalniki za ravninske in prostorske prikaze: (Encapsulated PostScript - GSView [8]; Scalable Vector Graphics - SVG [7]; VRML - CosmoPlayer [6]; MDLMOL - Rasmol [10]; Chime [5]; Kinemages - Mage [9]).

V nadaljevanju si bomo pogledali uporabo tega programa pri analizi rodovnikov.

4. Primer: analiza rodovnikov

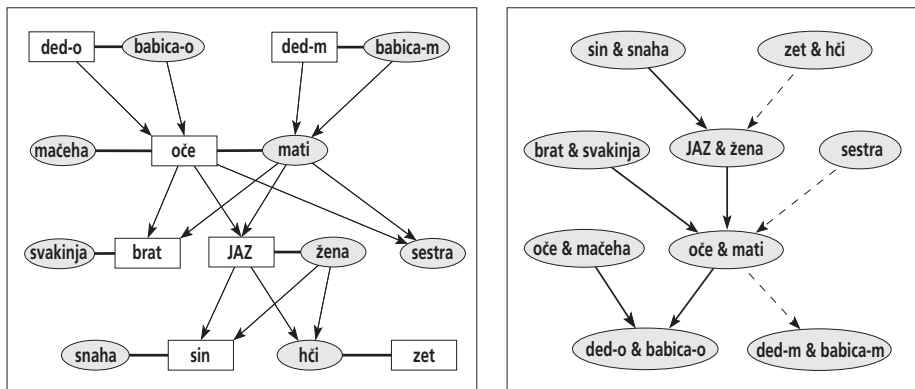
Kot smo omenili v uvodu, je eno od področij, kjer je na voljo večje število velikih omrežij, rodoslovje. V rodoslovju obstaja kar nekaj programov za vnos podatkov (GIM, Brother's Keeper, Family Tree Maker ...), ne obstajajo pa programi, ki bi omogočali njihovo analizo. To vrzel skuša zapolniti program Pajek.

Rodoslovne podatke lahko predelamo v dve vrsti grafov - rodovnikov, ki ju bomo imenovali *navadni* in *parni* rodovnik.

V *navadni* obliki (Ore, 1963) je vsaka oseba predstavljena s točko grafa, poroka je označena z neusmerjeno povezavo med točkama, usmerjene povezave pa vodijo od staršev do otrok (naprej v prihodnost). Tak je primer na levi strani slike 2. Na tej sliki so predstavljene štiri generacije. Zaradi splošnosti je pri izbrani osebi (očetu) predpostavljeno, da se je dvakrat poročila, zato sta z ustrežno točko (očetom) povezani dve neusmerjeni povezavi (povezavi z materjo in mačeho). Na tej sliki so tudi vse osebe (razen sestre) poročene.

V *parni* obliki (*p-graph*, *p* je prva črka francoske besede *parenté* - sorodstvo) so točke grafa lahko posamezniki ali pari. Več o tej obliki predstavitve najdemo na predstavitveni strani D. R. Whita z Univerze Irvine v Kaliforniji [4]. Če neka oseba še ni poročena, je v grafu predstavljena s svojo točko; če pa je poročena, je v grafu predstavljena skupaj s svojim zakoncem s skupno točko. V tej obliki imamo samo usmerjene povezave, ki vodijo od otrok do staršev (nazaj v preteklost). Ker so točke lahko tudi

pari, moramo posebej označiti, ali se povezava nanaša na moškega ali na žensko. Povezave, ki se nanašajo na izvirnega moškega (kažejo pa na njegove starše), so označene z neprekinjeno puščico, povezave, ki se nanašajo na izvorno žensko (kažejo pa na njene starše), pa s prekinjeno (pikčasto) puščico. Če je neka oseba večkrat poročena, se pojavi v toliko točkah, kolikor je porok. Tak primer grafa je podan na desni strani slike 2. Ta slika prikazuje povsem identično situacijo kot tista na levi strani, le da je namesto navadnega rodovnika uporabljen parni rodovnik.



Slika 2: Predstavitev rodovnika v navadni obliki (levo) in v parni obliki (desno)

4.1 Podatki

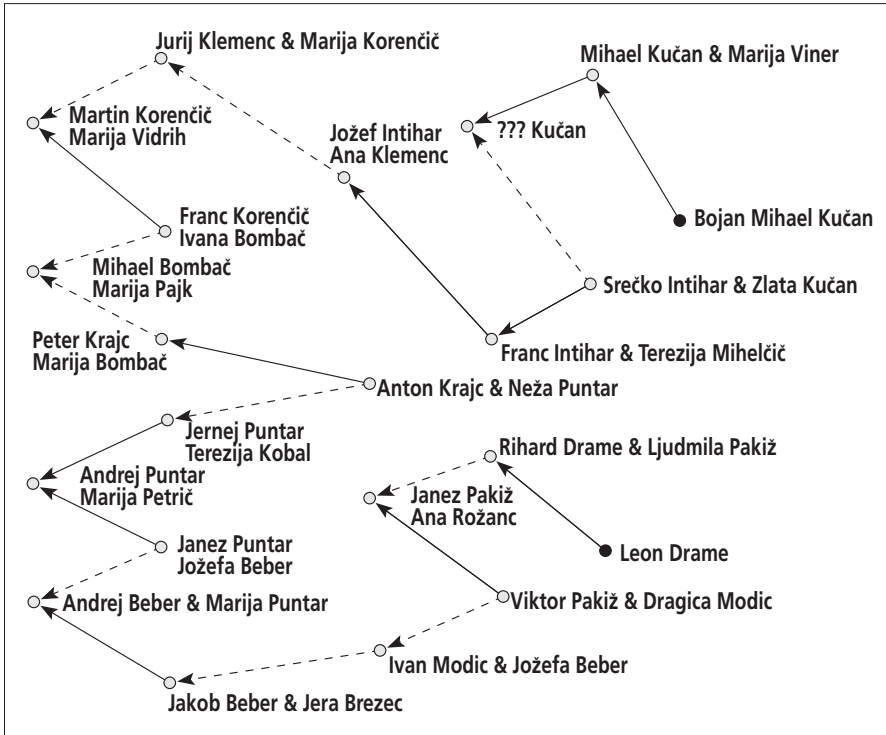
V velikih rodovnikih lahko odkrijemo zanimive stvari. V primerih, ki sledijo, bo uporabljen eden od (trenutno) največjih slovenskih rodovnikov - rodovnik Leona Drameta, v katerem se nahajajo podatki o okrog 30.000 osebah. Motivacija za začetek zbiranja teh podatkov je bila, da se čimdlje v preteklost popiše prednike vseh živečih oseb na določenem teritoriju. Najstareši zapis v tem rodovniku se nanaša na leto 1540.

4.2 Rezultati analize

V nadaljnjih podpoglavjih je prikazanih nekaj zanimivih podomrežij (lokalnih pogledov, oziroma izrezov), dobljenih iz navedenega rodovnika.

4.2.1 Iskanje najkrajše sorodstvene zveze med dvema osebama

Prvo vprašanje, ki si ga lahko zastavimo nad rodovnikom, je, ali sta dve naključno izbrani osebi v sorodu. Če sta v sorodu, bi seveda radi našli najkrajšo sorodstveno zvezo med njima.

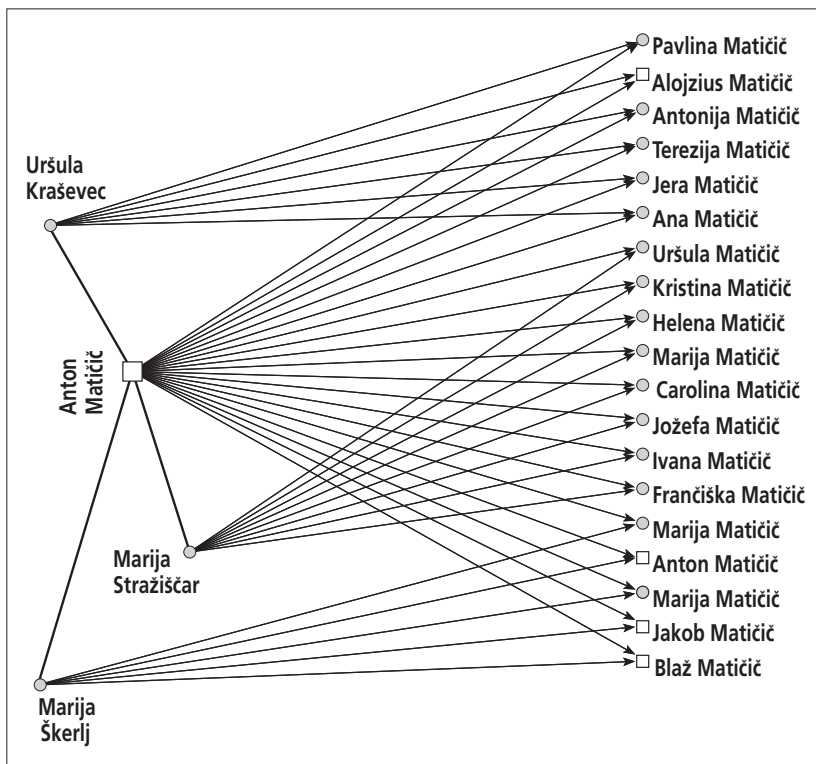


Slika 3: Najkrajša pot (parni rodovnik)

Slika 3 prikazuje najkrajšo pot med Leonom Drametom (zbirateljem podatkov v rodovniku) in Bojanom Mihaelom Kučanom (nečakom slovenskega predsednika), če uporabimo parni rodovnik. Slika je narisana po generacijskih nivojih, starejše generacije so narisane bolj levo, mlajše pa bolj desno. Tako lahko razberemo, da je Bojan Mihael Kučan za eno generacijo mlajši od Leona Drameta. Puščica, ki vodi iz točke, ki predstavlja Leona Drameta, kaže na njegove starše (Rihard Drame in Ljudmila Pakiž). Iz slike razberemo tudi, da naj bi Leon Drame ne bil poročen (v ustrezni točki je predstavljen brez svojega partnerja) ter da se v starejših generacijah pogosto lahko pojavljajo manjkajoči podatki (npr. imena oseb, glej na sliki oznako '??? Kučan').

4.2.2 Število neposrednih potomcev izbrane osebe

Zanimivo je tudi vprašanje o osebi (ali paru) z največjim številom potomcev. Slika 4 prikazuje osebo z največ zakonskimi partnerji in otroki (skupno glede na vse njegove poroke). V tem primeru je uporabljen navadni rodovnik (puščice torej kažejo od vsakega od staršev na njune otroke).



Slika 4: Posameznik z največ zakonskimi partnerji in otroci skupno (navadni rodovnik)

Iz slike razberemo, da se je Anton Maticič poročil trikrat in imel skupaj 19 otrok (šest z Uršulo Kraševac, osem z Marijo Stražiščar in pet z Marijo Škerlj). Na sliki so osebe moškega spola označene s kvadrati, osebe ženskega spola pa z krogi.

V oči nam hitro pade še ena zanimivost: obstajajo celo otroci istih staršev z enakimi imeni (npr. Marija Maticič). Če pogledamo v zapiske, lahko najdemo razlog: zaradi visoke umrljivosti otrok, so starši naslednjemu rojenemu otroku radi dali spet enako ime (še posebej, če je šlo za ohranjanje očetovega imena).

4.2.3 Iskanje 'zanimivih' porok

V tem podglavju bomo zaradi nedvoumnosti uporabljali pojem *družina* za partnersko zvezo med možem in ženo, pojem *rodbina* pa za podmnožico tistih oseb v rodovniku, ki imajo skupnega prednika.

V rodovniku si lahko postavimo še naslednji dve vprašanji:

- Ali so se pripadniki izbranih rodbin večkrat poročali med sabo? Tu gre za prepletenosti, kjer poročna partnerja nimata skupnega prednika (poročna prepletenost).

- Ali so se pripadniki iste rodbine poročali med sabo - krvne poroke: poroke, kjer imata partnerja skupnega prednika.

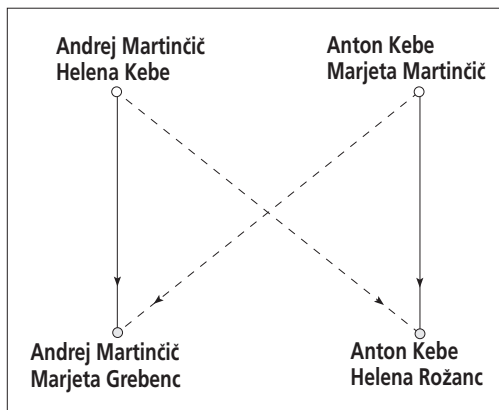
Na obe vprašanji lahko odgovorimo le, če rodovnik obravnavamo v parni obliki - to je tudi glavna analitska prednost parnega rodovnika pred navadnim rodovnikom (poleg manjšega števila točk in povezav). Za podani vzorec, ki predstavlja neko zanimivo situacijo, nam mora program vrniti vse pojavitve takega vzorca v rodovniku.

Oba tipa porok se na sliki izbranega dela parnega rodovnika kažeta kot sklenjeni verigi: če začnemo pot v neki točki in sledimo povezavam, pri čemer smeri povezav zanemarimo, se lahko vrnemo v začetno točko (glej sliki 5 in 6). Poglejmo si oba primera malo bolj podrobno.

Ugotavljanje poročne prepletenosti rodbin

Vprašanje o poročni prepletenosti rodbin lahko razbijemo na dve podvprašanji in sicer:

- Ali so se pripadniki izbranih dveh rodbin sploh kdaj poročili? Za neporočanja so lahko (ni pa nujno) kriva nesoglasja, ki se prenašajo preko generacij (tak primer je obdelan v znanem Shakespearjevem delu Romeo in Julija).
- Če smo na prvo vprašanje odgovorili pozitivno (obstaja vsaj ena poroka med rodbinama), si lahko postavimo nadaljnje vprašanje: ali obstajajo med temi porokami tudi sočasne poroke v okviru ene generacije. Do tega pride, če se brata poročita s sestrama iz druge rodbine, kjer gre torej za dvakratno poroko v okviru ene generacije med (istima) dvema rodbinama.



Slika 5: Primer poroke med sorodniki

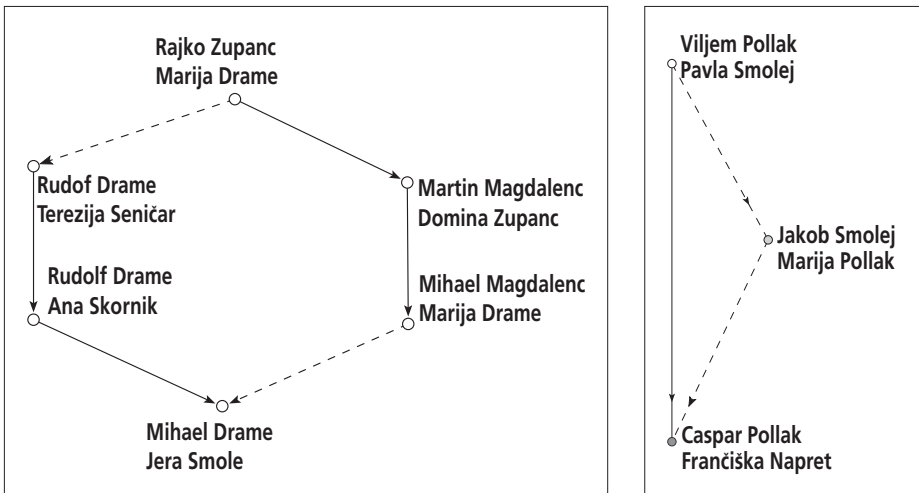
Slika 5 prikazuje primer poroke takega tipa. V tem primeru sta se brat in sestra (Andrej in Marjeta Martinič) poročila s sestro in bratom iz druge rodbine (Heleno in Antonom Kebetom). S porokami takega tipa, so rodbine ohranjale svoje premoženje (bogastvo) oziroma ga celo povečevale. Veliko takih navzkrižnih porok lahko zato zasledimo pri poročanju višjih družbenih slojev. Zanimiv je primer dubrovniškega plemstva, kjer se je zaradi teritorialne zaprtosti, pomanjkanja partnerjev iz ustreznih

plemiških rodbin in težnje po ohranjanju bogastva pojavljalo zelo veliko porok tega tipa (Dremelj et al., 1999).

Ugotavljanje krvnega sorodstva v porokah

Pri iskanju krvnih porok si lahko postavimo naslednji vezani vprašanji:

- Ali sta dve družini v krvnem sorodstvu (eden od partnerjev ima skupnega prednika z enim od partnerjev iz druge družine)?
- Ali obstaja med dvema družinama, ki sta v krvnem sorodstvu, kakšna poroka? Če da in če jih je več, katere izmed teh krvnih porok so (naj)bolj kočljive? Pri tem bomo za bolj kočljive šteli tiste poroke, kjer se poročijo bližnji krvni sorodniki - skupni prednik je torej generacijsko bliže (npr. obravnavamo njegove sinove in hčere), za manj pa tiste, kjer se poročijo bolj oddaljeni krvni sorodniki - skupni prednik se pojavi pred veliko generacijami (npr. obravnavamo njegove prapravnike in prapravnkinje).



Slika 6: Dva primera krvne poroke med sorodniki

Slika 6 prikazuje dva primera krvne poroke - levo je primer krvne poroke med sorodniki, kjer sta se poročila pravnuk (Rajko Zupanc) in pravnukinja (Marija Drame), desno pa primer, ko sta se poročila sin (Viljem Pollak) in vnukinja (Pavla Smolej).

Morda bi na tem mestu opozorili še na prisotnost napak v rodovnikih. Pri vnosu podatkov o daljnjih prednikih se lahko zelo hitro pojavijo napake, ko zaradi manjkajočih podatkov (predvsem o letnicah rojstva) rodoslovec vnese napačne sorodstvene povezave med osebami. To lahko povzroči, da se v rodovniku pojavijo omenjeni napačni vzorci - poroke med bližnjimi sorodniki. V tem primeru pa lahko spoznamo praktično uporabnost programa za analizo velikih omrežij - saj s pomočjo programa lažje najdemo napake pri vnosu.

Pri poroki med sinom in vnukinjo gre istočasno še za generacijski preskok. Do tega lahko pride zaradi različnih vzrokov. Najenostavnejši razlog je lahko velika razlika

med starostjo poročnih partnerjev. Do generacijskega preskoka pa lahko pride tudi, če so bile osebe iz različnih generacij v danem zgodovinskem času približno enako stare. Vzrok za to so lahko velike časovne razlike med rojstvi posameznih otrok in velike razlike med starostjo ob poroki posameznih otrok. Generacijskih preskokov lahko v preteklosti najdemo veliko. Pri večini generacijskih preskokov smo že dognali, da je običajno daljša veja po ženini strani (torej gre npr. večkrat za poroko sin-vnukinja kot za poroko vnuk-hči). To zakonitost lahko pojasnimo z dejstvom, da so se v preteklosti ženske poročale veliko prej kot moški. V primeru dubrovniškega plemstva so se ženske poročale stare v povprečju 18, moški pa 30 let.

Glede krvoskrunstva je zadnja situacija (situacija na desni strani slike 6) veliko bolj kočljiva. Podobnih situacij lahko najdemo v velikih rodovnikih zelo veliko. Za merjenje jakosti krvoskrunstva je bil vpeljan poseben indeks prepletenosti (*relinking index*). Definicijo tega indeksa najdemo v članku *Analyzing Large Kinship and Marriage Networks with Pgraph and Pajek* (White et al., 1999). Indeks lahko zavzame vrednosti med 0 in 1, kjer večja vrednost pomeni bližje sorodstvo oziroma močnejše krvoskrunstvo. Za primer povejmo, da ima bolj oddaljena krvna poroka na levi strani slike 6 indeks prepletenosti 0.2, desna (bližja) pa 0.5. Indeks doseže vrednost 1 v primeru poroke med bratom in sestro.

Naj omenimo še, da lahko z istim indeksom merimo tudi jakost prepletenosti, ko ne gre za krvne poroke (poroka med parom bratov in parom sester iz različnih rodbin). V omenjenem članku se nahajajo ta in še številni drugi primeri in podrobnosti, povezani z analizo rodovnikov, opisani pa so tudi koraki, kako pridemo do željenih rezultatov.

5. Sklepne misli

V članku smo prikazali, kako se lahko danes, s sistematično metodologijo in ob pomoči sodobnih informatičnih sredstev, lotevamo tudi zahtevnejših analitskih vprašanj nad velikimi (socialnimi) omrežji. S primeri smo posegli na področje empirične genealogije in za ponazoritev izvedli nekaj osnovnih vpogledov v Drametov rodovnik. Z uporabljenimi primeri seveda še zdaleč nismo izčrpali vseh vsebinsko zanimivih vprašanj, ki so možna glede na informacije v tej podatkovni bazi. V prihodnje bi zato kazalo (s pomočjo razpoložljivih časovnih in poklicnih podatkov) opraviti tudi kakšno longitudinalno študijo, na primer o trendnih spremembah v vzorcih statusne homogamije, v starostnih vzorcih poročanja, v številu potomcev, v številu priseljencev, itn. Posebna vrednost takšnih mikro-socialnih študij, zlasti nad rodovniki teritorialnih območij, je - poleg nazornosti - v njihovi morebitni posplošljivosti, saj iz historične analize lokalnih pojavov lahko zaslutimo obrise širših sprememb v vsakdanjem življenju ljudi.

Zahvala

Zahvaljujem se Antonu Krambergerju, ki je s svojimi pripombami v veliki meri prispeval k boljši razumljivosti in kakovosti članka.

Literatura

- Batagelj, V., Mrvar, A. (1998): Pajek - A Program for Large Network Analysis. *Connections*, **21**, (2), 47-57.
- Batagelj, V., Mrvar, A. (2000): Some Analyses of Erdős Collaboration Graph. *Social Networks*, **22**, 173-186.
- Datta, B. N. (1995): Numerical Linear Algebra and Applications. Brooks&Cole Publishing Company, Pacific Grove.
- Dremelj, P., Mrvar, A., Batagelj, V. (1999): Rodovnik dubrovniških plemiških družin med 12. in 16. stoletjem. *Drevesa - Bilten Slovenskega rodoslovnega društva*. Letnik 6, številka **1-4**, december 1999, 4-11.
- Golub, G. H., Van Loan, C. F. (1996): *Matrix Computations*. The John Hopkins University Press, Baltimore.
- Fruchterman, T. M. J., Reingold, E. M. (1991): Graph Drawing by Force-Directed Placement. *Software, Practice and Experience*, **21**, 1129-1164.
- Hummon, N. P., Doreian, P. (1989): Connectivity in a Citation Network: The Development of DNA Theory. *Social Networks*, **11**, 39-63.
- Hummon, N. P., Doreian, P., Freeman, L. C. (1990): Analyzing the Structure of the Centrality-Productivity Literature Created Between 1948 and 1979. *Knowledge: Creation, Diffusion, Utilization*, Vol. 11, June, No. **4**, 459-480.
- Kamada, T., Kawai, S. (1989): An Algorithm for Drawing General Undirected Graphs. *Information Processing Letters*, **31**, 7-15.
- Mrvar, A. (1999): *Analiza in prikaz velikih omrežij*. Fakulteta za računalništvo in informatiko: Doktorska disertacija.
- Mrvar, A., Batagelj, V. (1997): Pajek - program za analizo obsežnih omrežij. Uporaba v rodoslovju. *Drevesa - Bilten slovenskega rodoslovnega društva*. Letnik 4, številka **12**, december 1997, 4-6.
- Ore, O. (1963): *Graphs and Their Uses*. New York, Random House.
- Padgett, J. F., Ansell, C. K. (1989): Robust action and the rise of the Medici 1400-1434. *American Journal of Sociology*, **98**, 1259-1319.
- White, D. R., Jorion, P. (1996): Kinship Networks and Discrete Structure Theory: Applications and Implications. *Social Networks*, **18**, 267-314.
- White, D. R., Batagelj, V., Mrvar, A. (1999): Analyzing Large Kinship and Marriage Networks with Pgraph and Pajek. *Social Science Computer Review - SSCORE*, **17** (3), 245-274.

Spletni naslovi (izpis: marec 2001)

- [1] American Presidents GEDCOM file:
<ftp://www.dcs.hull.ac.uk/public/genealogy/>
- [2] GENDEX - WWW Genealogical Index:
<http://www.gendex.com/gendex/>
- [3] Graph Drawing Competition 1996, graf B:
www.research.att.com/conf/gd96/graphs/B
- [4] Parni rodovniki:
<http://eclectic.ss.uci.edu/~drwhite/pgraph/p-graphs.html>

- [5] Plug-in Chime:
<http://www.mdli.com/download/chimedown.html>
- [6] Plug-in Cosmo Player:
<http://cosmosoftware.com/>
- [7] Plug-in SVG:
<http://www.adobe.com/svg/viewer/install/>
- [8] Program GSView:
<ftp://ftp.cs.wisc.edu/pub/ghost/rjl/>
- [9] Program Mage:
<ftp://kinemage.biochem.duhe.edu/PCPrograms/Win95-98/2000>
- [10] Program RasMol (RASter MOLEcules):
<http://www.umass.edu/microbio/rasmol/getras.htm>
- [11] Rodovniki Mormonov:
<http://www.familytreemaker.com/00000116.html>
<http://www.genforum.com/mormon/>
- [12] Slovensko rodoslovno društvo:
<http://genealogy.ijp.si/slovrdrd.htm>
- [13] Tekmovanja v risanju grafov:
<http://vlado.fmf.uni-lj.si/pub/gd/gd00.htm>
- [14] Theoretical Computer Science Genealogy:
<http://sigact.acm.org/genealogy/>
- [15] Transportation Networks, National Transportation Atlas Database, Bureau of Transportation Statistics:
<http://www.bts.gov/gis/ntatlas/networks.html>

Avtorjev naslov

Dr. Andrej Mrvar,
Fakulteta za družbene vede Univerze v Ljubljani,
p.p. 2547, 1000 Ljubljana,
e-mail: andrej.mrvar@uni-lj.si



Program Pajek

Pajek (Batagelj in Mrvar, 1998; Mrvar, 1999) je programski paket, za okolje Windows (95/98/NT/2000), ki omogoča analizo in prikaz *velikih omrežij* (omrežij z več tisoč točkami). Program je prosto dostopen na naslovu: <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>
Avtorja programa sta Vladimir Batagelj in Andrej Mrvar. Program Pajek je še v razvoju. Zadnjo verzijo, različna omrežja, analize in druge zanimivosti, povezane z analizo omrežij je moč najti na njegovi predstavitveni strani.

Rokopis sprejet februarja 2001. Po mnenju uredništva je članek uvrščen v kategorijo: izvirni znanstveni članek (s kvantitativno argumentacijo).