

Omrežje sodelovanj med avtorji prispevkov iz Informatice in Uporabne informatike

Neli Blagus, Marko Bajec

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Večna pot 113, 1000 Ljubljana

neli.blagus@fri.uni-lj.si; marko.bajec@fri.uni-lj.si

Izvleček

Znanstveno sodelovanje je pomemben mehanizem razvoja in napredka v znanosti. Proučujemo ga lahko na različne načine, med drugim z opazovanjem soavtorstev raziskovalnih objav. V prispevku analiziramo sodelovanje slovenskih raziskovalcev s področja informatike, pri čemer se osredotočimo na prispevke iz revij *Informatica* in *Uporabna informatika*. Sodelovanja predstavimo v obliki družbenega omrežja, sestavljenega iz vozlišč, ki predstavljajo avtorje, ter povezav med njimi, ki pomenijo soavtorstvo pri vsaj eni objavi iz obeh revij. Podrobneje raziščemo lastnosti omrežja in sodelovanj med avtorji. Prav tako opazujemo razvoj omrežja skozi čas, pri čemer se osredotočimo predvsem na spreminjanje različnih značilnosti, kot so število vozlišč, premer, velikost največje komponente in povprečna razdalja med vozlišči. Pri analiziranju sodelovanj skozi čas pa opazujemo število člankov in število novih avtorjev vsako leto ter spreminjanje števila avtorjev na članek in števila člankov na avtorja od začetka izhajanja obeh revij.

Ključne besede: analiza omrežij, družbena omrežja, omrežje sodelovanj, informatika, razvoj omrežja.

Abstract

A Network of Collaboration: *Informatica* and *Uporabna Informatika*

Collaboration is an important mechanism of development in modern science. We can observe the patterns of scientific collaboration through analyzing the co-authorship network, formed by nodes (present authors) and links (denote the co-authorship of a paper between two authors). In this paper, we study the collaboration between the authors of papers published in *Informatica* and *Uporabna informatika*. We analyze the fundamental properties of the obtained network and the collaboration. Furthermore, we explore the evolution of the network and observe the changes of the network over time, for example, the number of nodes, the diameter, the size of the largest connected component and the average distance between nodes. We also observe the collaboration patterns over time, including the number of papers and the number of new authors each year.

Key words: Network analysis, Social networks, Collaboration network, Informatics, Network evolution

1 UVOD

Analiza družbenih omrežij (angl. social network analysis) (Carrington, Scott, & Wasserman, 2005; Newman, 2010) je pomembno sredstvo za razumevanje delovanja družbe (npr. razmerij med posamezniki) ter različnih procesov v družbi (npr. širjenje informacij ali bolezni). Družbena omrežja so sestavljena iz vozlišč, ki predstavljajo osebe, ter povezav, ki pomenijo interakcije med njimi. Vozlišča družbenega omrežja lahko opišemo z različnimi atributi (npr. ime, starost, spol), povezavam pa lahko določimo tip ali utež (npr. vrsta poznanstva, oblika sodelovanja ali število interakcij med dvema osebama).

Gradnja družbenih omrežij je lahko svojevrsten izziv, saj je težko podati točno in objektivno definicijo npr. poznanstva ali prijateljstva, ki določa povezave v omrežju (Newman, 2003). To pa ne velja za podvrsto družbenih omrežij, ki jim pravimo omrežja pripadnosti (angl. affiliation networks) (Breiger, 1974). Zgradimo jih iz podatkov o osebah, ki pripadajo različnim skupinam, povezave med osebami pa pomenijo pripadnost isti skupini. Primeri omrežij pripadnosti so omrežja filmskih igralcev (angl. network of film actors) (vozlišča so igralci, povezave med nji-

mi pa pomenijo, da se skupaj pojavijo v vsaj enem filmu (Zhang idr., 2006)), omrežja oseb, ki se pojavijo v sorodnem kontekstu (angl. co-appearance network) (vozlišča predstavljajo osebe, povezave pa pomenijo, da sta dve osebi skupaj omenjeni na primer na spletni strani, v časopisnem članku ali v knjigi (Knuth, 1993)), ter omrežja sodelovanj (angl. collaboration network) (vozlišča predstavljajo raziskovalce, povezave med njimi pa pomenijo sodelovanje oziroma soavtorstvo pri enem ali več znanstvenih člankih (Newman, 2001a, 2001b)).

V literaturi so najpogosteje med naštetimi raziskana omrežja sodelovanj med raziskovalci. Znana študija s tega področja analizira Erdösevo število raziskovalcev (angl. Erdős number) (de Castro & Grossman, 1999), ki pomeni razdaljo med matematikom Paulom Erdösem ter poljubnim drugim raziskovalcem v omrežju sodelovanj (npr. raziskovalec, ki je napisal prispevek skupaj z Erdösem, ima Erdösevo število 1, tisti, ki je sodeloval s tem raziskovalcem, a ne neposredno z Erdösem, ima Erdösevo število 2 in tako naprej). Sorodne študije nadalje raziskujejo povprečno Erdösevo število, spreminjanje števila skozi čas (Grossman & Ion, 1995) in druge lastnosti Erdösevega omrežja sodelovanj (Batagelj & Mrvar, 2000).

Veliko analiz omrežij sodelovanj se osredotoča na iskanje razlik pri sodelovanju raziskovalcev v različnih znanstvenih disciplinah, na določanje najpomembnejših raziskovalcev glede na znanstveno disciplino (Newman, 2004) ali znanstveno revijo (Martin, Ball, Karrer & Newman, 2013), na iskanje vzorcev sodelovanj (Hou, Kretschmer, & Liu, 2008) ter spreminjanje le-teh skozi različna časovna obdobja (Barabási, 2002). Sorodne študije raziskujejo tudi omrežja citiranj med prispevki, pri čemer vozlišča predstavljajo prispevke, povezave med njimi pa pomenijo, da je en prispevek citiral drugega. Iz rezultatov takšnih raziskav lahko sklepamo na primer na povezanost znanstvenih disciplin ali poiščemo najpomembnejše avtorje v omrežju glede na število citatov (Al, İrem, Gülten, 2012).

V prispevku se ukvarjamo z vprašanjem, kako pri pisanju prispevkov med seboj sodelujejo slovenski raziskovalci s področja informatike, pri čemer se osredotočimo na prispevke iz revij *Informatica* in *Uporabna informatika*. Glavni namen prispevka je analiza sodelovanj med raziskovalci, prikaz sodelovanj v obliki družbenega omrežja ter analiza razvoja sodelovanj skozi čas. Pri raziskavi nas je zanimalo,

kakšne so lastnosti in vzorci sodelovanj med avtorji prispevkov iz obeh revij, katere raziskovalne organizacije so najbolj zastopane v omrežju sodelovanj ter kako se trendi sodelovanja spreminjajo od začetka izhajanja obeh revij. Tako si v prvem delu prispevka ogledamo osnovne lastnosti omrežja in sodelovanj med avtorji ter primerjamo omrežje sodelovanj med avtorji prispevkov iz *Uporabne informatike* z omrežjem sodelovanj med avtorji prispevkov iz *Informatice*. V drugem delu prispevka podrobneje opazujemo razvoj omrežja in sodelovanj skozi čas.

Nadaljevanje prispevka je sestavljeno iz štirih delov. Najprej v razdelku 2 predstavimo omrežje ter njegove značilnosti. V razdelku 3 opazujemo spreminjanje omrežja in sodelovanj skozi čas, v četrtem razdelku sledi sklep.

2 OMREŽJE SODELOVANJ

2.1 Podatki in orodja

Omrežje sodelovanj smo sestavili na podlagi soavtorstev člankov, objavljenih v revijah *Uporabna informatika* (<http://www.uporabna-informatika.si/>) in *Informatica* (<http://www.informatica.si/>). Podatke o objavah smo pridobili iz kooperativnega online bibliografskega sistema in servisov (COBISS, <https://cobiss6.izum.si/>), ki vodijo bibliografije slovenskih raziskovalcev, evidentiranih pri Agenciji za raziskovalno dejavnost Republike Slovenije.

Sistem ima možnost iskanja bibliografije serijskih publikacij, ki omogoča izpis vseh v sistem zavedenih prispevkov iz določene publikacije. Za vsak prispevek so neposredno dostopni podatki o avtorjih, naslovu, letu izdaje, številu strani ter identifikacijska številka prispevka. V analizi smo uporabili podatke o avtorjih (ime in priimek), medtem ko smo podatke o raziskovalnih organizacijah, v katerih delujejo avtorji, pridobili iz informacijskega sistema o raziskovalni dejavnosti v Sloveniji (SICRIS, <http://www.sicris.si/>). Sistem SICRIS beleži podatke o raziskovalcih od leta 1998. Zapis o vsakem registriranem raziskovalcu med drugim vključuje raziskovalno šifro ter podatke o raziskovalni organizaciji. Podatki iz omenjenih virov so bili obdelani v programskem jeziku python, za analizo omrežij pa je bila uporabljena knjižnica *igraph* (<http://igraph.org/python/>). Analizirano omrežje je dostopno na <http://lovro.lpt.fri.uni-lj.si/support.jsp>.

Pri analizi smo upoštevali vse prispevke, ki so vpisani v sistem COBISS in so bili objavljeni v *Informatici*

v letih 1977 do 2013 ter v Uporabni informatiki v letih 1992 do 2013. Med njimi so tudi prispevki tujih avtorjev, če so bili med soavtorji slovenski raziskovalci.

Pri gradnji omrežja smo se soočili z dvema težavama pri določanju števila različnih avtorjev: 1) avtorja lahko imata enako ime in priimek ter 2) v sistemu so avtorji vneseni na več načinov (npr. z imenom in priimkom ali z začetnico imena in priimkom). Izkaže se, da uporaba začetnice imena ter celega priimka predstavlja spodnjo mejo pri ocenjevanju števila različnih avtorjev (Newman, 2001), zato smo se pri analizi odločili uporabiti ta pristop.



Slika 1: Omrežje sodelovanj med avtorji prispevkov iz Informatice in Uporabne informatike (INF&UI): celotno omrežje z 2288 vozlišči in 3530 povezavami (levo), največja komponenta omrežja z 553 vozlišči in 1599 povezavami (desno)

povezave (tj. več objav z istima avtorjema). Zanke smo izločili, saj nas zanimajo sodelovanja med raziskovalci. Pri nadaljnji analizi smo večkratne povezave upoštevali v obliki uteži na povezavah (utež predstavlja število sodelovanj med posameznima avtorjema).

V tabeli 1 so predstavljene osnovne lastnosti omrežij INF&UI, INF in UI. Analizirali smo število vozlišč in povezav omrežja (tj. število avtorjev in sodelovanj med njimi), število povezanih komponent (tj. povezano podomrežje celotnega omrežja, kjer obstaja pot po povezavah med poljubnima avtorjema) ter velikost največje in druge največje komponente (največja povezana komponenta običajno obsega večji del omrežja, npr. več kot 50 % vozlišč v omrežjih sodelovanj (Newman, 2001a)). Pogledali smo tudi, kakšna je povprečna razdalja med avtorji v omrežju (tj. število povezav med avtorji), kakšen je premer omrežja (tj. največja razdalja med avtorjema, ki jo merimo s številom povezav med njima) in nakopičenost omrežja (angl. clustering coefficient; tj.

2.2 Osnovne lastnosti omrežja

Iz dobljenih podatkov sestavimo omrežje sodelovanj tako, da vozlišča omrežja predstavljajo avtorje prispevkov, povezave med njimi pa pomenijo soavtorstvo pri vsaj eni objavi iz Informatice ali Uporabne informatike (omrežje imenujemo INF&UI). Prav tako sestavimo omrežje sodelovanj med avtorji prispevkov samo iz Informatice (INF) ter prispevkov samo iz Uporabne informatike (UI). Podrobneje smo analizirali lastnosti vseh treh omrežij.

Na sliki 1 je prikazano omrežje INF&UI brez zank (tj. objav z enim avtorjem), prikazane pa so večkratne

število trikotnikov v omrežju oziroma verjetnost, da med seboj sodelujeta sodelavca nekega avtorja).

V analizo smo zajeli 1751 objav, od tega 552 iz Uporabne informatike in 1199 iz Informatice. Različnih avtorjev, ki so objavljali v eni izmed revij, je 2288. Samo v Uporabni informatiki je objavilo prispevke 398 avtorjev, samo v Informatiki pa 1804 (86 avtorjev ima prispevke objavljene v obeh revijah).

V omrežju je v povprečju sedem povezav med parom avtorjev, kar pomeni, da se poljubna avtorja »poznata« v povprečju prek sedmih znancev. Največ povezav med dvema avtorjema je 20. Rezultati so primerljivi z rezultati podobnih študij (Newman, 2001; Perc, 2010), v katerih je povprečna razdalja v različnih omrežjih sodelovanj med 4 in 10, premer pa največ 20.

V primerjavi z omrežji iz obeh omenjenih študij so avtorji člankov iz Uporabne informatike in Informatice slabo povezani med seboj. Vseh komponent v omrežju je 704, pri čemer je 83 odstotkov vseh kompo-

ment sestavljenih iz največ treh vozlišč. Največja povezana komponenta obsega približno 24 odstotkov velikosti celotnega omrežja INF&UI, medtem ko je velikost največje komponente v omrežju INF manjša, v UI pa večja. V omrežju sodelovanj, sestavljenem iz vseh slovenskih raziskovalcev (Perc, 2010), je ta odstotek bistveno višji, saj največja komponenta obsega več kot 98 odstotkov celotnega omrežja, kar pomeni, da je velika večina raziskovalcev povezanih med seboj prek nekaj »znancev« in raziskovalci ne tvorijo veliko manjših izoliranih skupin. Za primerjavo dodajamo še podatek o velikosti druge največje komponente v omenjenem omrežju, ki obsega manj kot

odstotek celotnega omrežja (kar je običajno značilno za realna omrežja različnih tipov).

Precej visoka je nakopičenost analiziranih omrežij sodelovanj v primerjavi z omrežji iz analize sodelovanj slovenskih raziskovalcev v drugih znanstvenih disciplinah (Kronegger, Mali, Ferligoj, & Doreian, 2012; Perc, 2010). Za primer: le v sociologiji nakopičenost preseže vrednost 0,56, v fiziki, matematiki in biotehnologiji pa je manjša kot 0,50, kar nakazuje na to, da je na področju informatike velika verjetnost sodelovanja med dvema raziskovalcema, če sta v preteklosti sodelovala s skupnim znancem. Po drugi strani pa so na primer za sociologe takšna sodelovanja manj značilna.

Tabela 1: Osnovne lastnosti omrežij sodelovanj

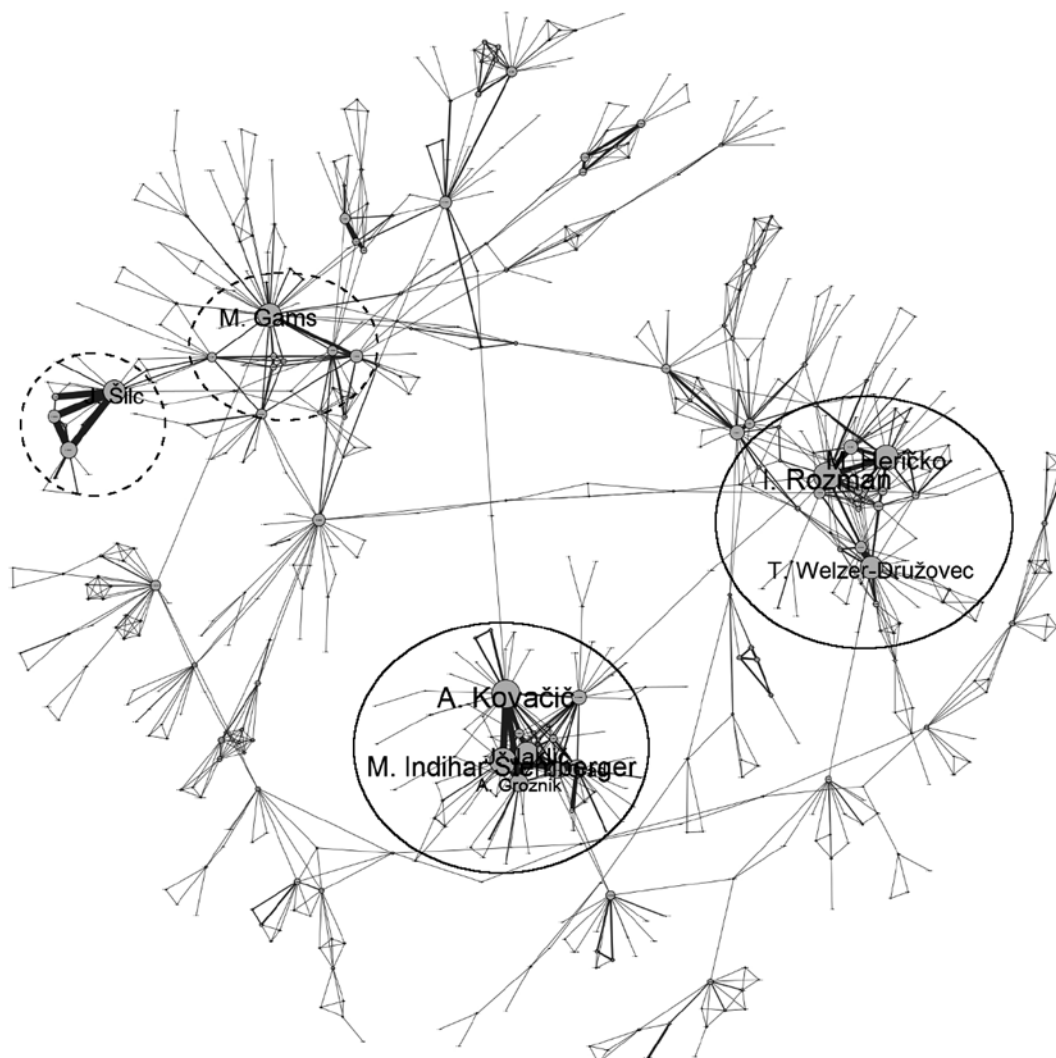
Lastnost	INF&UI	INF	UI
Število vozlišč	2288	1890	484
Število povezav (brez zank in večkratnih povezav)	2969	2455	580
Velikost največje komponente (število vozlišč, število povezav)	553, 1158	356, 732	185, 399
Velikost druge največje komponente (število vozlišč, število povezav)	27, 47	27, 47	8, 28
Premer omrežja	20	20	15
Nakopičenost omrežja	0,52	0,60	0,49
Nakopičenost največje komponente	0,37	0,40	0,45
Povprečna razdalja med avtorji v omrežju	7,1	7,4	6,2
Povprečna razdalja med avtorji v največji komponenti	7,2	7,7	6,3

V nadaljevanju se osredotočimo na največjo povezano komponento omrežja INF&UI in si oglejmo, kako je sestavljena (slika 2). Na sliki je velikost vozlišč sorazmerna z njihovo stopnjo (stopnja vozlišča pomeni število sodelovanj avtorja) in debelina povezav sorazmerna s številom sodelovanj med posameznima avtorjema. Opazimo, da izstopata dva dela omrežja (na sliki obkrožena s polno črto), ki vsebujeta večino avtorjev z visokimi stopnjami in v katerih avtorji več sodelujejo med seboj. Manjša označena dela (na sliki obkrožena s črtkano črto) vsebujeta malo avtorjev, ki pa so med seboj sodelovali pogosto (na sliki večja vozlišča ali debelejšje povezave). Prav tako opazimo, da so deli omrežja, v katerih so avtorji gosteje povezani med seboj, redkeje povezani z ostalimi avtorji ter drugimi deli omrežja. V veliko primerih je povezovalni člen med dvema gosteje povezanimi skupinama en avtor.

Podrobnejša analizo pomembnih vozlišč in delov največje povezane komponente omrežja smo predstavili v Blagus & Bajec (2014). Analizirali smo strukturo skupnosti omrežja sodelovanj INF&UI ter opazovali, kako skupnosti sovpadajo z raziskovalnimi

organizacijami, v katerih delujejo avtorji. Rezultati so pokazali, da so najpomembnejša vozlišča v omrežju pomembna glede na različne mere pomembnosti (stopnja, vmesna in dostopna središčnost). Vsak izmed desetih najpomembnejših avtorjev je bil vključen v vsaj 40 sodelovanj. Prav tako smo raziskali, v katerih raziskovalnih organizacijah delujejo avtorji iz največje komponente omrežja. Med avtorji, ki so objavljali članke v Informatiki in Uporabni informatiki, jih je 588 registriranih v sistemu SICRIS, od tega jih je 317 del največje povezane komponente omrežja. 1700 raziskovalcev nima raziskovalne šifre, med njimi so tako tujci kot tudi Slovenci, ki ne delujejo kot raziskovalci. Izmed vseh 1096 sodelovanj jih je 721 potekalo med registriranimi in neregistriranimi raziskovalci, ostala pa so sestavljena samo iz registriranih raziskovalcev. Pri tem omenimo še, da se tuji avtorji pojavljajo predvsem v manjših komponentah omrežja, manjši del jih je prisotnih v največji komponenti.

Podatke o raziskovalnih organizacijah raziskovalcev smo pridobili za 412 raziskovalcev iz celotnega



Slika 2: Pomembna vozlišča glede na stopnjo v največji povezani komponenti omrežja sodelovanj med avtorji prispevkov iz Informatice in Uporabne informatike (INF&UI)

omrežja, od tega 245 iz največje komponente. V tabeli 2 je navedenih šest raziskovalnih organizacij, ki imajo v omrežju največ raziskovalcev. V omenjeni analizi (Blagus & Bajec, 2014) smo raziskali tudi, kako sodelujejo raziskovalci znotraj posameznih organizacij in

kako med posameznimi organizacijami. Rezultati so pokazali, da raziskovalci iz posamezne organizacije veliko sodelujejo med seboj, med raziskovalnimi organizacijami pa so povezave redkeše, kar sovпада s skupnostmi (angl. communities) v omrežju.

Tabela 2: Raziskovalne organizacije s številom raziskovalcev v celotnem omrežju sodelovanja ter v njegovi največji komponenti

Raziskovalna organizacija	Zastopanost v omrežju (največji komponenti)
Fakulteta za elektrotehniko, računalništvo in informatiko, Univerza v Mariboru (FERI)	105 (79)
Inštitut Jožefa Stefana (IJS)	100 (46)
Fakulteta za računalništvo in informatiko, Univerza v Ljubljani (FRI)	51 (38)
Ekonomska fakulteta, Univerza v Ljubljani (EF)	29 (20)
Fakulteta za elektrotehniko, Univerza v Ljubljani (FE)	23 (5)
Fakulteta za matematiko in fiziko, Univerza v Ljubljani (FMF)	21 (8)

2.3 Osnovne lastnosti sodelovanj

V tabeli 3 so predstavljene osnovne lastnosti sodelovanj pri prispevkih iz obeh revij ter za vsako posebej. Približno dve tretjini vseh člankov v obeh revijah ima vsaj dva avtorja, v Uporabni informatiki je takih člankov malo več kot polovica. V povprečju sta pri eni objavi sodelovala dva avtorja; povprečje je manjše v Uporabni informatiki ter večje v Informatiki. Največ avtorjev na članek je 11. Rezultat se sklada z rezultati analize sodelovanj na različnih znanstvenih področjih (Newman, 2001), pri čemer je na področju računalništva povprečno število avtorjev na objavo 2,2. Za primerjavo: na področju fizike visokih energij pri enem članku v povprečju sodeluje devet avtorjev (Newman, 2001).

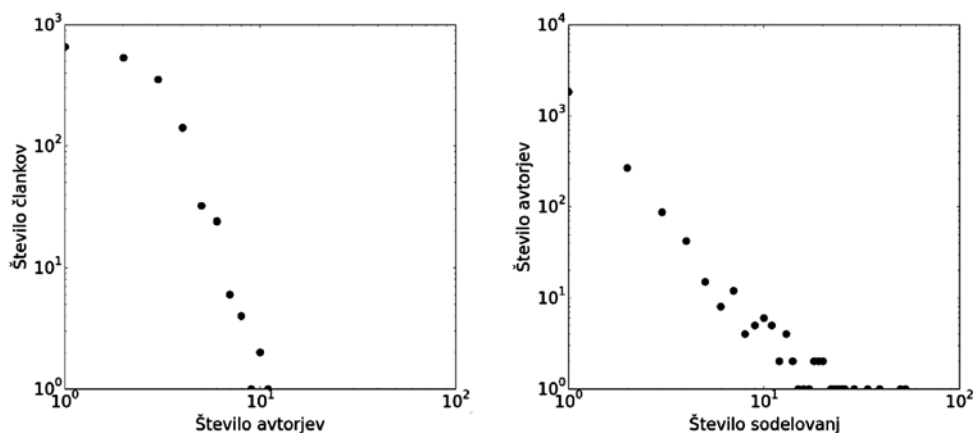
Povprečni avtor je v Uporabni informatiki in Informatiki objavil 1,64 prispevka, kar se najbolj sklada z analizo prispevkov s področja računalništva v raziskavi Newmana (2001), v kateri se izkaže, da so povprečno avtorji objavili po 2,55 prispevka. Nižje povprečje v naši raziskavi je lahko posledica manjšega števila analiziranih objav (1751 v primerjavi s 13169). Brez upoštevanja objav z enim avtorjem je povprečno število člankov na avtorja okrog

štiri. Relativno velika razlika v povprečnem številu prispevkov na avtorja z upoštevanjem enoavtorskih objav in brez njih je posledica velikega števila avtorjev, ki so sami napisali po en prispevek, in manjšega števila avtorjev, ki so napisali več prispevkov (glej tudi sliko 3).

Tabela 3: Osnovne lastnosti sodelovanj

Lastnost	INF&UI	INF	UI
Število vseh člankov	1751	1199	552
Število člankov z vsaj dvema avtorjema	1096	827	269
Povprečno število avtorjev na članek	2,1	2,3	1,8
Največ avtorjev na članek	11	10	11
Povprečno število člankov na avtorja	1,6	1,5	2,1
Povprečno število člankov na avtorja (brez enoavtorskih objav)	4,1	3,5	4,6

Na sliki 3 sta prikazani porazdelitvi števila avtorjev po članku in sodelovanj po avtorju na logaritemski skali na obeh oseh. Iz slik razberemo, da je veliko člankov z manj avtorji, malo pa je člankov z več avtorji. Še bolj izrazito je razmerje med številom sodelovanj in avtorjev, pri čemer večina avtorjev sodeluje pri manjšem številu objav.



Slika 3: Porazdelitev števila avtorjev po članku (levo) in števila sodelovanj po avtorju (desno)

3 RAZVOJ SKOZI ČAS

V drugem delu analize se osredotočimo na omrežje INF&UI ter na sodelovanja med avtorji prispevkov iz obeh revij. Opazujemo, kako se različne lastnosti omrežja in sodelovanj spreminjajo skozi čas.

3.1 Omrežje skozi čas

Slika 4 prikazuje spreminjanje lastnosti omrežja v letih 1977 do 2013. Število vozlišč omrežja ves čas

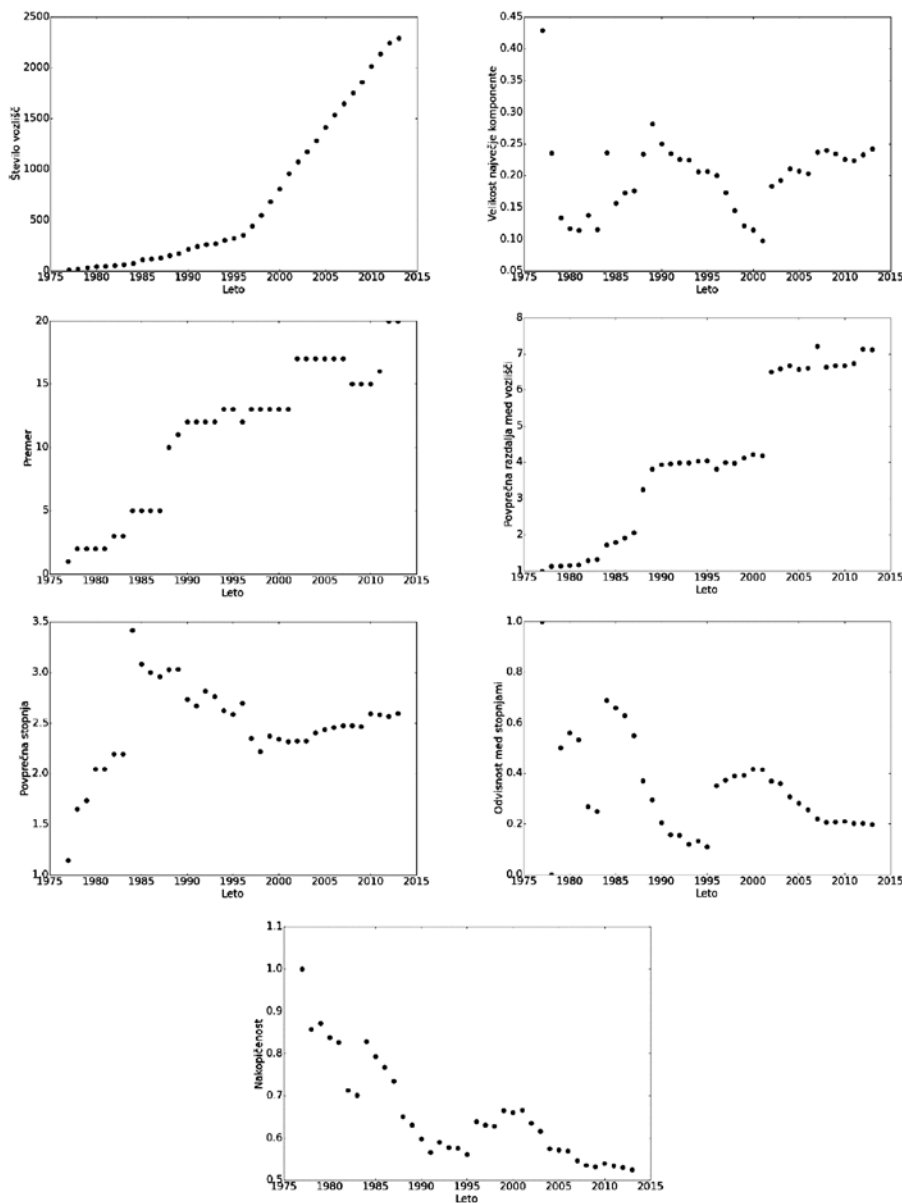
narašča, še posebno po letu 1995, kar sovpada z začetkom izhajanja Uporabne informatike. Velikost največje komponente se skozi leta spreminja in ne kaže stalnega trenda naraščanja ali padanja. Večino časa se giblje med 20 in 30 odstotki velikosti celotnega omrežja. Pri analizi omrežja sodelovanj vseh slovenskih znanstvenikov (Perc, 2010) je velikost največje komponente le v letih 1960 do 1975 (prvih 15 let opazovanja) manjša kot 40 odstotkov celotnega omrežja

ter začne po letu 1975 strmo naraščati do skoraj 99 odstotkov velikosti celotnega omrežja. To pomeni, da so raziskovalci povezani med seboj in se skoraj vsak »pozna« z vsakim drugim prek enega ali več znancev. Po drugi strani pa analizirano omrežje INF&UI vsebuje manjšo največjo povezano komponento, kar nakazuje na to, da raziskovalci raje sodelujejo v manjših povezanih skupinah.

Premer omrežja ter povprečna razdalja med avtorji kažeta podoben trend rasti z dvema skokoma, prvim po letu 1985 ter drugim po letu 2000. Oba skoka se zgo-

dita približno deset let po začetku izhajanja Informatice oziroma Uporabne informatike. Naraščanje premera omrežja in povprečne razdalje je posledica večanja samega omrežja (števila avtorjev v omrežju), hkrati pa nakazuje na to, da se novi avtorji povezujejo s starimi, novi pa z novimi sklepajo sodelovanja manj pogosto.

Povprečna stopnja vozlišč v omrežju prva leta (do 1985) narašča, kar pomeni, da so avtorji sodelovali z vedno več drugimi avtorji. Po letu 1985 začne povprečna stopnja padati, več se število prispevkov z enim avtorjem.

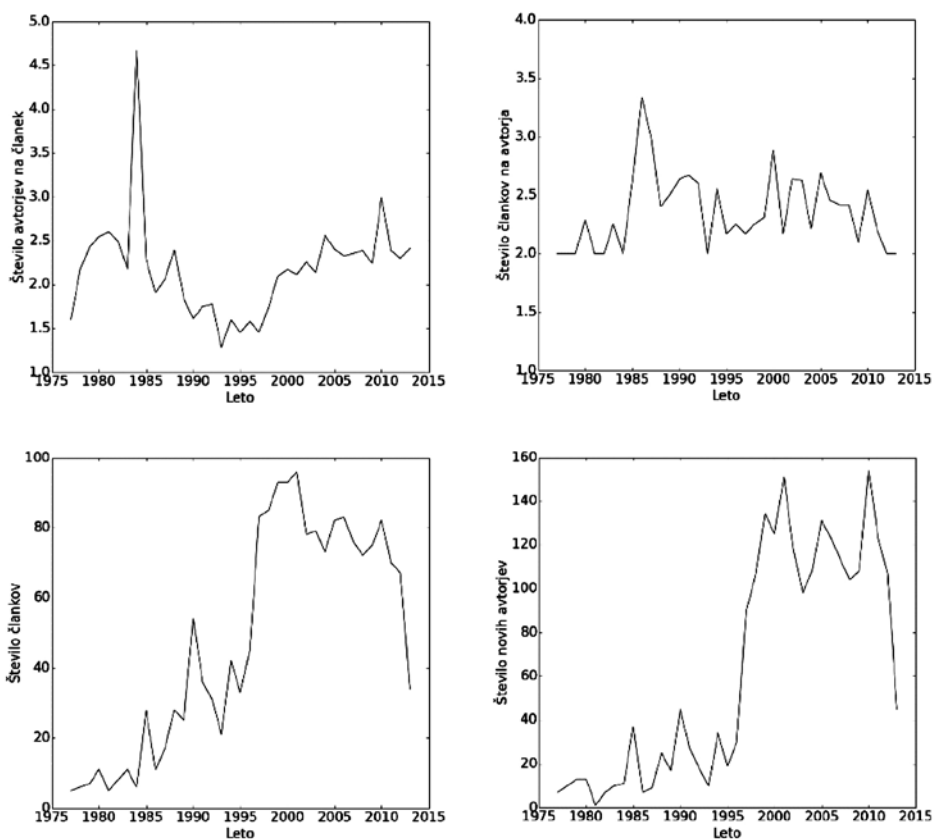


Slika 4: **Razvoj omrežja skozi čas: število vozlišč (prva vrsta levo), velikost največje komponente (prva vrsta desno), premer (druga vrsta levo), povprečna razdalja med vozlišči (druga vrsta desno), povprečna stopnja vozlišč (tretja vrsta levo), odvisnost med stopnjami vozlišč (tretja vrsta desno) ter nakopičenost omrežja (spodaj)**

Odvisnost med stopnjami (angl. degree mixing) pomeni korelacijo med stopnjami vozlišč, ki so povezane med seboj (tj. odvisnost blizu vrednosti 0 pomeni, da se vozlišča z večjo stopnjo povezujejo predvsem z vozlišči z manjšo stopnjo, odvisnost bliže vrednosti 1 pa pomeni, da se vozlišča s podobno stopnjo povezujejo med seboj). V primeru sodelovanj odvisnost med stopnjami pomeni, da se v primeru vrednosti bliže 1 med seboj povezujejo avtorji, ki sodelujejo z veliko drugimi avtorji. V primeru vrednosti bliže 0 pa je več povezav med neenakimi, tj. avtorji, ki sodelujejo z več različnimi avtorji, in tistimi, ki sodelujejo z manjšim številom drugih

avtorjev. Odvisnost med stopnjami je bila največja okrog leta 1985, kar pomeni, da so se takrat najraje povezovali enaki z enakimi. Zadnja leta odvisnost pada, kar nakazuje na porast sodelovanj med avtorji z več sodelovanji in tistimi, ki sodelujejo z manjšim številom drugih avtorjev.

Pri nakopičenosti omrežja opazimo trend padanja, vendar nakopičenost ne pade pod 0,5. Ta podatek nam pove verjetnost sodelovanja med dvema avtorjema, če sta oba sodelovala z nekom tretjim. Do leta 1980 je bila ta verjetnost nad 80 odstotki, ob začetku izhajanja Uporabne informatike se je približala 70 odstotkom, zadnja leta pa ostaja nad 50 odstotki.



Slika 6: Sodelovanja skozi čas: povprečno število avtorjev na članek (zgoraj levo), povprečno število člankov na avtorja (zgoraj desno), število člankov (spodaj levo), število novih avtorjev (spodaj desno)

3.2 Sodelovanja skozi čas

Zanimalo nas je še, kako se v letih 1977 do 2013 spreminjajo značilnosti sodelovanj. Slika 6 prikazuje spreminjanje povprečnega števila avtorjev na članek ter člankov na avtorja, števila člankov in koliko novih avtorjev se v omrežju pojavi vsako leto.

Povprečno število avtorjev na članek po strmi rasti do leta 1985, ko je pri enem članku sodelova-

lo povprečno skoraj pet avtorjev, začne padati in se približa enemu avtorju na članek v letu 1993. Po podrobnejšem pregledu objav v teh letih opazimo, da je bilo v Informatiki objavljenih letno tri do šest prispevkov, večina med njimi je imela enega ali dva avtorja. Leta 1984 pa sta bila objavljena članka s sedmimi avtorji, kar povzroči strmo rast na grafu. Uporabna informatika je na začetku izhajanja vsebovala

predvsem članke z enim ali dvema avtorjema, kar pojasni padec števila avtorjev na članek v teh letih. Kasneje začne število spet naraščati in se v zadnjih letih giblje okrog 2,5.

Število člankov na avtorja ostaja skozi leta zelo podobno, nekje med dvema in tremi, največ člankov so posamezni avtorji pisali okrog leta 1987, ko je bilo povprečje člankov na avtorja blizu 3,5. Pri spreminjanju števila člankov opazimo trend naraščanja do leta 1990, ko je bilo letno objavljenih skoraj 60 člankov. V letih 1990 do 1995 sledi padec in takoj nato strma rast, ki sovpada z ustanovitvijo Uporabne informatike leta 1992. Od leta 2000 naprej število člankov pada (letno je bilo v obeh revijah objavljenih med 70 in 100 člankov), še posebno v zadnjem letu, kar pa je lahko tudi posledica še ne vnesenih objav v sistem COBISS. Zelo podobno obnašanje je opazno pri številu novih avtorjev, ki strmo naraste po letu 1995, ko je novih avtorjev skoraj 150. V zadnjih letih se številka giblje nekje med 100 in 150, posebno malo pa je bilo novih avtorjev v letu 2013.

4 SKLEP

V prispevku smo predstavili sodelovanja med raziskovalci, ki so v letih 1977 do 2013 objavili članke v revijah *Informatica* ali *Uporabna informatika*. Sodelovanja smo analizirali v obliki omrežja sodelovanj. Ogledali smo si osnovne lastnosti treh omrežij – prvo je sestavljeno iz soavtorstev pri objavah iz obeh revij, drugo samo iz prispevkov iz Informatice ter tretje omrežje samo iz sodelovanj pri objavah v *Uporabni informatiki*. S poudarkom na opazovanju prvega in največjega omrežja smo ugotovili, da ima omrežje pričakovane lastnosti za to vrsto omrežij, izstopa pa predvsem v velikosti največje povezane komponente, ki je bistveno manjša kot v drugih podobnih omrežjih.

V drugem delu prispevka smo analizirali razvoj omrežja sodelovanj skozi čas. Pokazali smo, da število vozlišč, premer ter povprečna razdalja med vozlišči ves čas naraščajo. Po drugi strani pa nakopičenost omrežja pada. Prav tako smo raziskali spreminjanje števila člankov, avtorjev na članek, člankov na avtorja ter število novih avtorjev vsako leto. Prelomna točka pri razvoju sodelovanj je okrog leta 1995, ko začne izhajati *Uporabna informatika*. Sicer pa v zadnjih letih število člankov in novih avtorjev pada, po drugi strani pa narašča število avtorjev na članek, kar nakazuje na krepitev sodelovanj med

raziskovalci, ki so v preteklosti že sodelovali. Pri tem omenimo, da sodelovanja slovenskih raziskovalcev niso omejena le na sodelovanje pri objavi prispevkov pri analiziranih revijah, saj raziskovalci prav tako sodelujejo pri objavljanju prispevkov v tujih revijah ter na konferencah. Dokončni sklep o vzorcih sodelovanj med raziskovalci bi podala širša analiza sodelovanj tudi iz drugih virov, kar predstavlja eno glavnih možnosti nadaljnega dela.

Med analizo omrežja sodelovanj so se odprle še druge možnosti za nadaljnje raziskave. V raziskavo bi lahko vključili podrobnejše podatke o raziskovalcih ter analizirali vzorce sodelovanj med njimi z vidika lastnosti raziskovalcev (kot so npr. starost ali delovanje v različnih organizacijah). Podrobneje bi lahko raziskali povezovanja slovenskih raziskovalcev s tujimi ter obravnavali tudi drugačne oblike sodelovanj (npr. pri projektih) ali se osredotočili na sodelovanja na ravni raziskovalnih organizacij.

6 VIRI IN LITERATURA

- [1] Al, U., Soydal, Í., & Alir, G. (2012). Trends in research librarianship literature: a social network analysis of articles. *Liber quarterly*, 21, 429–444.
- [2] Barabási, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical mechanics and its applications*, 311, 590–614.
- [3] Batagelj, V., & Mrvar, A. (2000). Some analyses of Erdos collaboration graph. *Social Networks*, 22, 173–186.
- [4] Blagus, N., Bajec, M. (2014). Omrežje sodelovanj med slovenskimi raziskovalci na področju informatike. *Zbornik prispevkov 21. konference Dnevi slovenske informatike*.
- [5] Breiger, R. L. (1974). The duality of persons and groups. *Social forces*, 53, 181–190.
- [6] Carrington, P. J., Scott, J., & Wasserman, S. (2005). *Models and methods in social network analysis*. Cambridge University Press.
- [7] de Castro, R. & Grossman, J. W. (1999). Famous trails to Paul Erdős. *Mathematical Intelligencer*, 21, 51–63.
- [8] Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486, 75–174.
- [9] Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 35–41.
- [10] Grossman, J. W., & Ion, P. D. (1995). On a portion of the well-known collaboration graph. *Congressus Numerantium*, 129–132.
- [11] Hou, H., Kretschmer, H., & Liu, Z. (2008). The structure of scientific collaboration networks in Scientometrics. *Scientometrics*, 75, 189–202.
- [12] Knuth, D. (1993). *The Stanford GraphBase: A Platform for Combinatorial Computing* Addison-Wesley. Reading, MA.
- [13] Kronegger, L., Mali, F., Ferligoj, A., & Doreian, P. (2012). Collaboration structures in Slovenian scientific communities. *Scientometrics*, 90, 631–647.
- [14] Martin, T., Ball, B., Karrer, B., & Newman, M. E. J. (2013). Coauthorship and citation patterns in the Physical Review. *Physical Review E*, 88, 012814.

- [15] Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98, 404–409.
- [16] Newman, M. E. J. (2001a). Scientific collaboration networks. I. Network construction and fundamental results. *Physical review E*, 64, 016131.
- [17] Newman, M. E. J. (2001b). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical review E*, 64, 016132.
- [18] Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, 45, 167–256.
- [19] Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, 101, 5200–5205.
- [20] Newman, M. E. J. (2006). Finding community structure in networks using the eigenvectors of matrices. *Physical review E*, 74, 036104.
- [21] Newman, M. E. J. (2010). *Networks: an introduction*. Oxford University Press.
- [22] Perc, M. (2010). Growth and structure of Slovenia's scientific collaboration network. *Journal of Informetrics*, 4, 475–482.
- [23] Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, 31, 581–603.
- [24] Zhang, P.-P., Chen, K., He, Y., Zhou, T., Su, B.-B., Jin, Y., Chang, H., et al. (2006). Model and empirical study on some collaboration networks. *Physica A: Statistical Mechanics and its applications*, 360, 599–616.

■

Neli Blagus je mlada raziskovalka v laboratoriju za podatkovne tehnologije na Fakulteti za računalništvo in informatiko Univerze v Ljubljani. Raziskovalno se ukvarja z analizo omrežij.

■

Marko Bajec je izredni profesor na Fakulteti za računalništvo in informatiko Univerze v Ljubljani, kjer poučuje dodiplomske in podiplomske predmete s področja razvoja informacijskih sistemov in podatkovnih baz. Raziskovalno se ukvarja z metodami in pristopi k snovanju in razvoju informacijskih sistemov, obvladovanjem informatike ter v zadnjih letih predvsem s podatkovnimi tehnologijami za predstavitev, analizo in vizualizacijo podatkov. Leta 2009 je ustanovil laboratorij za podatkovne tehnologije ter prevzel njegovo vodenje. Je član številnih domačih in tujih združenj, komisij in odborov. V okviru fakultete je vodil več aplikativnih in raziskovalnih projektov. Svoje raziskovalne rezultate in dosežke iz prakse redno objavlja v domačih in mednarodnih znanstvenih in strokovnih krogih.