Analiza dvovrstnih omrežij in množenje omrežij

Monika Cerinšek

doktorska disertacija
predana
Fakulteti za računalništvo in informatiko
kot del izpolnjevanja pogojev za pridobitev naziva
doktor znanosti
s področja
računalništva in informatike

Ljubljana, 2015

HRUSKA.SI
SOČNE REŠITVE

Naložba v vašo prihodnost
Operacijo delno financira Evropska unija
Evropski socialni sklad

Analiza dvovrstnih omrežij in množenje omrežij

Monika Cerinšek

DOKTORSKA DISERTACIJA
PREDANA
FAKULTETI ZA RAČUNALNIŠTVO IN INFORMATIKO
KOT DEL IZPOLNJEVANJA POGOJEV ZA PRIDOBITEV NAZIVA
DOKTOR ZNANOSTI
S PODROČJA
RAČUNALNIŠTVA IN INFORMATIKE

Ljubljana, 2015

IZJAVA

Izjavljam, da sem avtor dela in da slednje ne vsebuje materiala, ki bi ga kdorkoli predhodno že objavil ali oddal v obravnavo za pridobitev naziva na univerzi ali na drugem visokošolskem zavodu, razen v primerih, kjer so navedeni viri.

— Monika Cerinšek —

januar 2015

Oddajo so odobrili

dr. Vladimir Batagelj
redni profesor za matematiko
mentor in član ocenjevalne komisije

dr. Gašper Fijavž
izredni profesor za matematiko
somentor in član ocenjevalne komisije

dr. Neža Mramor Kosta
redna profesorica za matematiko
predsednica ocenjevalne komisije

dr. Jure Leskovec
Assistant Professor of Computer Science
zunanji član ocenjevalne komisije
Stanford University

PREDHODNA OBJAVA

Izjavljam, da so bili rezultati obravnavane raziskave predhodno objavljeni/sprejeti za objavo v recenzirani reviji ali javno predstavljeni v naslednjih primerih:

1. V. Batagelj, M. Cerinšek. On bibliographic networks. *Scientometrics* 96(3): 845–864, 2013.
2. M. Cerinšek, V. Batagelj. Network analysis of Zentralblatt MATH data. *Scientometrics* 102(1): 977–1001, 2015.
3. M. Cerinšek, V. Batagelj. Semirings and Matrix Analysis of Networks. V: R. Alhajj, J. Rokne (ur.), *Encyclopedia of Social Network Analysis and Mining*, str. 1681–1687, Springer, 2014.
4. M. Cerinšek, V. Batagelj. Sources of Network Data. V: R. Alhajj, J. Rokne (ur.), *Encyclopedia of Social Network Analysis and Mining*, str. 1946–1954, Springer, 2014.
5. J. Bodlaj, M. Cerinšek, V. Batagelj. Visualization of traffic. V: V. Blondel, N. de Cordes, A. Decuyper, P, Deville, J. Raguenez, Z. Smoreda (ur.), *Mobile Phone Data for Development, Analysis of mobile phone datasets for the development of Ivory Coast*, str. 480–495, 2013.
6. M. Cerinšek, J. Bodlaj, V. Batagelj. Symbolic Clustering of Users and Antennae. V: V. Blondel, N. de Cordes, A. Decuyper, P, Deville, J. Raguenez, Z. Smoreda (ur.), *Mobile Phone Data for Development, Analysis of mobile phone datasets for the development of Ivory Coast*, str. 211–226, 2013.

V objavo je bil poslan še naslednji članek (čakamo na oceno popravljene različice):

1. M. Cerinšek, V. Batagelj. Generalized Two-mode Cores. *Social Networks*.

Rezultati so bili predstavljeni tudi na 13 mednarodnih konferencah in 7 drugih srečanjih.

Potrjujem, da sem pridobila pisna dovoljenja vseh lastnikov avtorskih pravic, ki mi dovoljujejo vključitev zgoraj navedenega materiala v pričujočo disertacijo. Potrjujem, da zgoraj navedeni material opisuje rezultate raziskav, izvedenih v času mojega podiplomskega študija na Univerzi v Ljubljani.

“Not all those who wander are lost.”
— J.R.R. Tolkien, *The Fellowship of the Ring*, 2003.

— Osebni arhiv.

Univerza *v Ljubljani*
Fakulteta *za računalništvo in informatiko*

Monika Cerinšek
Analiza dvovrstnih omrežij in množenje omrežij

POVZETEK

Tema disertacije je analiza dvovrstnih omrežij. Analize se lotimo iz dveh strani. Po eni strani razdelamo teoretično plat – metode (množenje omrežij, normalizacija omrežij in posplošene dvovrstne sredice). Po drugi strani prikažemo na primeru bibliografskih podatkov, kako lahko podatke pretvorimo v nabor usklajenih omrežij, iz katerih lahko z množenjem pridobimo različna izpeljana omrežja. Opisani pristop je uporaben veliko širše.

Pri metodah se podrobneje posvetimo množenju omrežij s poudarkom na velikih redkih omrežjih. Odgovorimo na vprašanje, kdaj je produkt dveh redkih omrežij tudi sam redko omrežje. Z uvedbo različnih polkolobarjev v množenje omrežij se odprejo novi pogledi na uporabnost produkta omrežij. Zato naredimo pregled polkolobarjev, ki so perspektivni za uporabo v analizi omrežij.

Razvita in analizirana je tudi nova metoda za iskanje pomembnih podomrežij v dvovrstnih omrežjih – posplošene dvovrstne sredice. Ta omogoča direktno analizo dvovrstnih omrežij, v katerih poiščemo podomrežja glede na izbrani lastnosti vozlišč na obeh množicah vozlišč dvovrstnega omrežja.

V disertaciji je narejen tudi pregled osnovnih pristopov za pridobivanje omrežnih podatkov. Podrobneje se posvetimo analizi bibliografskih omrežij, kjer raziščemo vpliv normalizacije v omrežjih sodelovanj in drugih izpeljanih omrežjih. Bibliografski podatki so še posebej zanimivi za analize, saj omogočajo vpogled v razvoj posameznega znanstvenega področja.

Ključne besede: omrežje, dvovrstno omrežje, veliko redko omrežje, matrika, polkolobar, množenje omrežij, izpeljana omrežja, viri omrežnih podatkov, bibliografska omrežja, omrežje sodelovanj, posplošene dvovrstne sredice, statistična analiza omrežij

University *of Ljubljana*
Faculty *of Computer and Information Science*

Monika Cerinšek
Analysis of two-mode networks and multiplication of networks

ABSTRACT

The topic of presented dissertation is the analysis of two-mode networks. We deal with the analysis from two aspects. On one hand we research the theoretical part – the methods (network multiplication, normalization of a network, and generalized two-mode cores). On the other hand we use an example of bibliographic data to present the conversion of a data into a set of compatible networks. From those we are able to get derived networks using a network multiplication. Described approach can also be used for other types of data.

We deal with the network multiplication especially multiplication of large sparse networks. We answer the question when the product of two sparse networks is also sparse. Using different semirings in a network multiplication we get new interpretations of a product of networks. Semirings that are perspective for the network analysis are presented in this dissertation.

A new method for determining important subnetworks in two-mode networks is presented – generalized two-mode cores. This method allows a direct analysis of two-mode networks and we can determine subnetworks according to different properties of links or vertices of a network with it.

In presented dissertation are also overviewed basic approaches for obtaining the network data. We made an analysis of bibliographic networks to explore the impact of normalizations in collaboration network and in other derived networks. Bibliographic data are especially interesting for the network analysis because with them we get an insight into a development of a particular scientific field.

Key words: network, two-mode network, large sparse network, matrix, semiring, network multiplication, derived networks, sources of data, bibliographic networks, collaboration network, generalized two-mode cores, statistical network analysis

ZAHVALA

Najprej bi se rada zahvalila mojemu mentorju, profesorju Vladimirju Batagelju, brez čigar pomoči ta disertacija ne bi bila mogoča. Predvsem sem hvaležna za vso podporo pri raziskovanju, pisanju člankov in ostalih prispevkov ter za moralno podporo. Seveda bi se rada zahvalila svojemu somentorju, izrednemu profesorju Gašperju Fijavžu, za diskusije, ki so mi podajale nove poglede na izzive.

Posebej bi se zahvalila tudi profesorici Neži Mramor Kosta, ki je sprejela odgovornost predsednice ocenjevalne komisije, ter profesorju Juretu Leskovcu za zasedanje sedeža zunanjega člana ocenjevalne komisije navkljub polnemu urniku in veliki potovalni razdalji med nami.

Zahvaljujem se kolektivu podjetij Hruška d.o.o. in Abelium d.o.o., ki so me sprejeli medse in mi omogočili delo mladega raziskovalca iz gospodarstva.

Posebej bi se zahvalila za moralno podporo in potrpežljivost moji družini, saj mi je omogočila pot, ki sem jo ubrala. Hvala tudi mojim prijateljem, saj mi stojijo ob strani v slabem in dobrem. Na koncu bi se pa rada zahvalila še Roku za nudenje pomoči pri premislekih in za potrpežljivost pri mojem razporejanju najinega časa.

— Monika Cerinšek, Ljubljana, januar 2015.

KAZALO

1

Uvod

1.1 Motivacija

Analiza omrežij je eden izmed načinov za analizo relacijskih podatkov. Analiziramo lahko svetovni splet, transportne sisteme, širjenje epidemije, človeški metabolizem, interakcije v skupini ljudi, električna vezja in še marsikaj drugega. Iz računalniškega vidika je analiza omrežij ena izmed uporab teorije grafov. Ta predstavlja teoretično podlago za analizo omrežij, ki ji dodamo vsebino glede na podatke, ki jih analiziramo.

V omrežju enote predstavimo z vozlišči, relacije med njimi pa s povezavami. Dvovrstna omrežja so omrežja, pri katerih je množica vozlišč sestavljena iz dveh ločenih podmnožic vozlišč, povezave pa imajo po eno krajišče v vsaki izmed njiju – vse povezave potekajo med tema množicama. Dvovrstna omrežja so posebej zanimiva, ker ponujajo alternativni omrežni pristop k analizi podatkov. Metod za neposredno analizo dvovrstnih omrežij je razmeroma malo. Zato se bom v doktorski disertaciji posvetila razvoju nekaterih metod za neposredno analizo dvovrstnih omrežij.

Graf $\mathscr{G}$ je urejen par množic $(\mathscr{V}, \mathscr{L})$ z množico *vozlišč* $\mathscr{V}$ in množico *povezav* $\mathscr{L}$. Vsaka povezava ima dve krajišči. Povezava je *usmerjena* ali *neusmerjena.*

(Navadno) *omrežje* $\mathscr{N} = (\mathscr{V}, \mathscr{L}, \mathscr{P}, \mathscr{W})$ sestoji iz grafa $\mathscr{G} = (\mathscr{V}, \mathscr{L})$, ki opisuje strukturo omrežja, in dodatnih podatkov: $\mathscr{P}$ je *množica lastnosti vozlišč* in $\mathscr{W}$ je *množica lastnosti* povezav. Kadar v navadnem omrežju dodamo pogoje glede strukture omrežja, dobimo posebne vrste omrežij.

Dvovrstno omrežje je omrežje $\mathscr{N} = ((\mathscr{I}, \mathscr{J}), \mathscr{L}, \mathscr{P}, \mathscr{W})$, ki ima množico vozlišč razdeljeno v dve ločeni množici vozlišč $\mathscr{I}$ in $\mathscr{J}$ ($\mathscr{V} = \mathscr{I} \cup \mathscr{J}, \mathscr{I} \cap \mathscr{J} = \emptyset$), vsaka povezava v $\mathscr{L}$ pa ima eno krajišče v $\mathscr{I}$ in drugo krajišče v $\mathscr{J}$.

V *večrelacijskem* omrežju $\mathscr{N} = (\mathscr{V}, \mathscr{L}, \mathscr{P}, \mathscr{W})$ množico povezav sestavlja več relacij $\mathscr{L} = (\mathscr{L}_1, \mathscr{L}_2, \dots, \mathscr{L}_r)$.

V *časovnem* omrežju $\mathscr{N} = (\mathscr{V}, \mathscr{L}, \mathscr{P}, \mathscr{W}, T)$ je dodana časovna komponenta T. Za vsako vozlišče in povezavo moramo določiti časovne intervale, v katerih so dejavni oz. prisotni. Skozi čas se lahko spreminjajo tudi lastnosti vozlišč in uteži povezav.

Vsako dvovrstno omrežje lahko predstavimo z matriko, pri čemer vozlišča v prvi podmnožici vozlišč predstavljajo vrstice matrike, vozlišča iz druge podmnožice omrežja pa stolpce matrike. Vrednosti v matriki so enake vrednostim povezav v omrežju. Če dve vozlišči nista povezani, v dano celico matrike zapišemo ničelno vrednost. Produkt omrežij je omrežje, ki ga dobimo iz matrike matričnega produkta ustreznih matrik obeh omrežij [1]. Seveda je produkt definiran le za usklajena omrežja – druga množica vozlišč prvega omrežja je enaka prvi množici vozlišč drugega omrežja. Pri množenju lahko namesto običajnega seštevanja in množenja števil uporabljamo tudi kak drug polkolobar [2–15], ki bolje odraža posebnosti problema, ki ga rešujemo.

Iz vsake podatkovne tabele (Enote × Spremenljivke) lahko za vsako spremenljivko ustvarimo dvovrstno omrežje med množicama (Enote, Zaloga vrednosti spremenljivke). Tako lahko na primer iz bibliografskih podatkov o člankih ustvarimo dvovrstna omrežja (Članki, Avtorji), (Članki, Ključne besede), (Članki, Države) idr. Ker imajo vsa ta omrežja skupno množico Enote, so v bistvu, če prvo omrežje transponiramo, usklajena. To nam omogoča, da z množenjem pridobimo cel kup izpeljanih omrežij. Če transponirano omrežje pomnožimo s samim seboj, dobimo običajno enovrstno omrežje, ki ga lahko analiziramo s standardnimi metodami analize omrežij [16].

Zgodba postane še zanimivejša, če imamo na voljo dodatna omrežja. Na primer pri člankih enovrstno omrežje $\mathscr{C}$ sklicevanj (citiranj) med članki. Tedaj, če z $\mathscr{A}$ označimo omrežje (Članki, Avtorji), je omrežje $\mathscr{A}^T \star \mathscr{C} \star \mathscr{A}$ omrežje, ki pove, kolikokrat se posamezni avtor sklicuje na drugega avtorja. Pri tem z $\mathscr{A}^T$ označimo transponirano omrežje $\mathscr{A}$ – obrnemo vrstni red obeh podmnožic vozlišč v omrežju in smeri povezav

Bibliografski podatki so še posebej zanimivi za analizo [17–23], saj je na voljo kar nekaj bibliografskih baz (Web of Science, Zentralblatt MATH, MathSciNet, Cobiss, Croatian Scientific Bibliography, Google Scholar idr.). Zberemo lahko podatke o člankih iz danega raziskovalnega področja, jih prevedemo v usklajena dvovrstna omrežja [24] in analiziramo. Iz njih lahko z množenjem pridobimo izpeljana omrežja, ki jih analiziramo dalje. Pri analizi bibliografskih podatkov lahko uporabimo standardne metode za analizo omrežij, ki pa jih moramo dopolniti s poznavanjem področja – rezultate analiz je potrebno vsebinsko razložiti.

Navadna sredica [25] je maksimalna podmnožica vozlišč v omrežju, ki imajo stopnjo znotraj sredice enako ali večjo od dane mejne vrednosti. Namesto stopnje lahko opazujemo tudi kako drugo lastnost vozlišč [26], kot so vsota vrednosti incidenčnih povezav, največja vrednost incidenčnih povezav, relativna gostota soseščine idr. V teh primerih govorimo o posplošenih sredicah. V članku [27] pa so bile za analizo dvovrstnih omrežij predlagane dvovrstne sredice. Pri definiciji posplošene sredice v dvovrstnem omrežju podamo dve mejni vrednosti – za vsako izmed obeh podmnožic vozlišč posebej, lahko pa opazujemo različni lastnosti za obe podmnožici.

Metoda posplošenih dvovrstnih sredic je uporabna, kadar iščemo močneje povezane dele dvovrstnega omrežja, pri čemer spodnjo mejo moči povezanosti določimo za vsako izmed obeh podmnožic omrežja posebej. Na primer iz podatkov o posameznih filmih lahko ustvarimo različna omrežja. Raziščemo lahko sodelovanja med režiserji in glavnimi igralci v filmih, torej režiserje in glavne igralce predstavimo s vozlišči, posamezen režiser pa je povezan s posameznim glavnim igralcem, če sta kdaj sodelovala pri snemanju filma. Vrednost povezave je enaka številu filmov, pri katerih sta sodelovala. V takšnem omrežju lahko s pomočjo dvovrstnih posplošenih sredic najdemo skupine režiserjev in glavnih igralcev, ki so sodelovali pri vsaj določenem številu filmov, pri čemer to mejo določimo posebej za režiserje in posebej za glavne igralce.

Pričujoča disertacija je torej razdeljena v tri sklope: množenje in polkolobarji, viri podatkov in bibliografska omrežja ter posplošene dvovrstne sredice. Vsi trije sklopi so se razvijali neodvisno eden od drugega. Kljub temu se vsebine sklopov vsaj delno prepletajo. Hitro množenje redkih omrežij iz prvega sklopa smo uporabili v analizah bibliografskih omrežij. V le-teh analizah smo prikazali metode za določanje pomembnih podomrežij, čemur je namenjena tudi metoda posplošenih dvovrstnih sredic. Le seznam polkolobarjev iz prvega sklopa se ne prepleta neposredno z ostalima sklopoma, saj smo se v analizah bibliografskih omrežij bolj posvetili normalizacijam omrežij. Uporaba različnih polkolobarjev v analizi omrežij je del načrtovanega prihodnjega dela.

1.2 Znanstveni prispevki

Kot že opisano, običajni pristop k analizi dvovrstnih omrežij temelji na pretvorbi dvovrstnega omrežja v navadno (enovrstno) omrežje, na katerem potem uporabimo poznane metode za analizo omrežij. S tem lahko izgubimo nekaj informacij o omrežju, zato sem to premostila z razvojem novih in s posplošitvijo ter razdelavo že obstoječih metod za neposredno analizo (velikih) dvovrstnih omrežij.

Na področju uporabe različnih polkolobarjev pri analizi omrežij je bilo narejenega že veliko dela. Moj prispevek na tem področju je razvoj novih uporab množenja omrežij nad ustreznimi polkolobarji pri reševanju različnih problemov analize omrežij. Pri tem sem si pomagala s podrobnim pregledom tovrstnih primerov, ki sem ga naredila na podlagi že obstoječih polkolobarjev. Posplošen in analiziran je bil postopek za hitro množenje redkih omrežij.

Na primeru bibliografskih omrežij so razdelane splošne metode za analizo družin omrežij, pridobljenih iz podatkovnih tabel. Tu nam množenje omrežij omogoča pridobivanje izpeljanih omrežij. Pokazan pa je tudi velik pomen ustrezne normalizacije omrežij za pridobitev smiselnih rezultatov.

Razvila in analizirala sem tudi metodo posplošenih dvovrstnih sredic. Le-te lahko uporabimo pri iskanju pomembnih podomrežij v dvovrstnem omrežju glede na izbrano(i) lastnost(i) vozlišč.

Za navedene probleme sem poskušala razviti podkvadratične (skoraj linearne) algoritme, ki jih je mogoče uporabljati tudi na velikih omrežjih, ki jih vse pogosteje srečujemo v praksi.

1.3 Metodologija

Metodologija raziskovalnega dela je zajemala predvsem študij literature, samostojno raziskovalno in razvojno delo ter programsko izvedbo in preizkušanje metod.

Preden smo posamezno metodo programsko izvedli, smo jo teoretično razdelali – njeno pravilnost in učinkovitost. Ob samem razvoju metode je zato dokazana njeno pravilnost in ocena učinkovitosti z analizo časovne zahtevnosti metode. Uporabnost metod je prikazana na primerih uporabe na realnih dvovrstnih omrežjih.

Zapiske o pravilnosti, učinkovitosti in uporabnosti metode smo zapisovali v digitalne dokumente. Del teh dokumentov smo preuredili v članke za objavo v znanstveni reviji. Vsi digitalni zapiski pa so povzeti in združeni v tej disertaciji.

Prototipna programska izvedba metod je potekala v programskem jeziku Python (verzija 2.7), končne različice metod pa so delno že sprogramirane tudi v programskem jeziku C++. Normalizacije omrežij so bile pripravljene z makroji v programu `Pajek` [28–30].

Uporabnost novo razvitih metod je prikazana z analizami velikih dvovrstnih omrežij, pridobljenih iz bibliografskih baz Web of Science in Zentralblatt MATH. Analize omrežij z že poznanimi metodami so bile opravljene s programom za analizo in prikaz velikih omrežij `Pajek` in s programskim jezikom R.

1.4 Pregled disertacije

Skozi celotno disertacijo se ukvarjamo z dvema pogledoma na analizo omrežij s poudarkom na analizi dvovrstnih omrežij. Predstavljena je tako analiza konkretnih podatkov kot tudi nove metode za analizo podatkov.

V poglavju 2 obravnavamo množenje omrežij. Pri tem se posvetimo množenju redkih omrežij, v katerih matrikah je veliko ničel in je zato običajno množenje preveč potratno. V analizi omrežij pa se vedno išče nove načine analize obstoječih podatkov, zato se v tem poglavju posvetimo tudi uporabi različnih polkolobarjev v množenju omrežij. Opisani so kombinatorični polkolobar, polkolobar regularnih izrazov, polkolobar najkrajših poti, polkolobar dosegljivosti in mnogi drugi.

V poglavju 3 je najprej opisanih nekaj najpogostejših pristopov k zbiranju podatkov oziroma virov podatkov: od opazovanja in vprašalnikov do semantičnih in naključnih omrežij. Nato sledita primera analize podatkov z metodami analize omrežij. V analizi bibliografskih podatkov iz podatkovne baze Web of Science so opisane različne izpeljave omrežij s pomočjo množenja omrežij in različnih normalizacij omrežij. V analizi bibliografskih podatkov iz podatkovne baze Zentralblatt MATH pa je podrobneje opisana struktura vhodnih podatkov, statistična analiza podmnožic pridobljenih dvovrstnih omrežij, analiza omrežja sodelovanj in vse te analize, ponovljene na podmnožici vhodnih podatkov, ki se nanaša na teorijo grafov.

V poglavju 4 je opisana nova metoda za neposredno analizo dvovrstnih omrežij. Podana je definicija posplošenih dvovrstnih sredic, lastnosti vozlišč, ki jih lahko uporabimo pri tej metodi, lastnosti dobljenih podomrežij in sam algoritem metode posplošenih dvovrstnih sredic.

1.4.1 Oznake

- $\mathbb{N} = \{0, 1, 2, 3, ...\}$ množica naravnih števil,
- $\overline{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$ množica naravnih števil s pozitivno neskončnostjo,
- $\mathbb{R}$ množica realnih števil,
- $\mathbb{R}^+$ množica pozitivnih realnih števil,
- $\mathbb{R}_0^+$ množica nenegativnih realnih števil,
- $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ množica realnih števil z negativno in pozitivno neskončnostjo,

Dodatno

- $\exists$ matematična oznaka za 'obstaja',
- $\forall$ matematična oznaka za 'vsak',
- $|\mathbf{A}|$ matematična oznaka za velikost množice $\mathbf{A}$,

- $\mathscr{P}(\mathbf{A})$ matematična oznaka za množico vseh podmnožic množice **A** (potenčna množica).

Pojmi iz analize omrežij so prevedeni s pomočjo slovarja, ki je na voljo na spletni strani `http://zvonka.fmf.uni-lj.si/netbook/doku.php?id=net:dic`.

2

Množenje omrežij in polkolobarji

Polkolobar je algebrska struktura z dvema operacijama, ustrezna za opis računanja z naravnimi števili in matrikami pa tudi za probleme določanja vrednosti sprehodov v omrežjih. Številni rezultati in algoritmi iz različnih področij uporabe teh izzivov se izkažejo za posebne primere uporabe ustreznih polkolobarjev.

2.1 Polkolobarji

Naj bo $\mathbb{K}$ množica ter a, b in c poljubni elementi iz te množice. *Polkolobar* [2, 4, 5] je algebrska struktura $(\mathbb{K}, \oplus, \odot, 0, 1)$ z dvočlenima operacijama seštevanja $\oplus$ in množenja $\odot$, za katero velja:

- $(\mathbb{K}, \oplus)$ je Abelov monoid z nevtralnim elementom 0 (*ničla*):

  komutativnost: $a \oplus b = b \oplus a$

  asociativnost: $(a \oplus b) \oplus c = a \oplus (b \oplus c)$

  obstoj ničle: $a \oplus 0 = a$

- $(\mathbb{K}, \odot)$ je monoid z nevtralnim elementom 1 (*enota*):

  asociativnost: $(a \odot b) \odot c = a \odot (b \odot c)$

  obstoj enote: $a \odot 1 = 1 \odot a = a$

- veljata distributivnostna zakona:

  $a \odot (b \oplus c) = a \odot b \oplus a \odot c$

  $(b \oplus c) \odot a = b \odot a \oplus c \odot a$

Pri zapisu izrazov privzamemo, da ima množenje prednost pred seštevanjem.

Opomba: Velja:

1. Operacija $\odot$ ni komutativna, torej se vrstni red povezav ohranja.
2. Ničlo in enoto polkolobarja lahko predstavlja tudi isti element množice polkolobarja.
3. Komutativnost in asociativnost $(\mathbb{K}, \oplus)$ pomeni, da ni pomembno, v kakšnem vrstnem redu gledamo na že izračunane uteži poti.

4. Nevtralni element $(\mathbb{K}, \oplus)$ predstavlja neobstoječo pot.

5. Asociativnost $(\mathbb{K}, \odot)$ pove, da ni pomembnen vrstni red računanja vrednosti sprehoda.

6. Če nevtralni element za katero operacijo ne obstaja, ga lahko dodamo kot element, za katerega velja za vsak $a \in \mathbb{K} : a \oplus 0 = a = 0 \oplus a$ in $a \odot 0 = 0 \odot a = 0$ oz. $a \odot 1 = 1 \odot a = a$.

7. Distributivnost pride v upoštev pri notranje (po povezavah) neločenih poteh, sicer nima pomena. Povezavo, ki se pojavi v več poteh hkrati, izračunamo le enkrat in ta izračun upoštevamo pri vseh naslednjih poteh, ki to povezavo vsebujejo.

Polkolobar $(\mathbb{K}, \oplus, \odot, 0, 1)$ je *poln* natanko takrat, ko je seštevanje dobro definirano za števne množice in zanj veljajo komutativnost, asociativnost in distributivnost. Te lastnosti v primeru števnih množic posplošimo. Distributivnost velja v naslednji obliki

$$\left(\bigoplus_{i\in I} a_i\right) \odot \left(\bigoplus_{j\in J} b_j\right) = \bigoplus_{i\in I}\left(\bigoplus_{j\in J}(a_i \odot b_j)\right) = \bigoplus_{i,j\in I\times J}(a_i \odot b_j).$$

Primer polnega polkolobarja je najpogosteje uporabljani polkolobar – *kombinatorični polkolobar* $(\overline{\mathbb{N}}, +, \cdot, 0, 1)$. Obstaja več različic kombinatoričnega polkolobarja glede na izbrano množico elementov: $\mathbb{N}, \overline{N}, \mathbb{R}, \mathbb{R}_0^+, \mathbb{Q}$.

Operacija seštevanja je *idempotentna* natanko takrat, ko velja $a \oplus a = a$ za vsak $a \in \mathbb{K}$. V tem primeru je polkolobar nad končno množico $\mathbb{K}$ poln.

Polkolobar $(\mathbb{K}, \oplus, \odot, 0, 1)$ je *zaprt*[31] natanko takrat, ko v njem obstaja dodatna (enočlena) operacija *zaprtja* *, za katero velja

$$a^* = 1 \oplus a \odot a^* = 1 \oplus a^* \odot a$$

za vsak $a \in \mathbb{K}$. V danem polkolobarju lahko obstaja več različnih zaprtij. Primer zaprtega polkolobarja je različica kombinatoričnega polkolobarja $(\mathbb{R}, +, \cdot, 0, 1)$.

Poln polkolobar je vedno zaprt za

$$a^* = \bigoplus_{k\in\mathbb{N}} a^k.$$

V zaprtem polkolobarju lahko definiramo *strogo zaprtje* $\bar{a}$ kot

$$\bar{a} = a \odot a^*.$$

V polkolobarju $(\mathbb{K}, \oplus, \odot, 0, 1)$ drži *absorpcijski zakon* natanko takrat, ko za katerekoli $a, b, c \in \mathbb{K}$ velja

$$a \odot b \oplus a \odot c \odot b = a \odot b.$$

Za veljavnost absorpcijskega zakona v polkolobarju je zaradi distributivnosti dovolj preveriti, ali za vsak $a \in \mathbb{K}$ velja $1 \oplus a = 1$.

2.2 Matrike

Matrika $\mathbf{A}$ velikosti $m \times n$ nad množico $\mathbb{K}$ je pravokotna tabela elementov iz množice $\mathbb{K}$, ki vsebuje m vrstic in n stolpcev. Vrednost v i-ti vrstici in j-tem stolpcu označimo z a_{ij}. Če je $m = n$, matriki $\mathbf{A}$ rečemo *kvadratna* matrika.

Matrika s samo ničelnimi vrednostmi je *ničelna* matrika in jo označimo z $\mathbf{0}_{mn}$. *Diagonalna* matrika je kvadratna matrika $\mathbf{A}$, v kateri so lahko le diagonalne vrednosti neničelne: $a_{ij} = 0$ za $i \neq j$. Če v diagonalni matriki velja $a_{ii} = 1, i = 1, \dots, n$, takšni matriki rečemo *enotska* matrika $\mathbf{I}_n$ reda n. Matrika $\mathbf{A}$ je *zgornje trikotna*, če velja $a_{ij} = 0$ za $i > j$, njej transponirana matrika pa je *spodnje trikotna*.

Naj bo $\mathscr{M}_{mn}(\mathbb{K})$ množica matrik reda $m \times n$ nad polkolobarjem $(\mathbb{K}, \oplus, \odot, 0, 1)$. Predpostavimo, da velja

$$\forall a \in \mathbb{K} : a \odot 0 = 0 \odot a = 0.$$

Naj bo $\mathscr{M}(\mathbb{K})$ množica vseh matrik nad $\mathbb{K}$. Operaciji $\oplus$ in $\odot$ lahko razširimo na $\mathscr{M}(\mathbb{K})$:

$$\mathbf{A}, \mathbf{B} \in \mathscr{M}_{mn}(\mathbb{K}) : \mathbf{A} \oplus \mathbf{B} = [a_{uv} \oplus b_{uv}]_{mn} \in \mathscr{M}_{mn}(\mathbb{K}),$$

$$\mathbf{A} \in \mathscr{M}_{mk}(\mathbb{K}), \mathbf{B} \in \mathscr{M}_{kn}(\mathbb{K}) : \mathbf{A} \odot \mathbf{B} = [\bigoplus_{t=1}^{k} a_{ut} \odot b_{tv}]_{mn} \in \mathscr{M}_{mn}(\mathbb{K}).$$

Potem velja:

- $(\mathscr{M}_{mn}(\mathbb{K}), \oplus, \mathbf{O}_{mn})$ je Abelov monoid,
- $(\mathscr{M}_{nn}(\mathbb{K}), \odot, \mathbf{I}_n)$ je monoid,

- $(\mathscr{M}_{nn}(\mathbb{K}), \oplus, \odot, \mathbf{O}_n, \mathbf{I}_n)$ je polkolobar.

Matrični polkolobar nad polnim polkolobarjem je tudi poln. Matrični polkolobar je zaprt za $\mathbf{A}^* = \bigoplus_{k=0}^{\infty} \mathbf{A}^k$. Zaprtje matrike dobimo tako, da izračunamo vrednosti sprehodov med posameznimi vozlišči.

Transponirana matrika matrike $\mathbf{A}$ je matrika $\mathbf{A}^T$, v kateri so vrstice matrike $\mathbf{A}$ zapisane kot stolpci matrike $\mathbf{A}^T : a_{ij}^T = a_{ji}$. Za matriki $\mathbf{A}$ in $\mathbf{B}$ velja

$$(\mathbf{A} \odot \mathbf{B})^T = \mathbf{B}^T \odot \mathbf{A}^T.$$

Kvadratna matrika $\mathbf{A}$ je *simetrična,* če velja $\mathbf{A} = \mathbf{A}^T$. Definirajmo še *simetrično zaprtje* matrike: $\mathbf{A}^\bullet = (\mathbf{A} \oplus \mathbf{A}^T)^*$.

2.3 Množenje omrežij

(Enostaven usmerjen) *graf* $\mathscr{G}$ je urejen par množic $(\mathscr{V}, \mathscr{A})$, pri čemer je $\mathscr{V}$ množica vozlišč in $\mathscr{A} \subseteq \mathscr{V} \times \mathscr{V}$ množica usmerjenih povezav. Privzamemo, da je množica vozlišč končna: $\mathscr{V} = \{v_1, v_2, \dots, v_n\}$. Naj bo $\mathscr{N} = ((\mathscr{I}, \mathscr{J}), \mathscr{A}, w)$ *enostavno dvovrstno omrežje,* pri čemer sta $\mathscr{I}$ in $\mathscr{J}$ ločeni (pod)množici vozlišč ($\mathscr{V} = \mathscr{I} \cup \mathscr{J}, \mathscr{I} \cap \mathscr{J} = \emptyset$), $\mathscr{A}$ je množica usmerjenih povezav od $\mathscr{I}$ do $\mathscr{J}$, preslikava $w : \mathscr{A} \to \mathbb{K}$ pa je *vrednost povezav,* ki ji pravimo tudi *utež.* Omrežju priredimo njegovo *vrednostno matriko* $\mathbf{W} = w_{(i,j)}$ z vrednostmi

$$w_{ij} = \begin{cases} w((i,j)) & (i,j) \in \mathscr{A}, \\ 0 & \text{sicer.} \end{cases}$$

V računalniški uporabi vrednostnih matrik njihova velikost lahko predstavlja velik problem. Vrednostne matrike velikih omrežij so običajno redke. Ker ni nobene potrebe po shranjevanju ničelnih elementov matrike, obstaja več različnih podatkovnih struktur za shranjevanje in delo z vrednostnimi matrikami: posebni slovarji in seznami. Zato pojem množenja matrik prenesemo tudi na omrežja.

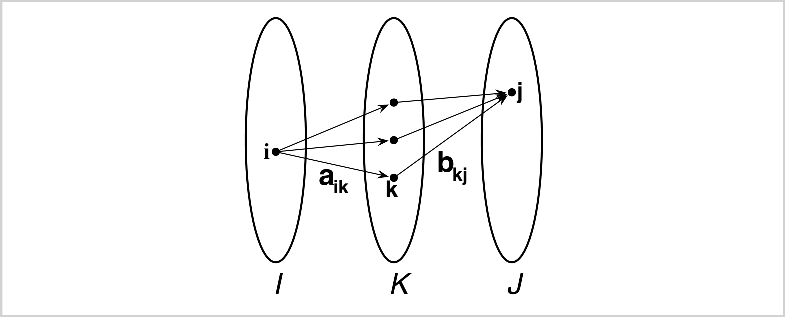
Naj bosta $\mathscr{N}_\mathbf{A} = ((\mathscr{I}, \mathscr{K}), \mathscr{A}_\mathbf{A}, w_\mathbf{A})$ in $\mathscr{N}_\mathbf{B} = ((\mathscr{K}, \mathscr{J}), \mathscr{A}_\mathbf{B}, w_\mathbf{B})$ dvovrstni omrežji s pripadajočima matrikama $\mathbf{A}_{\mathscr{I} \times \mathscr{K}}$ in $\mathbf{B}_{\mathscr{K} \times \mathscr{J}}$. Predpostavimo, da velja $w_\mathbf{A} : \mathscr{A}_\mathbf{A} \to \mathbb{K}$, $w_\mathbf{B} : \mathscr{A}_\mathbf{B} \to \mathbb{K}$ in da je $(\mathbb{K}, \oplus, \odot, 0, 1)$ polkolobar. Takšni omrežji oz. matriki sta *usklajeni* – druga množica prvega omrežja je enaka prvi množici drugega omrežja. *Produkt* $\mathscr{N}_\mathbf{A} \star \mathscr{N}_\mathbf{B}$ omrežij $\mathscr{N}_\mathbf{A}$ in $\mathscr{N}_\mathbf{B}$ je omrežje $\mathscr{N}_\mathbf{C} = ((\mathscr{I}, \mathscr{J}), \mathscr{A}_\mathbf{C}, w_\mathbf{C})$ za

$\mathscr{A}_\mathbf{C} = \{(i,j); i \in \mathscr{I}, j \in \mathscr{J}, c_{ij} \neq 0\}$ in $w_\mathbf{C}(i,j) = c_{ij}$ za $(i,j) \in \mathscr{A}_\mathbf{C}$, pri čemer je $\mathbf{C} = [c_{ij}]_{\mathscr{I}\times\mathscr{J}} = \mathbf{A} \odot \mathbf{B}$. Če so vse tri množice vozlišč enake ($\mathscr{I} = \mathscr{K} = \mathscr{J}$), imamo opravka z navadnimi (enovrstnimi) omrežji, ki jim pripadajo kvadratne matrike.

Kako so določene usmerjene povezave v produktnem omrežju? Oglejmo si definicijo vrednosti povezave v produktu matrik:

$$c_{ij} = \bigoplus_{k\in\mathscr{K}} a_{ik} \odot b_{kj}.$$

Usmerjena povezava $(i,j) \in \mathscr{A}_\mathbf{C}$ obstaja, če je vrednost c_{ij} neničelna. Torej mora biti



Slika 2.1

Množenje omrežij.

vsaj ena izmed vrednosti $a_{ik} \cdot b_{kj}$ v vsoti neničelna. To pomeni, da morata biti tako a_{ik} kot tudi b_{kj} neničelna za vsaj en $k \in \mathscr{K}$, torej mora veljati $(i,k) \in \mathscr{A}_\mathbf{A}$ in $(k,j) \in \mathscr{A}_\mathbf{B}$ za vsaj en $k \in \mathscr{K}$ (Slika 2.1). Vrednost v produktu matrik lahko torej napišemo kot

$$c_{ij} = \bigoplus_{k\in N_\mathbf{A}(i)\cap N^-_\mathbf{B}(j)} a_{ik} \odot b_{kj}.$$

kjer je $N_\mathbf{A}(i)$ množica *naslednikov* vozlišča i v omrežju $\mathscr{N}_\mathbf{A}$ in $N^-_\mathbf{B}(j)$ množica *prednikov* vozlišča j v omrežju $\mathscr{N}_\mathbf{B}$. Vrednost elementa c_{ij} je enaka vsoti vrednosti vseh poti (dolžine 2) od $i \in \mathscr{I}$ do $j \in \mathscr{J}$ skozi vozlišča $k \in \mathscr{K}$.

Predpostavimo, da operaciji polkolobarja zahtevata $O(1)$ časa za izvedbo. Standardni proces za izračun produkta matrik $\mathbf{A}_{\mathscr{I}\times\mathscr{K}}$ in $\mathbf{B}_{\mathscr{K}\times\mathscr{J}}$ zahteva $O(|\mathscr{I}|\cdot|\mathscr{K}|\cdot|\mathscr{J}|)$ časa in je zato prepočasen za uporabo na velikih omrežjih. Ker so matrike velikih omrežij ponavadi redke, lahko produkt dveh takih omrežij izračunamo veliko hitreje z upoštevanjem le neničelnih vrednosti. Predlaganih je bilo že nekaj algoritmov za množenje redkih omrežij [32–34]. Mi predlagamo algoritem, v katerem upoštevamo strukturo omrežja [7, 35], kakor je prikazano v algoritmu 2.1.

Algoritem 2.1

Algoritem za hitro množenje redkih matrik.

for $k \in \mathcal{K}$ *do*
  for $i \in N^-_{\mathbf{A}}(k)$ *do*
    for $j \in N_{\mathbf{B}}(k)$ *do*
      if $\exists c_{ij}$ *then*
        $c_{ij} \leftarrow c_{ij} \oplus a_{ik} \odot b_{kj}$
      else
        $c_{ij} \leftarrow a_{ik} \odot b_{kj}$
      end if
    end for
  end for
end for

Redkost omrežja oziroma matrike je definirana v odvisnosti od konteksta. Ponavadi privzamemo, da je omrežje redko, kadar je njegovo število povezav linearno odvisno od števila vozlišč: $m \approx cn$ za relativno majhno konstanto c. V splošnem je množenje velikih redkih omrežij nevarna operacija, saj lahko rezultat raznese – dobljeno omrežje ni nujno redko.

V algoritmu za množenje omrežij opazimo, da vsako vmesno vozlišče $k \in \mathcal{K}$ doda k produktnemu omrežju poln dvovrstni podgraf $K_{N^-_{\mathbf{A}}(k),N_{\mathbf{B}}(k)}$ (oz. v primeru $\mathbf{A} = \mathbf{B}$ poln podgraf $K_{N(k)}$). Če sta stopnji $\deg_A(k) = |N^-_{\mathbf{A}}(k)|$ in $\deg_{\mathbf{B}}(k) = |N_{\mathbf{B}}(k)|$ veliki, potem ima že izračun tega polnega podgrafa kvadratično (časovno in prostorsko) zahtevnost v odvisnosti od vhodnih podatkov – rezultat raznese.

Hitro opazimo, da če ima vsaj eno izmed obeh redkih omrežij $\mathcal{N}_{\mathbf{A}}$ in $\mathcal{N}_{\mathbf{B}}$ majhno največjo stopnjo v množici $\mathcal{K}$, je tudi rezultat množenja $\mathcal{N}_{\mathbf{C}}$ redko omrežje.

Pokažimo, da velja še močnejši rezultat: če v množici $\mathcal{K}$ redkih omrežjih $\mathcal{N}_{\mathbf{A}}$ in $\mathcal{N}_{\mathbf{B}}$ obstaja le nekaj vozlišč z visoko stopnjo in nobeno izmed teh nima visoke stopnje v obeh omrežjih, je produkt teh dveh omrežij $\mathcal{N}_{\mathbf{C}}$ tudi redko omrežje.

Za $k \in \mathcal{K}$ označimo

$$d_{min}(k) = \min(\deg_{\mathbf{A}}(k), \deg_{\mathbf{B}}(k)) \quad \text{in} \quad d_{max}(k) = \max(\deg_{\mathbf{A}}(k), \deg_{\mathbf{B}}(k)).$$

Potem velja

$$\deg_{\mathbf{A}}(k) \cdot \deg_{\mathbf{B}}(k) = d_{min}(k) \cdot d_{max}(k).$$

Definirajmo še $\Delta_{min} = \max_{k \in \mathcal{K}} d_{min}(k)$ in

$$\mathcal{K}(d) = \{k \in \mathcal{K} : d_{max}(k) \geq d\}.$$

Označimo $d^* = argmin_d(|\mathcal{K}(d)| \leq d)$ in $\mathcal{K}^* = \mathcal{K}(d^*)$. Potem velja $|\mathcal{K}^*| \leq d^*$ in za število neničelnih elementov v produktu velja

$$C \leq \sum_{k \in \mathcal{K}} \deg_{\mathbf{A}}(k) \cdot \deg_{\mathbf{B}}(k) = \sum_{k \in \mathcal{K}} d_{min}(k) \cdot d_{max}(k)$$

$$= \sum_{k \in \mathcal{K}^*} d_{min}(k) \cdot d_{max}(k) + \sum_{k \in \mathcal{K} \setminus \mathcal{K}^*} d_{min}(k) \cdot d_{max}(k)$$

$$\leq \Delta_{min} \cdot \sum_{k \in \mathcal{K}^*} d_{max}(k) + d^* \cdot \sum_{k \in \mathcal{K} \setminus \mathcal{K}^*} d_{min}(k)$$

$$\leq d^* \cdot (\Delta_{min} \cdot \max(|\mathcal{I}|, |\mathcal{J}|) + \min(|\mathcal{A}_{\mathbf{A}}|, |\mathcal{A}_{\mathbf{B}}|))$$

Torej velja:

> Če sta vrednosti Δ_{min} in d^* majhni za redki omrežji $\mathcal{N}_{\mathbf{A}}$ in $\mathcal{N}_{\mathbf{B}}$, potem je tudi njuno produktno omrežje $\mathcal{N}_{\mathbf{C}}$ redko.

Ta trditev je ekvivalentna gornjemu rezultatu.

2.4 Algebrajski problem poti

Uporaba pravega polkolobarja in množenja omrežij nas lahko vodita v samo bistvo problema najkrajše poti [4]. Tudi mnogo drugih problemov na omrežjih je rešljivih z zamenjavo običajnega seštevanja in množenja z operacijama iz primernega polkolobarja.

Naj bo $\mathcal{N} = (\mathcal{V}, \mathcal{A}, w)$ omrežje, za katero velja, da je $w : \mathcal{A} \to \mathbb{K}$ vrednost (utež) usmerjenih povezav in je $(\mathbb{K}, \oplus, \odot, 0, 1)$ polkolobar. Število vozlišč označimo z $n = |\mathcal{V}|$ in število usmerjenih povezav označimo z $m = |\mathcal{A}|$.

Končno zaporedje vozlišč $\sigma = (u_0, u_1, u_2, \dots, u_{p-1}, u_p)$ je *sprehod dolžine* p v $\mathcal{N}$ natanko takrat, ko je vsak par sosednjih vozlišč povezan: $(u_{i-1}, u_i) \in \mathcal{A}$ za $i = 1, \dots, p$.

Končno zaporedje σ je *polsprehod* ali veriga v $\mathscr{N}$ natanko takrat, ko je vsak par sosednjih vozlišč povezan ne glede na usmerjenost povezav: $(u_{i-1}, u_i) \in \mathscr{A} \vee (u_i, u_{i-1}) \in \mathscr{A}$ za $i = 1, \dots, p$. (Pol)sprehod je *sklenjen* natanko takrat, ko sta njegovi končni vozlišči isti: $u_0 = u_p$. Sprehod je *enostaven* natanko takrat, ko se vsako vozlišče v zaporedju pojavi le enkrat. Enostavnemu sprehodu pravimo tudi *pot*. *Obhod* je sprehod, v katerem krajišči sovpadata, razen tega pa nima drugih podvojenih vozlišč.

Vrednost w lahko razširimo na sprehode in množice sprehodov v $\mathscr{N}$ z naslednjimi pravili (Slika 2.2):

- Če je $\sigma_v = (v)$ ničeln sprehod v vozlišču $v \in \mathscr{V}$, potem je $w(\sigma_v) = 1$.
- Če je $\sigma = (u_0, u_1, u_2, \dots, u_{p-1}, u_p)$ sprehod dolžine $p \geq 1$ v $\mathscr{N}$, potem je

$$w(\sigma) = \bigodot_{i=1}^{k} w(u_{i-1}, u_i).$$

- Za prazno množico sprehodov $\emptyset$ velja $w(\emptyset) = 0$.
- Če je $\mathscr{S} = \{\sigma_1, \sigma_2, \dots\}$ množica sprehodov v $\mathscr{N}$, potem je (če vsota obstaja)

$$w(\mathscr{S}) = \bigoplus_{\sigma \in \mathscr{S}} w(\sigma).$$

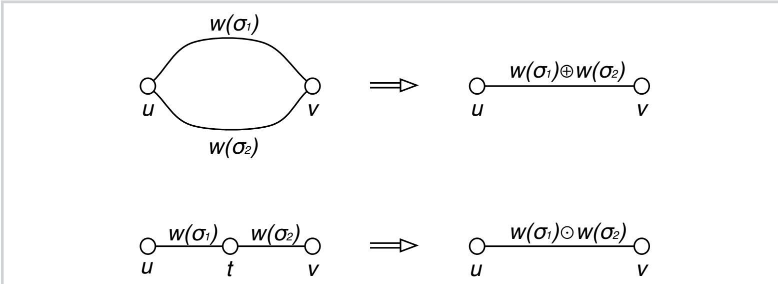
Naj bosta σ_1 in σ_2 usklajena sprehoda v $\mathscr{N}$ – zadnje vozlišče sprehoda σ_1 je enako prvemu vozlišču sprehoda σ_2 (spodnji del slike 2.2). Takšna sprehoda lahko staknemo v nov sprehod $\sigma_1 \circ \sigma_2$, katerega utež je definirana sledeče:

$$w(\sigma_1 \circ \sigma_2) = w(\sigma_1) \odot w(\sigma_2).$$

Če sprehoda σ_1 in σ_2 nista usklajena, velja $w(\sigma_1 \circ \sigma_2) = 0$.

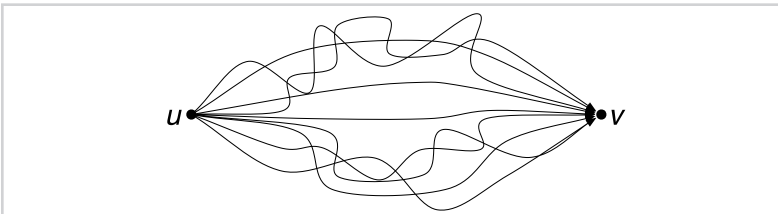
Naj bo $\mathscr{S}$ množica vseh sprehodov v $\mathscr{N}$. Uvedimo naslednje oznake:

- $\mathscr{S}_{uv}^{k}$ je množica vseh sprehodov od vozlišča u do vozlišča v dolžine k;
- $\mathscr{S}_{uv}^{(k)}$ je množica vseh sprehodov od vozlišča u do vozlišča v dolžine največ k;
- $\mathscr{S}_{uv}^{*}$ je množica vseh sprehodov od vozlišča u do vozlišča v (Slika 2.3);



Slika 2.2

Računanje vrednosti sprehodov.



Slika 2.3

Sprehodi med vozliščema u in v.

- $\overline{\mathscr{S}}_{uv}$ je množica vseh neničelnih sprehodov od vozlišča u do vozlišča v;
- $\mathcal{E}_{uv}$ je množica vseh enostavnih sprehodov (poti) od vozlišča u do vozlišča v.

Med temi množicami veljajo naslednje zveze:

$$\mathscr{S}^k_{uv} \subseteq \mathscr{S}^{(k)}_{uv} \subseteq \mathscr{S}^*_{uv},$$

$$k \neq l \Rightarrow \mathscr{S}^k_{uv} \cap \mathscr{S}^l_{uv} = \emptyset,$$

$$\mathscr{S}^{(k)}_{uv} = \bigcup_{i=0}^{k} \mathscr{S}^i_{uv}, \ \mathscr{S}^*_{uv} = \bigcup_{k=0}^{\infty} \mathscr{S}^k_{uv},$$

$$\text{za } k \geq |\mathscr{V}| - 1 \text{ velja } \mathcal{E}_{uv} \subseteq \mathscr{S}^{(k)}_{uv},$$

$$w(\mathscr{S}^{(k)}_{uv}) = \bigoplus_{i=0}^{k} w(\mathscr{S}^i_{uv}).$$

Naj bo $\mathscr{N}$ dano omrežje ter u in v izbrani vozlišči v tem omrežju. *Algebrajski problem poti* predstavlja določitev poti $\sigma \in \mathscr{S}_{uv}$, katere utež je enaka $w(\sigma) = \bigoplus_{\sigma \in \mathscr{S}_{uv}} w(\sigma)$.

Iščemo torej pot med vozliščema, ki je glede na polkolobar najboljša. Algebrajski problem poti lahko definiramo tudi kot iskanje takšne poti med vsemi pari vozlišč $u, v \in \mathscr{V}$.

Za običajen algebrajski problem poti uporabimo eno izmed različic *polkolobarja najkrajših poti* $(\mathbb{R} \cup \{\infty\}, \min, +, \infty, 0)$. V tem polkolobarju utež posamezne poti računamo postopoma: $w(\sigma_{sv}) = \min_{v \in V}\{\sigma_{su} + w(u,v)\}, \forall v \in V \setminus \{s\}$, pri čemer velja $w(\sigma_{ss}) = 0$.

Če označimo z $\mathscr{M}_n$ množico kvadratnih matrik, lahko tvorimo naslednji polkolobar: $(\mathscr{M}_n, \oplus, \odot, 0, 1)$, pri čemer sta operaciji definirani sledeče:

$$\mathbf{A} \oplus \mathbf{B} = \mathbf{C} \text{ za } [c_{ij}] = [a_{ij} \oplus b_{ij}],$$

$$\mathbf{A} \odot \mathbf{B} = \mathbf{C} \text{ za } [c_{ij}] = [\bigoplus_k a_{ik} \odot b_{kj}].$$

Problem iskanja najkrajše poti lahko v tem polkolobarju definiramo z enačbo $d^T = d^T \odot \mathbf{A} \oplus 1_s^T$, kjer z d označimo vektor uteži povezav od vozlišča s do vseh ostalih, $\mathbf{A}$ je utežena incidenčna matrika, 1_s pa enotski vektor z enico na mestu s. Problem najkrajših poti za vse pare vozlišč nato definiramo z enačbo $\mathbf{D} = \mathbf{DA} \oplus \mathbf{I}$, pri čemer je $\mathbf{D}$ matrika uteži najkrajših poti med vsemi pari vozlišč. Rešitev tega problema je $\mathbf{D} = \mathbf{A}^* = \bigoplus_{k=0}^{n-1} \mathbf{A}^k$. Če je v polkolobarju največji element ravno nevtralni element za multiplikativno operacijo, so uteži na vseh povezavah nenegativne.

Naj bosta $\mathscr{S}_1$ in $\mathscr{S}_2$ končni množici sprehodov. Potem velja

$$w(\mathscr{S}_1 \cup \mathscr{S}_2) \oplus w(\mathscr{S}_1 \cap \mathscr{S}_2) = w(\mathscr{S}_1) \oplus w(\mathscr{S}_2).$$

V posebnem primeru, ko je $\mathscr{S}_1 \cap \mathscr{S}_2 = \emptyset$, velja enakost $w(\mathscr{S}_1 \cup \mathscr{S}_2) = w(\mathscr{S}_1) \oplus w(\mathscr{S}_2)$. Stikanje sprehodov lahko posplošimo na množice sprehodov:

$$\mathscr{S}_1 \circ \mathscr{S}_2 = \{\sigma_1 \circ \sigma_2 : \sigma_1 \in \mathscr{S}_1, \sigma_2 \in \mathscr{S}_2, \sigma_1 \text{ in } \sigma_2 \text{ sta usklajena}\}.$$

Velja tudi: $\mathscr{S} \circ \emptyset = \emptyset \circ \mathscr{S} = \emptyset$.

Množica sprehodov $\mathscr{S}$ je *enolično razcepna* za množici sprehodov $\mathscr{S}_1$ in $\mathscr{S}_2$, če velja $\mathscr{S} = \mathscr{S}_1 \circ \mathscr{S}_2$ in za vse sprehode $\sigma_1, \sigma_1' \in \mathscr{S}_1$, $\sigma_2, \sigma_2' \in \mathscr{S}_2$, $\sigma_1 \neq \sigma_1'$, $\sigma_2 \neq \sigma_2'$ velja $\sigma_1 \circ \sigma_2 \neq \sigma_1' \circ \sigma_2'$.

Vsak sprehod dolžine k lahko za vsak $s, 0 < s < k$, enolično razcepimo na njegov začetek dolžine s in preostanek dolžine $k - s$.

Izrek 2.1: Naj bo končna množica $\mathscr{S}$ enolično razcepna na $\mathscr{S}_1$ in $\mathscr{S}_2$ ali naj bo polkolobar poln. Potem velja

$$w(\mathscr{S}_1 \circ \mathscr{S}_2) = w(\mathscr{S}_1) \odot w(\mathscr{S}_2).$$

Za katerokoli kvadratno matriko $\mathbf{W}$ nad $\mathbb{K}$ je njena k-ta stopnja $\mathbf{W}^k$ enolično določena zaradi asociativnosti.

Izrek 2.2: Vrednost w^k_{uv} k-te potence $\mathbf{W}^k$ vrednostne matrike $\mathbf{W}$ je enaka vrednosti vseh sprehodov od vozlišča u do vozlišča v dolžine k :

$$w(S^k_{uv}) = \mathbf{W}^k[u, v] = w^k_{uv}.$$

Zato za vrednostno matriko $\mathbf{W}$ acikličnega omrežja $\mathscr{N}$ velja:

$$\exists k_0 < n : \forall k > k_0 : \mathbf{W}^k = 0,$$

pri čemer je k_0 dolžina najdaljšega sprehoda v omrežju.

Torej, če je $\mathbf{W}$ matrika sosednosti omrežja nad kombinatoričnim polkolobarjem $(\mathbb{N}, +, \cdot, 0, 1)$, potem vrednosti w^k_{uv} predstavljajo število različnih sprehodov od vozlišča u do vozlišča v dolžine k.

Naj bo

$$\mathbf{W}^{(k)} = \bigoplus_{i=0}^{k} \mathbf{W}^i.$$

V idempotentnem polkolobarju velja $\mathbf{W}^{(k)} = (1 \oplus \mathbf{W})^k$.

Izrek 2.3:

$$w(S^{(k)}_{uv}) = \mathbf{W}^{(k)}[u, v] = w^{(k)}_{uv}.$$

Če je $\mathbf{W}$ matrika sosednosti omrežja nad kombinatoričnim polkolobarjem, potem vrednosti $w^{(k)}_{uv}$ predstavljajo število različnih sprehodov od vozlišča u do vozlišča v dolžine največ k.

Izrek 2.4: Naj bo $\mathbf{W}$ vrednostna matrika omrežja nad polnim polkolobarjem. Za njeno zaprtje $\mathbf{W}^*$ in strogo zaprtje $\overline{\mathbf{W}}$ velja:

$$w(\mathscr{S}^*_{uv}) = \mathbf{W}^*[u, v] = w^*_{uv} \qquad \text{in} \qquad w(\overline{\mathscr{S}}_{uv}) = \overline{\mathbf{W}}[u, v] = \overline{w}_{uv}.$$

Če je $\mathbf{W}$ matrika sosednosti omrežja nad polkolobarjem dosegljivosti, je matrika $\overline{\mathbf{W}}$ njeno tranzitivno zaprtje.

Če je $\mathbf{W}$ vrednostna matrika omrežja nad polkolobarjem najkrajših poti, so vrednosti w^*_{uv} enake vrednostim najkrajših poti od vozlišča u do vozlišča v.

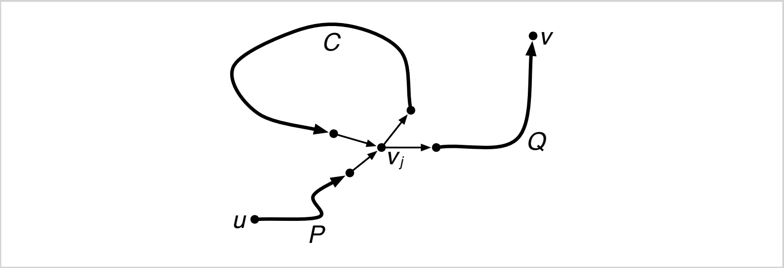
Naj bo $(\mathbb{K}, \oplus, \odot, 0, 1)$ absorpcijski polkolobar in naj bo σ sprehod, ki ni pot, iz množice sprehodov $\mathcal{S}^{(k)}_{uv}$. V σ se vsaj eno vozlišče v_j pojavi več kot enkrat. Del sprehoda med njegovo prvo in zadnjo pojavitvijo je sklenjen sprehod C (Slika 2.4). Celoten sprehod zato lahko zapišemo kot $\sigma = P \circ C \circ Q$, pri čemer je P začetni del sprehoda σ od vozlišča u do prve pojavitve vozlišča v_j in Q končni del sprehoda σ od zadnje pojavitve v_j do vozlišča v. Tudi $P \circ Q$ je sprehod. Vrednost obeh sprehodov skupaj je enaka

$$w(\{P \circ Q, P \circ C \circ Q\}) = w(P \circ Q).$$

Torej obhodi v sprehodih ne prispevajo k vrednosti sprehodov. Zato velja

$$w(\mathcal{S}^*_{uv}) = w(\mathcal{E}_{uv}).$$

Enakost velja tudi za $\mathcal{S}^*_{uv} = \emptyset$.



Slika 2.4
Primer sprehoda, ki ni pot.

Ker je množica vozlišč $\mathcal{V}$ končna, je tudi množica $\mathcal{E}_{uv}$ končna, kar nam omogoča izračun vrednosti $w(S^*_{uv})$. Vemo že, da velja $\mathbf{W}^* = \mathbf{W}^{(k)} = (\mathbf{1} + \mathbf{W})^k$ za dovolj velike vrednosti k. Za izračun matrike zaprtja $\mathbf{W}^*$ dane matrike nad polnim polkolobarjem $(\mathbb{K}, \oplus, \odot, 0, 1)$ lahko uporabimo Fletcherjev algoritem [36], ki je prikazan v algoritmu 2.2. Fletcherjev algoritem je posplošitev več algoritmov (Kleene, Warshall, Floyd, Roy) za izračun zaprtja nad specifičnim polkolobarjem.

Če iz algoritma odstranimo vrstico $c_k[k, k] := 1 \oplus c_k[k, k]$, dobimo algoritem za izračun strogega zaprtja $\overline{\mathbf{W}}$. Če je operacija seštevanja $\oplus$ idempotentna, lahko zaprtje matrike izračunamo na mestu – izpustimo indekse v matrikah $\mathbf{C}_k$.

Algoritem 2.2

Fletcherjev algoritem za izračun zaprtja matrike.

$\mathbf{C}_0 \leftarrow \mathbf{W}$
for $k \in \{1, \ldots n\}$ *do*
  for $i \in \{1, \ldots n\}$ *do*
    for $j \in \{1, \ldots n\}$ *do*
      $c_k[i,j] \leftarrow c_{k-1}[i,j] \oplus c_{k-1}[i,k] \odot (c_{k-1}[k,k])^* \odot c_{k-1}[k,j]$
    end for
    $c_k[k,k] \leftarrow 1 \oplus c_k[k,k]$
  end for
end for
$\mathbf{W}^* \leftarrow \mathbf{W}_n$

Urejeni polkolobarji so polkolobarji, za katere za $\mathbb{K}$ obstaja delna urejenost $\prec$ (refleksivnost, tranzitivnost, antisimetričnost). Dioid je polkolobar, ki je kanonično urejen, torej zanj velja $a \preceq b \Leftrightarrow \exists c \in \mathbb{K} : a \oplus c = b$. Če je operacija $\oplus$ idempotentna, je polkolobar dioid. Če je operacija $\oplus$ selektivna ($\forall a, b \in \mathbb{K} : a \oplus b = a \vee a \oplus b = b$), je polkolobar dioid s polno urejenostjo ($\forall a, b \in \mathbb{K} : a \preceq b \vee b \preceq a$). Urejenost se z operacijo $\odot$ ohranja.

Polkolobarje lahko ustvarimo iz že obstoječega polkolobarja $(\mathbb{K}, \boxplus, \boxdot)$ [4]:

- *Razširitev na vektorje in matrike:* $(\mathbb{K}, \oplus, \odot)$, kjer je $\oplus$ seštevanje po elementih in $\odot$ matrično množenje.
- *Konvolucija:* $(M, \star)$ je monoid z identiteto 1, torej funkcije $f : M \to \mathbb{K}$ tvorijo polkolobar z operacijama $(f \oplus g)(m) = f(m) \boxplus g(m)$ in $(f \odot g)(m) = \bigoplus_{m' \star m'' = m} f(m') \boxdot g(m'')$.
- *Redukcija:* $f : \mathbb{K} \to \mathbb{K}$ je redukcija za $f(0) = 0$, $f(1) = 1$, $f(a \oplus b) = f(f(a) \oplus b)$, $f(a \odot b) = f(f(a) \odot b) = f(a \odot f(b))$. $(\mathbb{K}_f, \oplus, \odot)$ je polkolobar za $\mathbb{K}_f = \{a \in \mathbb{K}; f(a) = a\}$, $a \oplus b = f(a \boxplus b)$, $a \odot b = f(a \boxdot b)$.
- *Endomorfizmi:* naj bo množica H množica endomorfizmov za $\mathbb{K}$ (torej velja linearnost). $(H, \oplus, \odot)$ je polkolobar za $(h_1 \oplus h_2)(a) = h_1(a) \oplus' h_2(a)$, $h_1 \odot h_2(a) =$

$h_2(h_1(a))$. $(\mathbb{K}, \oplus, \odot)$ je polkolobar za fiksirana $h_1, h_2 \in H$ in $a \oplus b = a \boxplus b$, $a \odot b = h_1(a) \boxdot h_2(b)$.

2.5 Množenje matrik in vektorjev

Naj bo e_i enotski vektor dolžine n – edini neničelni element je na i-tem mestu in le-ta je enak 1. Produkt enotskega vektorja in vrednostne matrike omrežja lahko uporabimo za izračun vrednosti sprehodov od vozlišča i do vseh ostalih vozlišč.

Naj bo

$$q_1^T = e_i^T \odot \mathbf{W}.$$

Vrednosti elementov vektorja q_1 so enake vrednostim sprehodov od vozlišča i do vseh ostalih vozlišč dolžine $1 : q_1[j] = w(\mathcal{S}_{ij}^1)$. Vrednosti vseh sprehodov dolžine s, $s = 2, 3, \ldots k$, ki se začnejo v vozlišču i, lahko izračunamo iterativno:

$$q_s^T = q_{s-1}^T \odot \mathbf{W}$$

ali $q_s^T = e_i^T \odot \mathbf{W}^s$ in $q_s[j] = w(\mathcal{S}_{ij}^s)$. Podobno dobimo $q^{(k)^T} = e_i^T \odot \mathbf{W}^{(k)}$, $q^{(k)}[j] = w(\mathcal{S}_{ij}^{(k)})$ in $q^{*T} = e_i^T \odot \mathbf{W}^*$, $q^*[j] = w(\mathcal{S}_{ij}^*)$.

To lahko posplošimo. Naj bo $\mathcal{I} \subseteq \mathcal{V}$ in naj bo $e_{\mathcal{I}}$ karakteristična vrednost množice $\mathcal{I}$ – za elemente $\mathcal{I}$ ima vrednost 1, za ostale pa 0. Potem recimo za $q_k^T = e_{\mathcal{I}}^T \odot \mathbf{W}^k$ velja $q_k[j] = w(\bigcup_{i \in \mathcal{I}} \mathcal{S}_{ij}^k)$.

2.6 Polkolobarji in omrežja

V nadaljevanju sledi opis polkolobarjev, ki so primerni za delo z omrežji. Več drugih primerov polkolobarjev je opisanih v [4, 11, 37–39]. Za omrežje $\mathcal{N} = ((\mathcal{I}, \mathcal{J}), \mathcal{A}, w)$ privzamimo okrajšano oznako $\mathcal{N} = \mathcal{I}\mathcal{J}$.

2.6.1 Kombinatorični polkolobar

Kombinatorični polkolobar $(\mathbb{N}, +, \cdot, 0, 1)$ je najpogosteje uporabljani polkolobar [11]. Namesto množice naravnih števil lahko uporabimo tudi kake druge množice: $\overline{\mathbb{N}}, \mathbb{R}, \mathbb{R}_0^+$ ali $\mathbb{Q}$. Če uporabimo množico $\overline{\mathbb{N}}$, je polkolobar zaprt za $a^* = \sum_{k \in \overline{\mathbb{N}}} a^k$, saj je ta polkolobar poln. Drugo zaprtje, ki velja v tem polkolobarju nad množico $\mathbb{R}$, je $a^* = 1/(1-a)$ za vsak $a \neq 1, \infty$. Za to zaprtje velja $0^* = 1, 1^* = \infty$ in $\infty^* = \infty$. Polkolobar je komutativen, saj v njem velja $a \cdot b = b \cdot a$ za vse a in b v množici elementov polkolobarja. Kombinatorični polkolobar ni idempotenten.

V produktu omrežij $\mathscr{A}\mathscr{B}$ in $\mathscr{B}\mathscr{C}$ je vrednost posamezne povezave enaka

$$w(a,c) = \sum_{b\in\mathscr{B}} (w(a,b) \cdot w(b,c)).$$

Ta vrednost pove število možnih sprehodov dolžine 2 od a do c preko $\mathscr{B}$.

Običajno pri množenju omrežij uporabljamo kombinatorični polkolobar. Imejmo omrežje soavtorstev $\mathscr{W}\mathscr{A}$, pri čemer so $\mathscr{W}$ dela in $\mathscr{A}$ raziskovalci, $w(w,a) = 1$, če je raziskovalec a soavtor dela w. Transponirano omrežje le-tega zmnožimo z omrežjem soavtorstev. Vrednosti povezav $w(a,c)$ v dobljenem produktu $\mathscr{A}\mathscr{W} \star \mathscr{W}\mathscr{A}$ so enake številu del, pri katerih sta sodelovala a in c. Vrednosti zank pa so enake številu del v soavtorstvu posameznega raziskovalca.

Naj bosta $\mathscr{D}\mathscr{Z}$ in $\mathscr{Z}\mathscr{T}$ omrežji, v katerih $\mathscr{D}$ označuje množico držav, $\mathscr{Z}$ množico znanj in $\mathscr{T}$ množico tehnologij. Vrednost povezave med državo in znanjem je enaka ena, če ima država na voljo znanje, vrednost povezave med znanjem z in tehnologijo t pa je enaka $w(z,t) = 1/\mathrm{indeg}(t)$, če je znanje z potrebno za razvoj tehologije t. Zanima nas, ali ima neka država izpolnjena vsa potrebna znanja za razvoj določene tehnologije. V ta namen izračunamo produkt $\mathscr{D}\mathscr{Z} \star \mathscr{Z}\mathscr{T}$, v katerem so vrednosti povezav definirane sedeče:

$$w(d,t) = \begin{cases} 1; & d \text{ ima izpolnjena vsa potrebna znanja za razvoj } t, \\ < 1; & d \text{ nima izpolnjenih vseh potrebnih znanja za razvoj } t, \\ 0; & d \text{ nima nobenih znanj za razvoj } t. \end{cases}$$

2.6.2 *Polkolobar regularnih izrazov*

Ta polkolobar ima oznako $(\mathscr{P}(\Sigma^*), \cup, \odot, \emptyset, \{\lambda\})$ [11], pri čemer je Σ^* množica končnih nizov nad abecedo Σ. Za unijo velja posplošena komutativnost, posplošena asociativnost in posplošena distributivnost, zato je polkolobar regularnih izrazov poln polkolobar. Zato je polkolobar zaprt za $a^* = \bigcup_{k\in\mathbb{N}} \underbrace{a \ldots a}_{k}$. Polkolobar je idempotenten, saj je unija dveh enakih množic spet ta množica. V njem ne velja absorpcijski zakon.

Če so v množici $\mathscr{A}$ omrežja $\mathscr{A}\mathscr{B}$ shranjeni vsa možna začetna stanja, v množici $\mathscr{C}$ omrežja $\mathscr{B}\mathscr{C}$ vsa možna končna stanja in v skupni množici $\mathscr{B}$ vsa možna vmesna stanja, dobimo s produktom $\mathscr{A}\mathscr{B} \star \mathscr{B}\mathscr{C}$ množico vseh možnih potekov v danem končnem avtomatu:

$$w(a,c) = \bigcup_{b\in\mathscr{B}} \odot(w(a,b), w(b,c)).$$

2.6.3 Polkolobar najkrajših poti

V polkolobarju $(\mathbb{R}_0^+, \min, +, \infty, 0)$ [11] velja komutativnost za množenje. Obstaja več različic polkolobarja najkrajših poti glede na izbiro množice: $\mathbb{R}_0^+, \overline{\mathbb{N}}, \mathbb{R} \cup \{\infty\}$. Polkolobar je zaprt za $a^* = 0$, saj je $\min(0, a + a^*) = 0$. Zaradi $\min(0, a) = 0$ velja tudi absorpcijski zakon. Kadar uporabimo množico $\overline{\mathbb{N}}$, se polkolobar imenuje tropski polkolobar. V primeru množice $\mathbb{R} \cup \{\infty\}$ je polkolobar izomorfen $(x \mapsto -x)$ polkolobarju max-plus $(\mathbb{R} \cup \{-\infty\}, \max, +, -\infty, 0)$.

Polkolobar najkrajših poti je idempotenten in v njem velja absorpcijski zakon.

S polkolobarjem najkrajših poti računamo najboljšo pot med izvirom in ponorom. Pri tem privzamemo $a_{ii} = 0$, $a_{ij} = \infty$, če med i in j ni povezave.

Računamo lahko tudi najkrajšo pot s časovno nehomogenimi povezavami. Za izvir s, destinacijo d in potovalni čas t_0 računamo najkrajši možni čas potovanja od s do d. Množico v polkolobarju sestavljajo endomorfizmi.

V produktu $\mathscr{A}\mathscr{B} \star \mathscr{B}\mathscr{C}$ so vrednosti povezav enake najmanjšim vsotam uteži povezav med a in c preko $\mathscr{B}$:

$$w(a, c) = \min_{b \in \mathscr{B}} (w(a, b) + w(b, c)).$$

Kot primer uporabe tega polkolobarja lahko vzamemo omrežje $\mathscr{A}\mathscr{B}$, kjer so $\mathscr{A}$ mesta in $\mathscr{B}$ podjetja, $w(a, b) =$ moč mesta a v omrežju podjetja b. Vrednost povezave v produktu $\mathscr{A}\mathscr{B} \star \mathscr{B}\mathscr{A}$ je enaka moči podjetja, ki je v obeh mestih skupaj (a in c) najšibkejše. Diagonalne vrednosti $w(a, a)$ so dvakratniki moči najšibkejšega podjetja v mestu a.

Še en primer uporabe polkolobarja najkrajše poti je analiza občutljivosti. Opazujemo lahko vpliv poznavanja toleranc povezav pri razumevanju preusmerjanja prometa. Preusmerjanje predstavlja preusmeritev prometa skozi povezave ali vozlišča, ki jih napadalec izbere, s čimer se te poti zdijo krajše od ostalih. To lahko povzroči zastoje na preusmerjenih delih omrežja, preusmerjen promet pa je slabši od optimalnega, saj lahko napadalec prometu prisluškuje, ga prilagodi ali spusti. Napadalec preusmeri promet na dva načina. S privlakom prometa se zdi, da ima povezava mnogo manjšo utež od ostalih. Sprememba teže pomeni preusmeritev prometa skozi to povezavo. Z odbojem prometa pa se zdi, da ima povezava mnogo večjo utež od ostalih. Poslabšanje uteži povezave povzroči odmik prometa.

2.6.4 Polkolobar dosegljivosti

Polkolobar dosegljivosti ali logični polkolobar $(\{0,1\}, \vee, \wedge, 0, 1)$ uporabljamo za reševanje vprašanj povezanosti v omrežjih [4]. Operacija množenja je komutativna in v polkolobarju velja absorpcijski zakon. Polkolobar dosegljivosti je zaprt za $a^* = 1 \vee a \wedge a^* = 1$ in je idempotenten.

Vrednosti povezav v produktu $\mathscr{AB} \star \mathscr{BC}$ so enake 1, če sta krajišči povezani vsaj preko enega skupnega $b \in \mathscr{B}$:

$$w(a,c) = \bigvee_{b \in \mathscr{B}} (w(a,b) \wedge w(b,c)).$$

Diagonalne vrednosti so enake nič za izolirana vozlišča.

Vzemimo bibliografsko omrežje $\mathscr{WK}$, pri čemer so $\mathscr{W}$ dela in $\mathscr{K}$ ključne besede, $w(w,k) = 1$, če je ključna beseda k uporabljena v delu w. Vrednost povezave v produktu $\mathscr{KW} \star \mathscr{WK}$ je enaka 1, če sta bili ključni besedi k_1 in k_2 (krajišči povezave) uporabljeni v vsaj enem skupnem delu. Diagonalna vrednost je enaka 0, če ključna beseda k ni bila nikoli uporabljena.

2.6.5 Verjetnostni polkolobar

Oznaka tega polkolobarja je $([0,1], +, \cdot, 0, 1)$ [11]. Utež povezave pove verjetnost prehoda iz začetnega vozlišča v končno vozlišče. Velja

$$\sum_{v \in N(u)} w(u,v) = 1.$$

Polkolobar je zaprt za $a^* = 1/(1-a)$ za $a \neq 1$. Polkolobar ni idempotenten in v njem aborpcijski zakon ne velja.

Ta polkolobar lahko uporabimo za reševanje posplošenega problema najkrajših poti na uteženem usmerjenem grafu. Utežene relacije med vozlišči v omrežju določajo metriko zaupanja. Uteži so določene z $w(u,v) : \mathscr{V} \times \mathscr{V} \to S$ (prostor mnenj; opinion space). Zanima nas, kakšno mnenje ima s o t, če upoštevamo mnenja na poti od s do t. Za določanje ugleda pa upoštevamo dve metriki: zaupanje (trust) in samozavest (confidence). Zaupanje predstavlja pravilnost (ali je nekdo dober človek), samozavest pa natančnost podatkov zaupanja.

Z uporabo množice $[0, \infty]$ lahko ta polkolobar uporabimo na markovskih verigah. Graf je usmerjen, uteži na povezavah so enake verjetnostim, da bo naključni sprehod vseboval posamezno povezavo. Vozlišča so torej stanja v markovski verigi, povezave pa verjetnosti korakov. Velja $\sum_v p_{uv} = 1, \forall i$, torej lahko tvorimo incidenčno matriko $\mathbf{A}$ tako, da vrednosti (u, v) v $\mathbf{A}^k$ predstavljajo verjetnosti, da smo v k korakih v stanju v, če smo začeli v u. Drugi primer je Fletcher postopek, pri katerem velja $\sum_v p_{uv} \leq 1$ in je verjetnost, da se ustavimo v u, enaka $1 - \sum_v p_{uv}$. Računamo pričakovano število obiskov naključnega sprehoda v vozlišču t, če smo sprehod začeli v s.

Vrednost

$$w(a,c) = \sum_{b \in \mathscr{B}} w(a,b) \cdot w(b,c)$$

v produktu $\mathscr{A}\mathscr{B} \star \mathscr{B}\mathscr{C}$ predstavlja verjetnost sprehoda iz a v c preko $\mathscr{B}$.

Vrednosti povezav v produktu nad verjetnostnim polkolobarjem se izračunajo enako kot nad kombinatoričnim polkolobarjem. Razlika je v omejitvi vrednosti. Pokažimo, da velja $w(a,c) \in [0,1]$. Vemo, da velja

$$w(a,b), w(b,c) \in [0,1], \forall b \in \mathscr{B}.$$

$$\Rightarrow w(a,b) \cdot w(b,c) \leq w(a,b), w(b,c)$$

Ker je omrežje dvovrstno, velja

$$N(a) \subseteq \mathscr{B} \Rightarrow \sum_{b \in \mathscr{B}} w(a,b) = 1.$$

$$\Rightarrow 1 = \sum_{b \in \mathscr{B}} w(a,b) \geq \sum_{b \in \mathscr{B}} w(a,b) \cdot w(b,c) \geq 0.$$

Preverimo še pogoj z vsoto vrednosti povezav v soseščino nekega vozlišča $a \in \mathscr{A}$:

$$\begin{aligned}
\sum_{c \in \mathscr{C}} w(a,c) &= \sum_{c \in \mathscr{C}} \sum_{b \in \mathscr{B}} w(a,b) \cdot w(b,c) \\
&= \sum_{b \in \mathscr{B}} \sum_{c \in \mathscr{C}} w(a,b) \cdot w(b,c) \\
&= \sum_{b \in \mathscr{B}} \left(w(a,b) \cdot \sum_{c \in \mathscr{C}} w(b,c) \right) \\
&= \sum_{b \in \mathscr{B}} w(a,b) \cdot 1 \\
&= 1.
\end{aligned}$$

2.6.6 Polkolobar razdalje

Operaciji v $([0,\infty]\times[0,1],\oplus,\odot,(0,0),(0,1))$ [4] sta definirani sledeče:

$$(a_1,b_1)\oplus(a_2,b_2)=(a_1+a_2,b_1+b_2),$$

$$(a_1,b_1)\otimes(a_2,b_2)=(a_1b_2+a_2b_1,b_1b_2).$$

Polkolobar ni idempotenten in v njem ne velja absorpcijski zakon.

Ta polkolobar predstavlja posplošitev verjetnostnega polkolobarja: poleg verjetnosti imamo na povezavah podane še njihove uteži.

2.6.7 Polkolobar max-min

V polkolobarju $(\overline{\mathbb{R}}_0^+,\max,\min,0,\infty)$ [11] so dovoljene tudi druge številske množice: $[L,R]\subseteq\overline{\mathbb{R}}$ za $L<R$ in $\{0,1\}$. Polkolobar z množico $\overline{\mathbb{R}}_0^+$ je zaprt za $a^*=\infty$, je idempotenten in v njem velja absorpcijski zakon. Imenuje se tudi algebra ozkega grla.

Uporabimo ga lahko za računanje najboljše poti med izvirom in ponorom glede na dano metriko ali več metrik. Primer: iskanje najširše poti med s in t, torej poti z največjo kapaciteto (glede na ozko grlo). Polkolobar je komutativen.

Polkolobar max-min z množico $\{0,1\}$ lahko uporabimo tudi za določanje determinističnega neuspeha oz. dosegljivost. Preverjamo torej povezanost v grafu, pri čemer najprej tvorimo incidenčno matriko z elementi

$$a_{ij}=\begin{cases}1; & (i,j) \text{ je povezava,}\\ 0; & sicer\end{cases}$$

nad polkolobarjem. Polkolobar max-min z izbrano množico $\{0,1\}$ je absorbirajoči polkolobar z ničlo 1 in enoto 0.

Vrednost posamezne povezave v produktu $\mathscr{A}\mathscr{B}\star\mathscr{B}\mathscr{C}$ izračunamo sledeče:

$$w(a,c)=\max_{b\in\mathscr{B}}(\min(w(a,b),w(b,c))).$$

Za vsak $b\in\mathscr{B}$ vzamemo manjšo od vrednosti povezav od/do a in c ter iz dobljenih vrednosti izberemo najvišjo. Vrednost povezave $w(a,c)$ pove, koliko vsaj lahko prenesemo od a do c, če izbiramo med vsemi potmi med njima. Pot med dvema vozliščema je močna toliko, kolikor je močan njen najšibkejši člen.

2.6.8 Polkolobar min-max

V polkolobarju $(\overline{\mathbb{R}}_0^+, \min, \max, \infty, 0)$ [11] so dovoljene tudi druge številske množice: $[L, R] \subseteq \overline{\mathbb{R}}$ za $L < R$, $\{0, 1\}$. Polkolobar z izbrano množico $\overline{\mathbb{R}}_0^+$ je zaprt za $a^* = 0$, je idempotenten in v njem velja absorpcijski zakon.

Ta polkolobar je izomorfen polkolobarju max-min s preslikavo $x \mapsto \frac{1}{x}$.

V usmerjenem in povezanem omrežju lahko s pomočjo tega polkolobarja poiščemo minimalno (uteženo) vpeto drevo. Povezava (i, j) je v tem drevesu natanko takrat, ko velja $\forall \sigma \in \mathcal{S}_{ij} : w(i, j) < \max_{(e,f)\in\sigma} w(e, f)$ (ko je na katerikoli drugi poti od i do j povezava z večjo utežjo).

Vrednost povezave v produktu $\mathcal{A}\mathcal{B} \star \mathcal{B}\mathcal{C}$ izračunamo sledeče:

$$w(a, c) = \min_{b\in\mathcal{B}}(\max(w(a, b), w(b, c))).$$

Ta vrednost je enaka najmanjši vrednosti izmed močnejših povezav v paru (a, b) in (b, c) za vsak $b \in \mathcal{B}$.

2.6.9 Polkolobar verjetnosti najboljše poti

Polkolobar $([0, 1], \max, \cdot, 0, 1)$ [11] lahko uporabimo za računanje najboljše poti med izvirom in ponorom glede na dano metriko ali več metrik. Primer: iskanje najbolj zanesljive poti glede na verjetnost uspešnosti prenosa.

Polkolobar je zaprt za $a^* = 1$, je idempotenten in v njem drži absorpcijski zakon.

Če omrežje predstavlja strukturo in anotacijo spletne strani, so vrednosti povezav enake zanesljivosti posameznih internetnih poti. V tem primeru lahko polkolobar verjetnosti najboljše poti uporabimo za ugotavljanje stopnje kvalitete spletne strani za vsakega uporabnika. Temu polkobarju zato rečemo tudi semantični omrežni polkolobar.

Vrednosti povezav v produktu $\mathcal{A}\mathcal{B} \star \mathcal{B}\mathcal{C}$ izračunamo kot

$$w(a, c) = \max_{b\in\mathcal{B}}(w(a, b) \cdot w(b, c)).$$

Če so uteži povezav enake verjetnosti, da neka pot vsebuje določeno povezavo, ali pa ta verjetnost odraža kakšno drugo lastnost omrežja, potem je vrednost $w(a, c)$ enaka najboljši možni izbiri poti od a do c preko $\mathcal{B}$. To velja ob predpostavki, da najboljšo izbiro predstavlja najvišja dovoljena vrednost.

2.6.10 Polkolobar vozlišč sprehoda

Množica polkolobarja $(\mathscr{P}(\mathscr{V} \cup \mathscr{V}^2 \cup \ldots \cup \mathscr{V}^n), \cup, \circ, \emptyset, \emptyset)$ [11] je družina podmnožic sprehodov v omrežju. Operacija množenja je definirana kot $W_1 \circ W_2 = \{w_1 \circ w_2; w_1 \in W_1, w_2 \in W_2,$ zadnje vozlišče v w_1 je prvo vozlišče v $w_2\}$. Polkolobar je idempotenten in v njem ne velja absorpcijski zakon.

Ta polkolobar lahko uporabimo za iskanje sprehoda med dvema vozliščema v omrežju. Množenje predstavlja stikanje sprehoda iz dveh kosov: prvi del je iz W_1, drugi pa iz W_2. Ta polkolobar lahko uporabimo tudi za iskanje sprehoda, ki zadošča nekem kriteriju.

2.6.11 Polkolobar najboljše poti

Množica W polkolobarja $(W^k, \oplus, \odot, (\max(W), \ldots, \max(W)), (0, \ldots, 0))$ [4] je množica uteži povezav, operaciji sta definirani sledeče:

$$(a_1, a_2, \ldots, a_k) \oplus (b_1, b_2, \ldots, b_k) = k\!-\min\{a_1, a_2, \ldots, a_k, b_1, b_2, \ldots, b_k\},$$

$$(a_1, a_2, \ldots, a_k) \odot (b_1, b_2, \ldots, b_k) = k\!-\min\{a_i + b_j; i, j = 1, 2, \ldots, k\},$$

pri čemer operator $k\!-\min$ pomeni izbiro k najmanjših elementov iz dane množice. Ničla in enota sta vektorja dolžine k. Polkolobar je idempotenten in v njem velja absoprcijski zakon.

S tem polkolobarjem računamo najboljšo pot med izvirom in ponorom glede na dano metriko ali več metrik. Poiščemo lahko k najkrajših poti.

2.6.12 Časovni polkolobar

Množica $\mathbb{K}$ polkolobarja $(\mathbb{K}, \min, \text{kompozitum}, \max(W), 1)$ [4] je množica vrednosti povezav, enota je enaka vrednosti poti dolžine 0. Polkolobar je idempotenten in v njem absorpcijski zakon velja.

S tem polkolobarjem lahko poiščemo najkrajšo pot s časovno nehomogenimi povezavami. Za izvir s, destinacijo d in potovalni čas t_0 računamo najkrajši možni čas poti od s do d. Z vsako povezavo (i, j) asociiramo nenaraščajočo $w(i, j) : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$, $\lim_{t \to \infty} w(i, j)(t) = \infty$. $w(i, j)(t)$ je čas prispetja do j z zamikom $w(i, j)(t) - t$. To je polkolobar endomorfizmov.

2.6.13 Prerezni polkolobar

S polkolobarjem $(2^{\mathscr{A}} \cup \{0\}, \cap, \cup)$ [4] za $\mathscr{A}$ množico povezav lahko poiščemo prerezne povezave. Povezava e predstavlja prerezno povezavo, če $\exists s, t : e$ leži na vseh poteh od s do t. Če namesto podane množice uporabimo $2^{\mathscr{V}} \cup \{0\}$, lahko s tem polkolobarjem poiščemo prerezna vozlišča.

2.6.14 Družina polkolobarjev - bijekcija, homomorfizem

Naj bo $B \subseteq \overline{\mathbb{R}}$ takšna, da je $(B, +, \cdot, 0, 1)$ ali $(B, \min, +, U, 0)$ polkolobar za $U = \max(B)$. Torej velja $0 \in B$ in $1 \in B$. Naj bo $A \subseteq \mathbb{R}$ takšna, da je $g : A \to B$ bijekcija. Definiramo operacije $\oplus, \triangledown, \odot$ tako, da je g homomorfizem:

$$g(a \oplus b) = g(a) + g(b),$$
$$g(a \triangledown b) = \min(g(a), g(b)),$$
$$g(a \odot b) = g(a) \cdot g(b).$$

Te definicije so enakovredne naslednjim:

$$a \oplus b = g^{-1}(g(a) + g(b)),$$
$$a \triangledown b = g^{-1}(\min(g(a), g(b))),$$
$$a \odot b = g^{-1}(g(a) \cdot g(b)).$$

Tudi funkcija g^{-1} je homomorfizem: označimo $\alpha = g(a), \beta = g(b)$. Potem velja $a = g^{-1}(\alpha), b = g^{-1}(\beta)$, torej velja

$$a \oplus b = g^{-1}(\alpha) \oplus g^{-1}(\beta),$$
$$a \oplus b = g^{-1}(g(a) + g(b)) = g^{-1}(\alpha + \beta),$$

$$a \odot b = g^{-1}(\alpha) \odot g^{-1}(\beta),$$
$$a \odot b = g^{-1}(g(a) \cdot g(b)) = g^{-1}(\alpha \cdot \beta),$$

$$a \triangledown b = g^{-1}(\alpha) \triangledown g^{-1}(\beta),$$
$$a \triangledown b = g^{-1}(\min(g(a), g(b))) = g^{-1}(\min(\alpha, \beta)).$$

Če je g strogo naraščajoča funkcija, velja $a \triangledown b = g^{-1}(\min(g(a), g(b))) = \min(a, b) :$

$$\min(g(a), g(b)) \in \{g(a), g(b)\} \subseteq B,\ g^{-1}(g(x)) = x, x \in B.$$

Ker homomorfizem ohranja algebrske lastnosti, je $(A, \oplus, \odot, g^{-1}(0), g^{-1}(1))$ polkolobar za $A \subseteq \mathbb{R}, g$ bijekcijo in homomorfizem.
Dokaz: [Za polkolobar]

- Komutativnost prve operacije: Očitno.
- Asociativnost prve operacije:

$$\begin{aligned}a \oplus (b \oplus c) =& g^{-1}(g(a) + g(g^{-1}(g(b) + g(c))))\\ =& g^{-1}(g(a) + g(b) + g(c))\\ =& g^{-1}(g(g^{-1}(g(a) + g(b))) + g(c))\\ =& (a \oplus b) \oplus c\end{aligned}$$

- Obstoj ničle:

$$\begin{aligned}a \oplus 0 =& g^{-1}(g(a) + g(0)) \overset{0=g^{-1}(0)}{=} g^{-1}(g(a))\\ =& a\end{aligned}$$

- Asociativnost druge operacije:

$$\begin{aligned}a \odot (b \odot c) =& g^{-1}(g(a) \cdot g(g^{-1}(g(b) \cdot g(c))))\\ =& g^{-1}(g(a) \cdot g(b) \cdot g(c))\\ =& g^{-1}(g(g^{-1}(g(a) \cdot g(b))) \cdot g(c))\\ =& g^{-1}(g(a \odot b) \cdot g(c))\\ =& g^{-1}(g(a) \cdot g(b)) \cdot g(c))\\ =& (a \odot b) \odot c\end{aligned}$$

- Obstoj enote:

$$\begin{aligned}a \odot 1 = g^{-1}(g(a) \cdot g(1)) \quad &= \quad g^{-1}(g(1) \cdot g(a)) = 1 \odot a\\ = g^{-1}(g(a) \cdot g(1)) \overset{1=g^{-1}(1)}{=} \quad & g^{-1}(g(a)) \quad = \quad a\end{aligned}$$

- Distributivnostna zakona:

$$\begin{aligned}a \odot (b \oplus c) \quad &= \quad g^{-1}(g(a) \cdot g(g^{-1}(g(b) + g(c))))\\ &= \quad g^{-1}(g(a) \cdot (g(b) + g(c)))\\ &= \quad g^{-1}(g(a) \cdot g(b) + g(a) \cdot g(c))\\ &\overset{g^{-1} homo\varphi}{=} g^{-1}(g(a) \cdot g(b)) \oplus g^{-1}(g(a) \cdot g(c))\\ &= \quad (a \odot b) \oplus (a \odot c)\end{aligned}$$

$$\begin{aligned}(b \oplus c) \odot a &= g^{-1}(g(g^{-1}(g(b)+g(c))) \cdot g(a)) \\ &= g^{-1}((g(b)+g(c)) \cdot g(a)) \\ &= g^{-1}(g(b) \cdot g(a) + g(c) \cdot g(a)) \\ &\overset{g^{-1} homo\varphi}{=} g^{-1}(g(b) \cdot g(a)) \oplus g^{-1}(g(c) \cdot g(a)) \\ &= (b \odot a) \oplus (c \odot a)\end{aligned}$$

■

Polkolobarji Minkowskega 1

Polkolobarji $(\mathbb{R}_0^+, \min, \boxed{r}, \infty, 0)$ se imenujejo tudi polkolobarji Pathfinder [13]. Ti polkolobarji so podmnožica zgornje družine polkolobarjev za $g(x) = x^r, g^{-1}(y) = \sqrt[r]{y}$. Množenje je definirano kot operacija Minkowskega: $a \boxed{r} b = \sqrt[r]{a^r + b^r}$. Iz članka [40] izhaja, da je struktura $(\mathbb{R}_0^+, \min, \boxed{r}, \infty, 0)$ poln polkolobar. Primeri izbire vrednosti r:

- $r = 1 : a \boxed{r} b = a + b$,
- $r = 2 : a \boxed{r} b = \sqrt{a^2 + b^2}$,
- $r = \infty : a \boxed{r} b = \max(a, b)$.

Ti polkolobarji so zaprti za $a^* = 0$, so idempotentni in v njih velja absorpcijski zakon.

Dokaz: [Za polkolobar] Poglejmo le en distributivnostni zakon; ostalo je očitno.

$$\begin{aligned}a \boxed{r} (\min(b,c)) &= \sqrt[r]{a^r + \min(b,c)^r} \\ &= \sqrt[r]{a^r + \min(b^r, c^r)} \\ &= \min(\sqrt[r]{a^r + b^r}, \sqrt[r]{a^r + c^r}) \\ &= \min(a \boxed{r} b, a \boxed{r}, c)\end{aligned}$$

■

Ta polkolobar uporabimo pri poenostavljanju uteženih omrežij z odstranjevanjem povezav, ki ne zadoščajo trikotniški neenakosti. To pomeni, da iz omrežja odstranimo tiste povezave, katerih krajišča so povezana tudi preko neke krajše poti. Osnovna ideja postopka Pathfinder je v tem, da so različnosti med vozlišči predstavljene kot uteži usmerjenih povezav, torej ohranimo le tiste povezave, ki tvorijo najkrajše poti med vozlišči glede na dano mero različnosti. Kot matriko različnosti lahko uporabimo uteženo matriko sosednosti. Ko tvorimo matriko različnosti, vedno opazujemo različnosti med vsemi pari vozlišč, mera različnosti pa se navezuje na število skupnih sosedov.

Postopek Pathfinder: izračunaj matriko različnosti $\mathbf{W}$, nato izračunaj q-to potenco te matrike nad polkolobarjem Minkowskega ($\mathbf{W}^{(q)}$). Za vsako povezavo v omrežju preveri, ali sta njegovi uteži v $\mathbf{W}$ in $\mathbf{W}^{(q)}$ enaki (oz. se razlikujeta kvečjemu za $\varepsilon > 0$ zaradi zaokrožitvenih napak). Če da, je ta povezava vključena v poenostavljeno omrežje, sicer ne.

Vrednost r za operator Minkowskega določimo glede na izbrano mero različnosti. Pri izbiri $r = 1$ dobimo polkolobar najkrajših poti, pri izbiri $r = \infty$ pa polkolobar min-max.

Vrednost povezave v produktu $\mathscr{A}\mathscr{B} \star \mathscr{B}\mathscr{C}$ izračunamo sledeče:

$$w(a,c) = \min_{b \in \mathscr{B}} \sqrt[r]{w(a,b)^r + w(b,c)^r}.$$

Vrednost $w(a,c)$ je enaka dolžini najkrajše dovoljene poti od a do c preko nekega $b \in B$, pri čemer za izračun dolžine uporabimo r-to metriko.

Polkolobarji Minkowskega 2

Polkolobarji $(\mathbb{R}, \boxed{\mathrm{r}}, \cdot, 0, 1)$ so tudi podmnožica zgornje družine polkolobarjev za $g(x) = x^r, g^{-1}(y) = \sqrt[r]{y}$. Za seštevanje vzamemo operacijo Minkowskega, za množenje pa kar običajno množenje števil. Polkolobar je zaprt za $a^* = 0$, ni idempotenten in v njem ne velja absorpcijski zakon.
Dokaz: [Za polkolobar] Poglejmo le en distributivnostni zakon; ostalo je očitno.

$$\begin{aligned} a \cdot (b \boxed{\mathrm{r}} c) &= a \cdot \sqrt[r]{b^r + c^r} = \\ &= \sqrt[r]{a^r} \cdot \sqrt[r]{b^r + c^r} = \\ &= \sqrt[r]{a^r \cdot b^r + a^r \cdot c^r} = \\ &= (a \boxed{\mathrm{r}} b) \cdot (a \boxed{\mathrm{r}} c) \end{aligned}$$

■

Vrednost r za operator Minkowskega določimo glede na izbrano mero različnosti. Pri izbiri $r = 1$ dobimo kombinatorični polkolobar, pri izbiri $r = \infty$ pa polkolobar

$(\max, \cdot)$. Ti polkolobarji so izomorfni polkolobarjem Minkowskega $(\mathbb{R}_0^+, \min, \boxed{\text{r}}, \infty, 0)$ s preslikavo $x \mapsto \ln x$.

Vrednosti povezav v produktu $\mathscr{A}\mathscr{B} \star \mathscr{B}\mathscr{C}$ izračunamo sledeče:

$$w(a,c) = \sqrt[r]{\sum_{b \in \mathscr{B}} (w(a,b) \cdot w(b,c))^r}.$$

Na vrednosti povezav lahko gledamo kot na število možnih prehodov iz začetnega v končno vozlišče. Tako pri izbiri $r = 1$ dobimo novo utež, ki je enaka številu vseh možnih prehodov iz a do c.

2.6.15 Polkolobarji za uravnoteženost in razvrščanje

Označen graf je urejen par $(\mathscr{G}, \sigma)$, kjer je $\mathscr{G} = (\mathscr{V}, \mathscr{A})$ usmerjen graf brez zank, $\sigma : \mathscr{A} \to \{p, n\}$ pa je označitvena funkcija. Povezave z znakom p so pozitivne, povezave z znakom n pa negativne. Množico vseh pozitivnih povezav označimo z $\mathscr{A}^+$, množico vseh negativnih povezav pa z $\mathscr{A}^-$. V neusmerjenem grafu lahko vsako povezavo nadomestimo z dvema nasprotno usmerjenima povezavama, ki imata isti znak kot originalna povezava. Označen graf je uravnotežen natanko takrat, ko lahko množico vozlišč $\mathscr{V}$ razdelimo v dve podmnožici tako, da sta krajišči vsake pozitivne povezave znotraj ene podmožice, krajišči vsake negativen povezave pa v dveh podmnožicah. Označen graf je razcepen natanko takrat, ko lahko množico vozlišč $\mathscr{V}$ razdelimo v vsaj dve paroma ločeni skupini tako, da sta krajišči vsake pozitivne povezave znotraj ene skupine, krajišči vsake negativne povezave pa v obeh skupinah. Sprehod v označenem grafu je pozitiven natanko takrat, ko vsebuje sodo število negativnih povezav, sicer je negativen. Veljata naslednja izreka [5]:

Izrek 2.5: Označen graf $(\mathscr{G}, \sigma)$ je uravnotežen natanko takrat, ko je vsak sklenjen sprehod pozitiven.

Izrek 2.6: Označen graf $(\mathscr{G}, \sigma)$ je razcepen natanko takrat, ko G ne vsebuje sklenjenega sprehoda z natanko eno negativno povezavo.

Polkolobar za uravnoteženost

Ta polkolobar ima oznako $(\{0, n, p, a\}, \oplus, \odot, 0, p)$ [5]. Vsak graf označimo z eno izmed naslednjih oznak:

- 0 – v grafu ni sprehodov,
- n – vsi sprehodi v grafu so negativni,
- p – vsi sprehodi v grafu so pozitivni,
- a – obstaja vsaj eden negativen in obstaja vsaj eden pozitiven sprehod v grafu.

Potem sta operaciji polkolobarja definirani kot:

$\oplus$	0	n	p	a
0	0	n	p	a
n	n	n	a	a
p	p	a	p	a
a	a	a	a	a

$\odot$	0	n	p	a
0	0	0	0	0
n	0	p	n	a
p	0	n	p	a
a	0	a	a	a

Ta polkolobar je zaprt za $0^* = p, n^* = a, p^* = p, a^* = a$, je idempotenten in v njem ne velja absorpcijski zakon.

V družbenih omrežjih pozitivna utež mnogokrat predstavlja prijateljstvo, negativna pa sovraštvo. Polkolobar predstavlja polkolobar za identifikacijo uravnoteženih grafov. Označen graf je uravnotežen natanko takrat, ko ima vsak njegov obhod težo 1 (je pozitiven), kar je enakovredno naslednjemu izreku.

Izrek 2.7 ([5]): Označen graf $(\mathscr{G}, \sigma)$ je uravnotežen natanko takrat, ko so vsi diagonalni elementi uravnoteženega zaprtja matrike $\mathbf{W}^{\bullet}$ enaki p.

Polkolobar za razvrščanje

Ta polkolobar ima oznako $(\{0, n, p, a, q\}, \oplus, \odot, 0, p)$ [5]. Vsako omrežje označimo z eno izmed naslednjih oznak:

- 0 – v grafu ni sprehodov,
- n – obstaja vsaj en sprehod z natanko eno negativno povezavo, ni sprehodov s samo pozitivnimi povezavami,
- p – obstaja vsaj en sprehod s samo pozitivnimi povezavami in nič sprehodov z natanko eno negativno povezavo,

- a – obstaja vsaj en sprehod s samo pozitivnimi povezavami, obstaja vsaj en sprehod z natanko eno negativno povezavo,
- q – vsak sprehod ima vsaj dve negativnipovezavi.

Potem operaciji polkolobarja definiramo sledeče:

$\oplus$	0	n	p	a	q
0	0	n	p	a	q
n	n	n	a	a	n
p	p	a	p	a	p
a	a	a	a	a	a
q	q	n	p	a	q

$\odot$	0	n	p	a	q
0	0	0	0	0	0
n	0	q	n	n	q
p	0	n	p	a	q
a	0	n	a	a	q
q	0	q	q	q	q

Polkolobar je zaprt za $0^* = p, n^* = a, p^* = p, a^* = a, q^* = p$, je idempotenten in v njem ne velja absorpcijski zakon.

Polkolobar predstavlja polkolobar za identifikacijo razcepnih grafov. Označen graf je razcepen natanko takrat, ko ne vsebuje nobenega obhoda z natanko eno negativno povezavo, kar je enakovredno naslednjemu izreku.

Izrek 2.8 ([5]): Označen graf $(\mathscr{G}, \sigma)$ je razcepen natanko takrat, ko so vsi diagonalni elementi razcepnega zaprtja matrike $\mathbf{W}^\bullet$ enaki p.

2.6.16 Polkolobar geodezičnih razdalj in števcev

Operaciji v $((\mathbb{R}_0^+ \cup \{\infty\}) \times (\mathbb{N} \cup \{\infty\}), \oplus, \odot, (\infty, 0), (0, 1))$ [5] sta definirani takole:

$$(a, i) \oplus (b, j) = (\min(a, b), \begin{cases} i & a < b \\ i + j & a = b \\ j & a > b \end{cases}),$$

$$(a, i) \odot (b, j) = (a + b, i \cdot j).$$

Ta polkolobar je zaprt za

$$(a, i)^* = \begin{cases} (0, \infty) & a = 0, i \neq 0 \\ (0, 1) & sicer. \end{cases}$$

Polkolobar ni idempotenten in v njem velja absorpcijski zakon.

Ta polkolobar je posplošitev polkolobarja najkrajših poti. Z njim hkrati izračunamo dolžino in število najkrajših poti med pari vozlišč. Prvi del uporabljamo za iskanje dolžine najkrajših poti med vsemi pari vozlišč. Operacija $\odot$ v drugem delu pride do izraza, kadar naredimo zaporedno vezavo najkrajših poti v nekem vozlišču t. Če je najkrajših poti od u do t i, najkrajših poti od t do v pa j, je vseh najkrajših poti od u do v ravno $i \cdot j$. Operacija $\oplus$ v drugem delu deluje pri vzporedni vezavi najkrajših poti - glede na dolžine različnih najkrajših poti od u do v določimo število najkrajših poti. Če je poti od vozlišča u do vozlišča v dolžine a i in poti dolžine b j, je novo število najkrajših poti enako i, če je $a < b$, enako j, če velja $a > b$, če sta pa vrednosti a in b enaki, pa je novo število najkrajših poti enako $i + j$.

S tem polkolobarjem lahko izračunamo indeks središčnosti vsakega vozlišča glede na vmesnost: $C_B(t) = \sum_u \sum_v \frac{n_{uv}(t)}{n_{uv}}$, pri čemer je n_{uv} število najkrajših poti od u do v, $n_{uv}(t)$ pa število teh najkrajših poti, ki gredo skozi vozlišče t. Ta indeks izračunamo preko matrike $[(d_{uv}, n_{uv})]$, saj je

$$n_{uv}(t) = \begin{cases} n_{ut}n_{tv}; & d_{ut} + d_{tv} = d_{uv} \\ 0; & \text{sicer} \end{cases},$$

kjer je d_{uv} dolžina najkrajših poti od u do v. To matriko lahko izračunamo iz zaprtja relacijske matrike nad polkolobarjem geodezičnih razdalj in števcev. Elementi relacijske matrike so

$$(d, n)_{uv} = \begin{cases} (1, 1); & (u, v) \text{ v relaciji} \\ (\infty, 0); & \text{sicer} \end{cases},$$

ničla in enota polkolobarja sta $0 = \{(\infty, 0)\}$ in $1 = \{(0, 1)\}$.
Dokaz: [Za polkolobar]

- Komutativnost prve operacije:

$$(a, i) \oplus (b, j) = (b, j) \oplus (a, i)\text{: Očitno.}$$

- Asociativnost prve operacije (privzamemo lahko, da velja $a \leq c$):

$$((a,i) \oplus (b,j)) \oplus (c,k) = (a,i) \oplus ((b,j) \oplus (c,k))$$

za vsak primer posebej:

$$\begin{array}{lrl}
a = b = c: & (a, i+j+k) & (a, i+j+k) \\
a < b = c: & (a,i) & (a,i) \\
a = b < c: & (a, i+j) & (a, i+j) \\
a < b < c: & (a,i) \oplus (c,k) = (a,i) & (a,i) \oplus (b,j) = (a,i) \\
a < b, b > c: & (a,i) \oplus (c,k) & (a,i) \oplus (c,k)
\end{array}$$

- Obstoj ničle:

$$(a,i) \oplus (\infty, 0) = (\infty, 0) \oplus (a,i) = (a,i)$$

- Asociativnost druge operacije:

$$\begin{array}{rcl}
((a,i) \odot (b,j)) \odot (c,k) & = & (a+b, i \cdot j) \odot (c,k) \\
& = & (a+b+c, i \cdot j \cdot k \cdot) \\
(a,i) \odot ((b,j) \odot (c,k)) & = & (a,i) \odot (b+c, j \cdot k) \\
& = & (a+b+c, i \cdot j \cdot k \cdot))
\end{array}$$

- Obstoj enote:

$$(a,i) \odot (0,1) = (0,1) \odot (a,i) = (a+0, i \cdot 1) = (a,i)$$

$$(a,i) \odot (\infty, 0) = (\infty, 0) \odot (a,i) = (\infty + a, 0 \cdot i) = (\infty, 0)$$

- Pri dokazovanju distributivnostnega zakona lahko zaradi komutativnosti $\oplus$ privzamemo $b \leq c$ za spodnji splošen primer:

$$\begin{array}{rcl}
(a,i) \odot ((b,j) \oplus (c,k)) & = & (a,i) \odot (b, \begin{cases} j & b < c \\ j+k & b = c \end{cases}) \\
& = & (a+b, \begin{cases} i \cdot j & b < c \\ i \cdot (j+k) & b = c \end{cases}) \\
(a,i) \odot (b,j) \oplus (a,i) \odot (c,k) & = & (a+b, i \cdot j) \oplus (a+c, i \cdot k) \\
& = & (a+b, \begin{cases} i \cdot j & b < c \\ i \cdot j + i \cdot k & b = c \end{cases})
\end{array}$$

Drugi distributivnostni zakon dokažemo podobno.

■

Za dano dvovrstno omrežje $\mathcal{N} = (\mathcal{I}, \mathcal{J}, \mathcal{L}, w)$ tvorimo matriko $[(d,n)_{i,j}]_{\mathcal{I}\times\mathcal{J}}$:

$$(d,n)_{i,j} = \begin{cases} (w(i,j), n(i,j)) & (i,j) \in \mathcal{L} \\ (\infty, 0) & \text{sicer,} \end{cases}$$

pri čemer je $w(i,j)$ utež povezave (i,j), $n(i,j)$ pa je enako 1. Če dovolimo večkratne povezave, je $w(i,j)$ utež najcenejše povezave od i do j, $n(i,j)$ pa število povezav (i,j) s to utežjo.

Vrednosti povezav v produktu $\mathcal{AB} \star \mathcal{BC}$ torej izračunamo kot

$$(d,n)_{a,c} = \left(\min_{b\in\mathcal{B}}(w(a,b) + w(b,c)), \sum_{\substack{b\in\mathcal{B} \\ w(a,b)+w(b,c) \text{ je min.}}} n(a,b)\cdot n(b,c) \right).$$

Prvi element predstavlja vrednost najcenejšega prehoda od a do c preko $\mathcal{B}$ (enako vrednost dobimo tudi pri uporabi polkolobarja najkrajših poti), drugi elementa pa število najcenejših prehodov od a do c preko $\mathcal{B}$.

2.6.17 *Semantični polkolobar*

Polkolobar $([0,1], \min, \cdot, 1, 1)$ [4] je idempotenten in v njem ne velja absorpcijski zakon.

Uporabimo ga lahko za reševanje posplošenega problema najkrajših poti na uteženem usmerjenem grafu. Semantični omrežni polkolobar uporabimo za ugotavljanje stopnje kvalitete spletne strani za vsakega uporabnika glede na strukturo in anotacijo vsebine spletne strani.

2.6.18 *Polkolobar geodezičnih množic*

Ta polkolobar ima oznako $(\{(i, \mathcal{P}, j, d); i, j \in \mathcal{V}, \mathcal{P} \subseteq \mathcal{V}, d \in \mathbb{N}\}, \oplus, \odot, \{(i, \{i\}, i, 0); i \in \mathcal{V}\}, \{(\#, \emptyset, \#, 0)\}$ [5], v množici je i začetek, j konec, d dolžina, $\mathcal{P}$ pa množica vozlišč na poteh od i do j s krajiščema vred. Znak # v enoti polkolobarja je poljubno vozlišče. Operaciji sta definirani sledeče:

$$s_1 \oplus s_2 = \text{Redukcija}(s_1 \cup s_2)$$
$$\text{Redukcija}(s) = [(s \setminus \{a_i, a_j\}) \cup (\{a_i\} \oplus \{a_j\}), i < j]$$
$$a_1 \oplus a_2 = \{(i_1, \mathscr{P}_1, j_1, d_1)\} \oplus \{(i_2, \mathscr{P}_2, j_2, d_2)\} =$$
$$\begin{cases} \{(i_1, \mathscr{P}_1, j_1, d_1)\}; & i_1 = i_2, j_1 = j_2, d_1 < d_2 \\ \{(i_1, \mathscr{P}_2, j_1, d_2)\}; & i_1 = i_2, j_1 = j_2, d_1 > d_2 \\ \{(i_1, \mathscr{P}_1 \cup \mathscr{P}_2, j_1, d_1)\}; & i_1 = i_2, j_1 = j_2, d_1 = d_2 \\ \{(i_1, \mathscr{P}_1, j_1, d_1)\} \cup \{(i_2, \mathscr{P}_2, j_2, d_2)\}; & \text{sicer} \end{cases}$$
$$S_1 \otimes S_2 = \bigoplus_{w_1 \in S_1, w_2 \in S_2} w_1 \odot w_2$$
$$(i_1, \mathscr{P}_1, j_1, d_1) \odot (i_2, \mathscr{P}_2, j_2, d_2) = \begin{cases} (i_1, \mathscr{P}_1, j_1, d_1); & d_2 = 0, j_1 = i_2 \\ (i_2, \mathscr{P}_2, j_2, d_2); & d_1 = 0, j_1 = i_2 \\ (i_1, \mathscr{P}_1 \cup \mathscr{P}_2, j_2, d_1 + d_2); & (i_1 \neq j_1 \neq j_2, \\ & \mathscr{P}_1 \cap \mathscr{P}_2 = \{j_1\}, \\ & j_1 = i_2) \vee \\ & (j_1 \neq i_1 = j_2, \\ & \mathscr{P}_1 \cap \mathscr{P}_2 = \{i_1, j_1\}, \\ & j_1 = i_2) \\ 1; & \mathit{sicer} \end{cases}$$

Polkolobar geodezičnih množic uporabimo za iskanje množice vozlišč na najkrajši poti med vsakima dvema s, t v usmerjenem neuteženem grafu. Množico vozlišč na najkrajši poti med dvema vozliščema izračuna Boylov operator: $u \star v = \{z \in \mathscr{V}; z$ na najkrajši poti od u do $v\}$. Zanj velja: $u \star v \subset u \star t \Rightarrow v \star t \subset u \star t$ in $u \star v = w \star t \Rightarrow u, v, w, t$ so v nekem obhodu, vendar pa za Boylov operator ne velja asociativnost. Rezultat Boylovega operatorja lahko izračunamo preko tega polkolobarja tako, da relacijo pretvorimo v matriko z elementi $(\mathscr{P}, p)_{uv} = \begin{cases} (\{u, v\}, 1); & (u, v) \text{ v relaciji} \\ (\emptyset, \infty); & \text{sicer} \end{cases}$, kjer je $\mathscr{P}$ množica vozlišč na najkrajši $u - v$ poti, p pa dolžina najkrajših poti.

2.7 Zaključek

Množenje omrežij je ena izmed metod, ki se v analizi omrežij pogosto pojavlja. Omogoča namreč izračun izpeljanih omrežij, s katerimi pridobimo nov pogled na podatke in drugačne rezultate. Izračune izpeljanih omrežij lahko dodatno razširimo, kadar v množenje omrežij vpeljemo uporabo še katerega drugega polkolobarja poleg kombinatoričnega.

V tem poglavju je najprej podana definicija polkolobarja, opisane pa so posamezne lastnosti polkolobarjev, ki jih potrebujemo pri množenju omrežij. Ker si množenje omrežij lahko predstavljamo kot množenje matrik teh omrežij, v nadaljevanju poglavja razširimo polkolobarje na matrične polkolobarje.

Eden izmed problemov na omrežjih, v katerem pride do izraza uporaba polkolobarjev, je algebrajski problem poti. Za reševanje tega problema uporabimo polkolobar najkrajših poti, na voljo pa je še več razširitev tega polkolobarja, ki so v tem poglavju tudi opisane.

Poleg različic polkolobarja najkrajših poti so v tem poglavju podani opisi več drugih polkolobarjev, v katerih vidimo uporabnost v analizi omrežij. Seznamu teh polkolobarjev nameravamo v prihodnje dodati še primere uporabe na realnih podatkih.

3

Zbiranje podatkov za analize

Podatke v obliki omrežja najdemo v marsikaterem področju vsakdanjega življenja:

- mesta so povezana s cestami,
- skupina ljudi je povezana z izmenjavo elektronskih in telefonskih sporočil,
- dela iz posameznega raziskovalnega področja so povezana preko citatov,
- raziskovalci so povezani med sabo preko skupnih raziskav,
- atomi so v molekulah povezani s kemijskimi vezmi,
- besede povezujemo med sabo glede na njihova sopojavljanja v povedih besedila,
- rodovniki so omrežja, v katerih so ljudje med sabo povezani s porokami in vezmi starš-otrok.

Kadar gradimo omrežje, moramo najprej definirati vozlišča in relacijo, s katero ta vozlišča povežemo – pojavi se problem *mej omrežja* [41, 42]. Množico vozlišč omejimo v skladu z načrtom analize omrežja. Hkrati z vozlišči in povezavami izberemo tudi njihove lastnosti. Odločiti se moramo, ali bo omrežje enovrstno ali dvovrstno in katere lastnosti so bistvene za načrtovane analize. Pri povezavah moramo odgovoriti na več vprašanj: Ali so povezave usmerjene? Ali v omrežju obstaja več različnih vrst povezav (več relacij)? Ali lahko med dvema vozliščema obstaja več povezav? Kaj je pomen uteži povezav? Je omrežje statično ali dinamično (se spreminja skozi čas)?

Včasih seznam vozlišč poznamo že vnaprej (recimo učence v razredu). Velikokrat pa množico vozlišč gradimo tekom zbiranja podatkov. V tem primeru definiramo kriterij pripadnosti, ki za vsako možno vozlišče odloči, ali pripada omrežju ali ne.

Za zbiranje omrežnih podatkov velikorat uporabimo postopek *snežne kepe* [43]. Najprej izberemo (majhno) množico vozlišč, ki predstavljajo začetne kandidate. Nato zberemo podatke o vsakem kandidatu in določimo njemu sosednja vozlišča. Tista sosednja vozlišča, ki so nova, dodamo v seznam kandidatov. Vključitev novih vozlišč lahko določa tudi kakšno drugačno pravilo, recimo razdalja od najbližjega začetnega vozlišča. Postopek končamo, ko je seznam kandidatov izrabljen ali ko dosežemo želeno število opazovanih vozlišč.

Tekom določanja množice vozlišč se lahko pojavi problem *identifikacije* vozlišč. Enota, ki ustreza vozlišču, ima lahko več različnih poimenovanj (*sinonimi*) ali pa posamezno poimenovanje označuje več različnih enot (*homonimi* ali *dvoumnost*). Primer iz bibliografije matematike so imena avtorjev v bibliografski bazi Zentralblatt MATH; zapisi imen Borštnik, N. S. Mankoč; Mankoč Borštnik, N.; Mankoč-Borštnik, Norma; Mankoč Borštnik, Norma Susana; Mankoc-Borstnik, N.S. in Mankoč Borštnik, N.S. pripadajo isti avtorici. Na drugi strani v tej bazi obstajata zapisa, ki pripadara vsaj dvema različnima avtorjema Smith, John W., saj v bazi najdemo zapise publikacij s tem imenom avtorja v obdobju od 1868 do 2007. V matematični bazi MathSciNet Database je shranjen zapis imena Wang, Li, ki pripada vsaj 57 različnim matematikom. Uredniki te baze se že od leta 1985 trudijo z reševanjem identifikacije imen avtorjev [44], kar izvajajo med vnosom podatkov v bazo. Problem bi lahko rešili z enolično 'številko raziskovalca', k čemur stremita sistema ResearcherID in ORCID.

Problem identifikacije se pojavi tudi pri pridobivanju enot iz enostavnega besedila. Na primer besedni zvezi 'glavni izvršni direktor podjetja Apple' in 'Tim Cook' trenutno označujeta isto osebo. Za reševanje te plati problema moramo podati ekvivalenčno relacijo, s katero določimo sezname ekvivalentnih besednih zvez. Drugi vir problema identifikacije so slovnična pravila jezika, v katerem je besedilo napisano. Na primer glagol 'iti' se lahko v besedilu v slovenščini pojavi v oblikah 'grem', 'greš', 'pojdite', 'sva šla' idr. Za reševanje tega problema uporabimo krnjenje in lematizacijo s pomočjo orodij za obdelavo naravnih jezikov, kakor sta NLTK[1] in MontyLingua[2].

Poseben pristop k zbiranju podatkov za analizo omrežij je ustvarjanje *egocentričnih* omrežij [45]. Ta pristop uporabimo, kadar je obseg populacije, ki nas zanima, prevelik. Iz populacije izberemo vzorec enot (*ego*) in zberemo podatke o njih, njihovih sosedih (*alter*) in povezavah med njimi. Primer takega omrežja je omrežje prijateljstva izbranih oseb na Facebooku.

Kadar zbiramo podatke, moramo paziti na pravne (pravice kopiranja) in etične omejitve [46–49].

3.1 Metode za pridobivanje podatkov

Omrežne podatke lahko pridobimo na več različnih načinov:

- z opazovanjem,
- z anketami ali intervjuji,

[1] http://www.nltk.org/
[2] http://web.media.mit.edu/ hugo/montylingua/

- iz arhivov in podatkovnih baz,
- iz podatkov, ki so že organizirani v strukturo omrežja,
- z izpeljavo iz zbirk podatkov,
- iz pomenskega spleta (semantic web),
- z generiranjem naključnih omrežij itd.

Vsaka izmed naštetih metod za pridobivanje omrežnih podatkov je podrobneje opisana v nadaljevanju.

3.1.1 Opazovanje

Načini pridobivanja podatkov so se spreminjali skozi čas glede na razvoj tehnologije. Osnovni pristop je opazovanje [50]. Opazovanje je človeška dejavnost, ki je sestavljena iz pridobivanja informacij o zunanjem svetu skozi čutila ali iz zapisovanj podatkov s pomočjo znanstvenih pripomočkov in vključuje tudi zbiranje podatkov. Za izboljšanje človeške moči opazovanja so bili razviti znanstveni pripomočki, recimo tehtnice, ura, teleskop, mikroskop, termometer, kamera, snemalnik zvoka, in tudi za pretvorbo za človeška čutila nezaznavnih dogodkov v bolj zaznavne oblike, kot so voltmeter, spektometer, infrardeče kamere, osciloskop, interferometer, geigerjev števec, rentgen in radijski sprejemnik [51].

Najbolj natančna metoda za merjene posameznih spremenljivk je neposredno merjenje, ki pa je lahko omejeno s pomanjkanjem tehnologije. Najpogostejši alternativni pristop k temu je pridobitev podatkov o dejavnostih drugih.

Primer pridobivanja omrežnih podatkov z opazovanji je opisan v doktorski disertaciji Samuela F. Sampsona [52]. Naredil je etnografsko študijo sprememb strukture skupnosti v samostanu v Novi Angliji – 18 novicev je razdelil v štiri skupine v petih različnih časovnih obdobjih glede na njegova opazovanja in analize. Drugi primer uporabe opazovanja je zaznavanje molekularne strukture organskih molekul.

3.1.2 Ankete

Anketiranje je metoda za zbiranje podatkov, ki aktivno vključuje tudi opazovane [53]. Uporaba anket omogoča študij obnašanja, prepričanj in ostali značilnostih. Z dobro pripravljenim vprašalnikom lahko zberemo ogromne količine kvalitetnih podatkov.

Odgovori so lahko *zaprtega* tipa – izbor vnaprej pripravljenih odgovorov. Tovrstne odgovore je lažje analizirati. Odgovori *odprtega* tipa pa niso podani vnaprej in omogočajo pridobitev večjega obsega informacij. Anketo lahko izvedemo na več različnih načinov: v živo, s papirjem in pisalom, preko telefona, preko elektronske pošte ali preko spletne aplikacije. Danes večinoma izvajamo ankete v digitalni obliki (spletne ankete), ki jih lahko hitro prilagodimo, takoj preverimo podane odgovore in dopolnimo še z nekaj vsebinskimi (opazovalnimi) podatki.

Z uporabo neposrednega opazovanja v kombinaciji z anketiranjem dobimo dodatne informacije. S kombinacijo teh metod lahko potrdimo ali zanikamo informacije, ki smo jih pridobili z anketami. Kot za opazovanje se moramo tudi za opazovanje v kombinaciji z anketami pripraviti. Opazovalec lahko uporabi primerna merila, kontrolne sezname in ostale pripomočke, ki jih izbere v skladu z vprašanji in morebitnimi zaprtimi odgovori v anketi.

Primer zanimivega omrežja, ki je nastalo z uporabo anket, je Edinburgh Associative Thesaurus[3].

Ankete predstavljajo najpogosteje uporabljeno metodo za zbiranje podatkov o družbenih omrežjih. Uporabimo jih lahko tudi za raziskovanje notranjeorganizacijskih vezi [54].

3.1.3 *Arhivi in podatkovne baze*

Arhiv je zbirka zgodovinskih podatkov ali prostor, v katerem je ta zbirka shranjena [55]. Arhivi imajo zgodovinsko, kulturno in dokazno vrednost. Najdemo jih povsod, kjer so shranjeni podatki. Vsaka organizacija ima arhiv preteklih dejavnosti, univerze imajo arhive dosežkov in raziskav bivših študentov, varnostna kopija podatkov na osebnem računalniku je arhiv pretekle uporabe računalnika idr. Arhivi se vedno bolj pretvarjajo in tvorijo v digitalni obliki, saj se pisarniško delo vedno bolj seli na računalnike. K temu veliko pripomore tudi hitra širitev uporabe interneta.

Podatkovna baza je zgrajena v obliki zapisov. Za vsak zapis pa je shranjenih tudi nekaj lastnosti [56]. Zaradi organizacije podatkov je te podatke zelo lahko pretvoriti v zbirko (pogosto dvovrstnih) omrežij. Manjše količine podatkov pa lahko predstavimo v tabelarični obliki, recimo s preglednicami.

Bibliografske baze (Web of Science, Scopus, Zentralblatt Math idr.) hranijo podat-

[3]http://www.eat.rl.ac.uk/

ke o objavljenih člankih in knjigah. Obstaja tudi arhiv svetovnega spleta[4] (World Wide Web) – svetovni splet je delno shranjen kot arhiv za prihodnje raziskovalce, zgodovinarje in javnost.

Arhivi različnih vrst so poceni in prednostni viri podatkov za analizo že obstoječih ali preteklih družbenih omrežij [53]. Tako lahko povezave med korporacijami analiziramo na podlagi podatkov o članih upravnih odborov teh korporacij.

Zgodovinski arhivi pomagajo raziskovalcem pri zbiranju znanja o razvoju nekega področja – ekonomije, šolstva, vojske idr. Recimo podatki iz druge svetovne vojne omogočajo analizo premikov enot vojsk tekom vojne, prevozov beguncev ali zapornikov, prevozov orožja idr. Drugi primer uporabe arhivskih podatkov je analiza zavezništev med močnejšimi državami tekom izbranega časovnega intervala.

Arhivske podatke o prebivalcih nekega mesta ali področja lahko uporabimo za rodoslovne analize. V rodoslovju iščemo značilne vzorce porok in njihove nepravilnosti. Tako iščemo poroke med sorodniki, ki omogočajo ohranjanje bogastva znotraj družine, ali poroke med pomembnejšimi družinami, kar omogoča povečanje vpliva družine. Rodoslovni podatki so pogosto na voljo v formatu GEDCOM[5]. Večje zbirke družinskih rodovnikov so na voljo na spletni strani Genealogy Forum[6]. 'Znanstveni' rodovniki, ki se uporabljajo v antropoloških raziskavah, so na voljo na spletni strani KinSource[7].

Dejavnosti na svetovnem spletu (elektronska pošta, klepet, forumi idr.) pustijo sled, ki jo lahko uporabimo kot vir podatkov za analizo omrežij. Dobro poznani primer je Enronova elektronska pošta[8].

Arhiv svetovnega spleta je še posebej zanimiv za področje analize omrežij. Za zbiranje podatkov iz svetovnega spleta lahko uporabimo spletne pajke (web crawlers), ki imajo na začetku shranjen dani seznam naslovov URL spletnih strani. Ko obiščejo spletno stran z naslovom iz tega seznama, identificirajo vse povezave na njej in dodajo njihove naslove v seznam naslovov. Internet Archive[9] je največja organizacija za arhiviranje spleta, ki zbira podatke s spletnimi pajki. Narodne knjižnice, narodni arhivi in podobne organizacije se osredotočajo na zbiranje kulturno pomembnih podatkov na

[4] http://webarchives.cdlib.org/

[5] http://homepages.rootsweb.ancestry.com/ pmcbride/gedcom/55gctoc.htm

[6] http://www.genealogyforum.com/gedcom/

[7] http://kinsource.net/csac/wiki/kinsrc/KinSources/

[8] http://www.isi.edu/~adibi/Enron/Enron.htm

[9] http://archive.org/index.php

svetovnem spletu.

Facebook, Twitter, LinkedIn, Google+ in ostali ponudniki družabnih omrežij so tudi primeri zbirateljev ogromnih količin arhivskih podatkov. Te organizacije zbirajo podatke o svojih uporabnikih, njihovih objavah in drugih dejavnostih. Podatki o uporabnikih niso javno dostopni, posamezen uporabnik lahko dobi podatke le o svoji pretekli dejavnosti in podatke, dostop do katerih mu dovolijo drugi uporabniki.

Internet Movie Database[10] (IMDb), Amazon[11] in last.fm[12] so nadaljni primeri ogromnih podatkovnih zbirk. Podatke iz teh podatkovnih zbirk lahko pretvorimo v dvovrstna omrežja, z analizo teh ter njihovih kombinacij pridobimo informacije o sodelovanjih med igralci, producenti in skladatelji, podobnosti med filmi glede na izbrane metrike idr.

Z razvojem tehnologije so se pojavile različne vrste podatkovnih baz. Vrsto podatkovne baze definiramo kot način, na katerega so podatki shranjeni v njej. Zaradi rasti količine podatkov so se razvila podatkovna skladišča. Podatkovna skladišča shranjujejo podatke neposredno iz virov. Predstavljajo osrednji vir podatkov za upravitelje, ki jih potrebujejo za izdelavo statističnih interaktivnih analitičnih plošč in poročil. Druga zelo popularna vrsta podatkovne baze je podatkovna baza v oblaku [57].

V analizi omrežij je zelo praktična uporaba grafovske podatkovne baze [58]. Ta vrsta podatkovne baze uporablja grafovsko strukturo za shranjevanje in predstavitev podatkov. Posebne grafovske podatkovne baze uporabljajo omrežni model, ki omogoča bolj prilagodljivo predstavitev objektov in relacij med njimi.

V zadnjih letih se vsak dan zbere velika količina podatkov. Zato se je pojavil pojem velikih podatkovij [59], ki predstavljajo zbirke zelo velikih množic podatkov. Te množice podatkov so tako velike in kompleksne, da jih je težko obdelovati z uveljavljenimi aplikacijami. Zato so se razvile nove in prilagojene metode na področju razvrščanja, strojnega učenja, nevronskih mrež, zaznavanja vzorcev in anomalij.

Primeri zbirk omrežij in drugih podatkovnih zbirk so: Repositories of Datasets[13], KDnuggets Datasets for Data Mining[14], Data Surfing on the World Wide Web[15],

[10]http://www.imdb.com/

[11]http://www.amazon.com/

[12]http://www.last.fm/

[13]http://www.trustlet.org/wiki/Repositories_of_datasets

[14]http://www.kdnuggets.com/datasets/index.html

[15]http://it.stlawu.edu/~rlock/datasurf.html

Public Data Sets on Amazon Web Services[16], TunedIT[17], The Internet2 Observatory Data Collections[18], Infochimps[19], CAIDA (The Cooperative Association for Internet Data Analysis) Data[20] in Network Data Sources[21] na spletni strani programa `Pajek`.

Določenim dejavnostim lahko sledimo z njihovim beleženjem. Telefonski operaterji shranjujejo podatke o uporabi telefonov, zbirajo se podatki iz vremenskih postaj, ponudniki spletnih družabnih omrežij zbirajo podatke o svojih uporabnikih [60], tvorijo se omrežja podatkov iz različnih senzorjev, vedno bolj zanimiva so omrežja peer-to-peer (P2P) , z identifikatorji radijskih frekvenc (RFID) lahko sledimo njihovim lastnikom idr. Tovrstne podatke lahko uporabimo za napovedno analizo in za analizo obnašanja uporabnikov.

3.1.4 Skoraj omrežni podatki

Nekateri podatki že imajo obliko omrežja. Prevozniško omrežje je omrežje cest, cevi, ulic ali druge podobne strukture, ki dovoljuje prevoz oz. prenos določene vrste. Te strukture so predstavljene kot povezave, njihova križišča pa kot vozlišča. Še eno področje, ki se ukvarja s podatki v skoraj omrežni obliki, je kemija. Struktura molekule je omrežje atomov kot vozlišč in kovalentnih kemijskih vezi kot povezav med njimi. Za analizo omrežij najbolj zanimive kemijske strukture so organske molekule: proteini, lipidi, hidrokarbonati in DNK. Veliko tovrstnih podatkov o molekulah je na voljo na spletni strani Protein Data Bank[22].

Kadar želimo tovrstne podatke analizirati z orodji analize omrežij, jih moramo najprej pretvoriti v ustrezen datotečni format.

Včasih potrebujemo za pretvorbo podatkov posebno programsko rešitev. Recimo pretvorbo opisa zemljevida z mejami med državnimi administrativnimi enotami (občine idr.) formata ESRI v relacijo sosednosti administrativnih enot lahko naredimo s kratkim programom v programskem jeziku R z uporabo funkcije `poly2nb` iz paketa `spdep`.

---

[16] http://aws.amazon.com/publicdatasets/

[17] http://tunedit.org/repo

[18] http://www.internet2.edu/observatory/archive/data-collections.html

[19] http://infochimps.com/

[20] http://www.caida.org/data/

[21] http://pajek.imfm.si/doku.php?id=data:urls:index

[22] http://www.rcsb.org/pdb/home/home.do

3.1.5 Omrežja, pridobljena iz podatkov

Določeni viri podatkov zahtevajo bolj zapletene postopke za pretvorbo podatkov v pripadajoča omrežja.

Za analizo lahko dobimo zanimive podatke iz dnevnih arhivov tiskovnih agencij (Agence France-Presse, Reuters, United Press International, American Press Agency, Xinhua, ITAR-TASS idr.). Posamezna novica je (označeno) besedilo, ki ga lahko analiziramo s postopki *računalniško podprte analize besedila* [61]. Eden izmed glavnih pristopov k tej vrsti analize besedila je semantična analiza besedila. Enote besedila so kodirane v skladu z modelom Chomskega *osebek-povedek-predmet*, ki ga enostavno pretvorimo v časovno večrelacijsko omrežje z osebki in predmeti kot vozlišči in povedki kot relacijami. Primer uporabe tega pristopa so Kansas Event Data System[23], International Relations Data Site[24] Paula Hensela in Correlates of War[25]. Pregled tega pristopa je opisan v knjigi From Words to Numbers [62]. Zanimiv je tudi pristop Steva Cormena Centering Resonance Analysis[26].

Še en primer takšnih podatkov so sosedska omrežja. Naj bo $\mathscr{V}$ množica (multivariatnih) enot in naj bo $d(u, v)$ *različnost*, ki je definirana na tej množici. Definiramo dve vrsti omrežij: omrežje *k-najbližjih sosedov* $\mathscr{N}(k) = (\mathscr{V}, \mathscr{A}, w)$, v katerem so povezave definirane kot:

$$(u, v) \in \mathscr{A} \Leftrightarrow v \text{ je med } k \text{ najbližjimi sosedi vozlišča } u, \quad w(u, v) = d(u, v);$$

in omrežje *r-sosedov* $\mathscr{N}(r) = (\mathscr{V}, \mathscr{E}, w)$, v katerem so povezave definirane kot:

$$(u : v) \in \mathscr{E} \Leftrightarrow d(u, v) \leq r, \quad w(u, v) = w(v, u) = d(u, v).$$

Takšna omrežja omogočajo povezavo med (multivariatno) analizo podatkov in analizo omrežij. Pri večjih množicah enot se pojavi problem oblikovanja učinkovitega algoritma za določanje najbližjih sosedov. David M. Mount je v Approximate Nearest Neighbor Library[27] opisal hitri algoritem za (aproksimacijsko) iskanje najbližjih sosedov. V programskem jeziku R so tovrstni algoritmi na voljo z uporabo funkcije `ann` iz paketa `yaImpute`.

[23] http://web.ku.edu/keds/

[24] http://www.paulhensel.org/data.html

[25] http://www.correlatesofwar.org/

[26] http://www.crawdadtech.com/

[27] http://www.cs.umd.edu/~mount/ANN

3.1.6 Pomenski splet

Pomenski splet (semantic web) [63] je nadgradnja in razširitev navadnega omrežja. Omogoča uporabo podatkovne plasti v svetovnem spletu in v spletnih storitvah. Osnova pomenskega spleta je semantični opis spletne vsebine z uporabo metapodatkov in ontologij. Uporablja se za pretvorbo omrežja nestrukturiranih dokumentov v omrežje podatkov. To olajša analizo teh podatkov, saj so s tem podatki že pretvorjeni v obliko omrežja.

Pomenski splet se opira na spletne tehnologije: Uniform Resource Identifier (URI), Resource Description Framework (RDF) in Web Ontology Language (OWL). URI je niz, ki identificira ime ali vir in omogoča interakcijo s predstavitvijo vira v omrežju z uporabo posebnega protokola. RDF je standardni model W3C za izmenjano podatkov na svetovnem spletu. Uporablja se za konceptualno opisovanje ali modeliranje informacij iz spletnega vira in za iskanje znanja z računalniki. RDF je osnova za predelavo metapodatkov; omogoča notranjo operabilnost med aplikacijami, ki si izmenjujejo strojno razumljive informacije na svetovnem spletu. OWL je družina jezikov za zapis ontologij.

V RDF je delček znanja predstavljen v obliki trojice osebek-povedek-predmet. Osebek označuje vir, predikat označuje vidik vira in izrazi relacijo med osebkom in predmetom. Viri so vedno poimenovani z URIji in morebitnimi dodanimi IDji (podmnožici sta množici URLjev in URNjev). Trojice tvorijo večrelacijsko omrežje z osebki in predmeti kot vozlišči in predikati kot vrstami relacij. Večje zbirke trojic RDF so shranjene v Linked Data - Connect Distributed Data across the Web[28], Freebase[29] in DBpedia[30].

Obstaja več različnih sintaktičnih formatov, ki se v kompleksnosti precej razlikujejo: N3, N-Triples, TRiG, TRiX, Turtle, RDF/XML, RDFa, JSON-LD. Namen RDF-a je zagotavljanje kodirnega in razlagalnega mehanizma, s katerim lahko vire opišemo v takšni obliki, da jih z njim skladna programska oprema razume. Nekateri formati so namenjeni predvsem strojnemu branju in so zato človeku težko berljivi. Primer poizvedbenega jezika za RDF je SPARQL[31].

[28] http://linkeddata.org/

[29] http://www.freebase.com/

[30] http://en.wikipedia.org/wiki/DBpedia

[31] http://www.w3.org/TR/rdf-sparql-query/

3.1.7 Naključna omrežja

Tvorjenje naključnih omrežij [64, 65] je pomembno za raziskovanje kompleksnih sistemov, recimo električnega omrežja, družabnih povezav, svetovnega spleta in interneta, omrežja sodelovanj in citiranj med znanstveniki idr. Naključna omrežja uporabljamo za modeliranje razredov grafov.

Paul Erdős in Alfréd Rényi sta v [66] predlagala pristop za formalizacijo naključnega grafa. Model *Erdős-Rényi,* ki ga označimo z $\mathcal{G}(n,m)$ za n vozlišč in m povezav, tvori naključni graf z n vozlišči in m povezavami, ki so (enakomerno) naključno izbrane med $\frac{n(n-1)}{2}$ možnimi povezavami.

Podoben Erdős-Rényijevemu modelu je *Gilbertov* model $\mathcal{G}(n,p)$ [67] za n vozlišč in verjetnost p za izbiro posamezne povezave. V tem modelu je vsaka izmed $\frac{n(n-1)}{2}$ možnih povezav preprostega neusmerjenega grafa $G(n,p) \in \mathcal{G}(n,p)$ vključena neodvisno od ostalih povezav z verjetnostjo p.

Watts in Strogatz sta predstavila model *majhnih svetov* [68]. Ta razred naključnih omrežij temelji na dveh strukturnih funkcijah. Koeficient nakopičenosti je visok, povprečna razdalja med pari vozlišč pa je majhna. Karakteristike majhnih svetov imajo družabna omrežja, internet in genska omrežja.

Porazdelitev stopenj naključnega grafa iz Erdős-Rényijevega ali Gilbertovega modela je skoncentrirana okoli povprečne stopnje – Poissonova porazdelitev. V večini omrežij iz dejanskega življenja pa porazdelitev stopenj v grobem sledi potenčnemu zakonu. Za takšna omrežja rečemo, da so *brezlestvična.* Barabási in Albert [69] sta opisala postopek *prednostnega dodajanja,* s katerim tvorimo omrežja s to lastnostjo. Prednostno dodajanje v vsakem trenutku ustvari eno vozlišče in ga poveže s podanim številom obstoječih vozlišč. Verjetnost izbire nekega soseda je sorazmerna z njegovo trenutno stopnjo.

Različne razrede naključnih grafov lahko opišemo tudi kot *verjetnostne induktivne razrede* grafov [70].

3.2 Analiza bibliografskih podatkov iz baze Web of Science

Omrežje sodelovanj je ponavadi definirano takole: Množica *vozlišč* je množica *avtorjev. Povezava* (neusmerjena) med avtorjema u in v obstaja natanko takrat, ko sta avtorja sodelovala pri izdelavi skupnega *dela* (članek, knjiga, poročilo idr.). *Utež* te povezave $w(u,v)$ je enaka številu skupnih del avtorjev u in v.

Omrežje sodelovanj temelji na še preprostejšem *dvovrstnem* omrežju, v katerem so dela povezava z avtorji. V tem omrežju obstaja *usmerjena povezava* od dela p do avtorja u natanko takrat, ko je u (so)avtor dela p. Če to omrežje soavtorstev $\mathscr{W}\mathscr{A}$ predstavimo z matriko **WA**, lahko izračunemo matriko omrežja sodelovanj kot $\mathbf{WA}^T \star \mathbf{WA}$. Ker lahko vsako omrežje predstavimo z matriko, bomo v nadaljevanju uporabljali matrične oznake omrežij: namesto $\mathscr{W}\mathscr{A}$ bomo za to omrežje uporabili oznako **WA**.

Matrike velikih omrežij zasedejo v standardni predstavitvi preveč prostora glede na to, da je večina elementov teh matrik ničelnih. Zato uporabimo 'vzporedno' operacijo množenja omrežij, ki ničelnih elementov v matrikah ne upošteva (glej podpoglavje 2.3).

Za dano množico del lahko poleg omrežja soavtorstev **WA** tvorimo tudi druga dvovrstna omrežja: **WK** za dela × ključne besede, **WC** za dela × klasifikacije, **WJ** za dela × revije idr. Ker imajo vsa ta omrežja isto prvo množico (množico del), lahko z množenjem omrežij dobimo različna *izpeljana* omrežja. Na primer z množenjem $\mathbf{WA}^T \star \mathbf{WK} = \mathbf{AK}$ dobimo dvovrstno omrežje **AK** za avtorje × ključne besede, v katerem so uteži $w(u,k)$ usmerjenih povezav enake številu del, v katerih je avtor u uporabil ključno besedo k. Dodatna izpeljana omrežja dobimo, če uporabimo tudi enovrstno omrežje citiranj med deli **Ci**.

V tem podpoglavju najprej pokažemo, kako lahko pretvorimo katerokoli tabelo podatkov v zbirko pripadajočih dvovrstnih omrežij. Ker imamo zbirko usklajenih dvovrstnih omrežij (vsa omrežja imajo eno množico vozlišč isto), lahko z množenjem omrežij dobimo različna izpeljana omrežja. Glavni del tega podpoglavja se ukvarja z ustrezno normalizacijo uteži povezav v izpeljanih omrežjih, kar prikažemo na primeru omrežja sodelovanj. Opisani pristop lahko uporabimo tudi za ostala izpeljana omrežja. Na koncu podpoglavja so predstavljena še nekatera druga izpeljana omrežja za primer bibliografskih omrežij.

Predstavljene pristope uporabimo na zbirki omrežij `SN5`. Podatki za ta omrežja so bili zbrani leta 2008 iz bibliografske baze Web of Science z iskalnim nizom `` `social network*' `` in razširjeni z obstoječimi opisi najpogostejših referenc in bibliografij približno 100 raziskovalcev s področja družbenih omrežij. Z uporabo programa `WoS2Pajek` [24] smo to zbirko podatkov pretvorili v zbirko omrežij: omrežje dela × avtorji, dela × ključne besede idr., omrežje citiranj; razbitje del po letu objave in razbitje del na tista s polnim opisom in samo citirane. Velikosti množic so naslednje: dela $|\mathbf{W}| = 193376$, dela s polnim opisom $|\mathbf{C}| = 7950$, avtorji $|\mathbf{A}| = 75930$, revije $|\mathbf{J}| = 14651$, ključne besede $|\mathbf{K}| = 29267$. Analize omrežij so bile narejene z orodjem za analize in prikaze velikih omrežij `Pajek` [29].

3.2.1 *Dvovrstna omrežja iz tabel podatkov*

Kakor je že omenjeno v podpoglavju 3.1, lahko omrežja pridobimo med drugim tudi iz tabel podatkov. *Tabela podatkov* T je množica *zapisov* $T = \{T_k : k \in \mathcal{K}\}$, pri čemer je $\mathcal{K}$ množica *ključev*. Vsak zapis je oblike $T_k = (k, q_1(k), q_2(k), \ldots, q_r(k))$, pri čemer je $q_i(k)$ vrednost *lastnosti* (atributa) $\mathbf{q}_i$ za ključ k.

Predpostavimo, da ima lastnost $\mathbf{q}$ zalogo vrednosti $2^{\mathcal{Q}}$. Na primer za SNA [41] :

Avtorji(SNA) = { S. Wasserman, K. Faust },
LetoObjave(SNA) = { 1994 },
KljučneBesede(SNA) = { network, centrality, matrix, …}, …

ali za [20, 26, 30, 41, 71]:

delo	avtorji	leto
…		
SNA	S. Wasserman, K. Faust	1994
GenCores	V. Batagelj, M. Zaveršnik	2011
Islands	M. Zaveršnik, V. Batagelj	2004
ESNA2	W. de Nooy, A. Mrvar, V. Batagelj	2012
IFCS09	N. Kejžar, S. Korenjak, V. Batagelj	2010
…		

V tej tabeli je delo ključ, avtorji in leto pa sta lastnosti.

Če je zaloga vrednosti $\mathcal{Q}$ končna, lahko lastnosti $\mathbf{q}$ priredimo dvovrstno omrežje $\mathcal{K} \times \mathbf{q} = ((\mathcal{K}, \mathcal{Q}), \mathcal{A}, w)$, pri čemer je $(k, v) \in \mathcal{A}$ natanko takrat, ko je $v \in q(k)$ in $w(k, v) = 1$. Množico $\mathcal{Q}$ lahko vedno pretvorimo v končno množico z uporabo razbitja vrednosti na podmnožice (intervale). Lastnosti z eno samo vrednostjo lahko bolj zgoščeno predstavimo z razbitjem.

Iz podatkov v bibliografski bazi Web of Science (Knowledge)[32] lahko pripadajoča omrežja dobimo z uporabo programa `WoS2Pajek`. Podobni programi obstajajo tudi za druge bibliografske baze in formate: BiBTeX, DBPL, IMDB, Zentralblatt Math idr.

[32] http://home.izum.si/izum/ft_baze/wos.asp

3.2.2 Sodelovanja

Omrežje soavtorstev

Naj bo **WA** dvovrstno omrežje soavtorstev za dela × avtorji; $wa_{pi} \in \{0,1\}$ opisuje avtorstvo avtorja i dela p. Za vsako delo $p \in \mathbf{W}$ velja:

$$\sum_{i \in \mathbf{A}} wa_{pi} = \mathrm{outdeg}(p).$$

Izhodna stopnja $\mathrm{outdeg}(p)$ je enaka številu avtorjev dela p.

Naj bo **N** normalizirana različica omrežja soavtorstev z utežmi n_{pi}, ki opisujejo delež prispevka avtorja i k delu p, tako da za vsako delo $p \in \mathbf{W}$ velja:

$$\sum_{i \in \mathbf{A}} n_{pi} \in \{0,1\}.$$

Vsota je enaka nič za dela brez navedenih avtorjev.

Prispevki n_{pi} so lahko določeni s podanimi pravili ali pa s predpostavko, da vsak avtor prispeva enakovreden delež k delu, torej jih lahko iz **WA** izračunamo kot

$$n_{pi} = \frac{wa_{pi}}{\max(1, \mathrm{outdeg}(p))}.$$

Podobna normalizacija uteži sodelovalnih povezav je bila predlagana že v [72]. Razlika je v tem, da v predlagani normalizaciji avtor uporabi $\mathrm{outdeg}(p) - 1$ namesto $\mathrm{outdeg}(p)$. Avtor te normalizacije uteži interpretira kot delež časa, ki ga posamezen avtor porabi za sodelovanje z vsakim soavtorjem.

Normalizacija omrežja po vrsticah je omrežje $n(\mathcal{N})$, ki ga dobimo iz $\mathcal{N}$ tako, da utež vsake povezave a delimo z vsoto uteži vseh povezav z istim začetim krajiščem kot a. Za dvojiško omrežje **A** z množicama vozlišč $\mathbf{I} \times \mathbf{J}$ izračunamo normalizacijo kot

$$n(\mathbf{A}) = \mathrm{diag}\left(\frac{1}{\max(1, \mathrm{outdeg}(i))}\right)_{i \in \mathbf{I}} \star \mathbf{A}.$$

Torej lahko normalizirano omrežje soavtorstev dobimo kot

$$\mathbf{N} = n(\mathbf{WA}).$$

Definiramo lahko povratno transformacijo – *binarizacijo* $b(\mathcal{N})$ omrežja $\mathcal{N}$: rezultat je izvorno omrežje, v katerem so vse uteži postavljene na ena. Velja

$$\mathbf{WA} = b(\mathbf{N}).$$

Velja tudi $\mathbf{WA} = b(n(\mathbf{WA}))$.

Še ena koristna transformacija je transponiranje. *Transponirano* omrežje $\mathcal{N}^T$ oziroma $t(\mathcal{N})$ je omrežje, ki ga iz omrežja $\mathcal{N}$ dobimo tako, da vsem povezavam obrnemo usmerjenost in pri dvovrstnih omrežjih zamenjamo vrstni red množic vozlišč. Za bibliografska omrežja uvedemo okrajšani zapis $\mathbf{AW} = \mathbf{WA}^T$, $\mathbf{KW} = \mathbf{WK}^T$ itd.

Prvo omrežje sodelovanj

Po standardni poti dobimo omrežje sodelovanj **Co** iz omrežja soavtorstev z množenjem:

$$\mathbf{Co} = \mathbf{AW} \star \mathbf{WA}.$$

Iz

$$co_{ij} = \sum_{p \in \mathbf{W}} wa_{pi} wa_{pj} = \sum_{p \in N(i) \cap N(j)} 1$$

vidimo, da je posamezna utež co_{ij} enaka številu del, ki sta jih avtorja i in j napisala skupaj.

Uteži so v prvem omrežju sodelovanj simetrične:

$$co_{ij} = \sum_{p \in \mathbf{W}} wa_{pi} wa_{pj} = \sum_{p \in \mathbf{W}} wa_{pj} wa_{pi} = co_{ji}.$$

Z utežmi co_{ij} lahko lahko določimo Saltonovo kosinusno mero podobnosti [73] ali Ochiaijev koeficient med posameznima avtorjema i in j kot

$$\cos(i, j) = \frac{co_{ij}}{\sqrt{co_{ii} co_{jj}}}, \qquad \text{za } co_{ij} > 0.$$

Ena izmed značilnih skupin avtorjev, ki jo lahko poiščemo v omrežju sodelovanj, je skupina avtorjev z največ soavtorji. Tovrstni avtorji iz množice `SN5` so navedeni v tabeli 3.1. V tabeli 3.2 pa je prikazana porazdelitev izhodne stopnje avtorjev v množici `SN5`. Izhodna stopnja avtorja je enaka številu del, pri katerih je sodeloval.

Tabela 3.1

Seznam avtorjev z največ sodelovanji v podatkih SN5.

Zap. št.	Avtor	Št. soavtorjev	Zap. št	Avtor	Št. soavtorjev
1	Snijders, T.	77	11	Rothenberg, R.	58
2	Krackhardt, D.	71	12	Doreian, P.	56
3	Wasserman, S.	65	13	Breiger, R.	56
4	Ferligoj,.A.	63	14	Valente, T.	52
5	Berkman,.L.	63	15	Butts, C.	52
6	van Duijn, M.	63	16	Goodreau, S.	52
7	Donovan, D.	62	17	Draper, D.	51
8	Friedman, S.	60	18	Batagelj, V.	51
9	Latkin, C.	59	19	Barabasi, A.	51
10	Faust, K.	59	20	Kelly, J.	50

Tabela 3.2

Porazdelitev izhodne stopnje v **WA**(SN5).

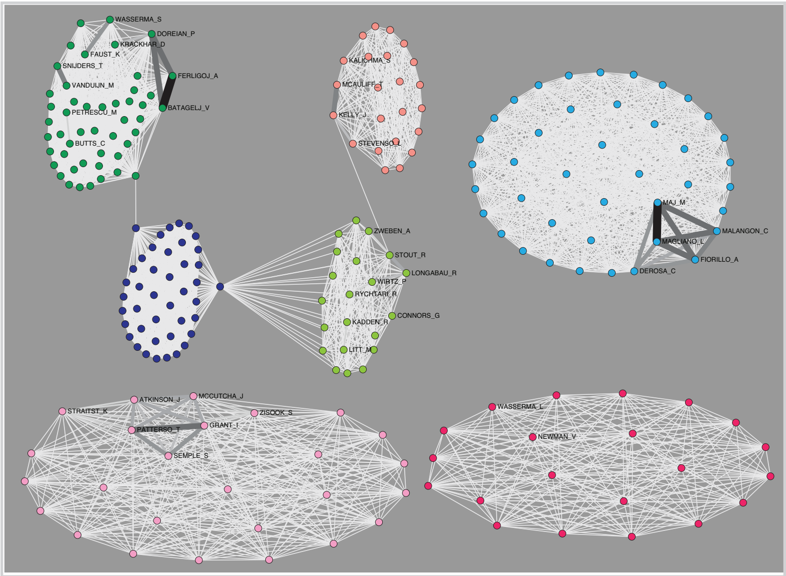
Izhodna stopnja	Frekvenca	Izhodna stopnja	Frekvenca	Članek
1	2637	12	8	
2	2143	13	4	
3	1333	14	3	
4	713	15	2	
5	396	21	1	[5]
6	206	22	1	[1]
7	114	23	1	[3]
8	65	26	1	[6]
9	43	41	1	[4]
10	24	42	1	[2]
11	10	48	1	[7]

Med raziskovanjem omrežja sodelovanj se pojavi tudi očitno vprašanje o tem, kdo je najbolj sodelovalen avtor. Standardni odgovor na to vprašanje dobimo s pomočjo navadnih sredic [26].

Podmnožica $\mathscr{U} \subseteq \mathscr{V}$ vozlišč določa *(navadno) sredico* reda k $\mathscr{C} = (\mathscr{U}, \mathscr{L}|\mathscr{U})$ v omrežju $\mathscr{N} = (\mathscr{V}, \mathscr{L})$ natanko takrat, ko za vsako vozlišče $u \in \mathscr{U}$ velja $\deg_{\mathscr{C}}(u) \geq k$ in je množica $\mathscr{U}$ maksimalna takšna množica. Podmnožica povezav $\mathscr{L}|\mathscr{U}$ je sestavljena iz povezav v $\mathscr{L}$, ki imajo obe krajišči v množici $\mathscr{U}$.

V omrežju sodelovanj je navadna sredica reda k največje podomrežje, v katerem velja, da je vsak avtor sodeloval z vsaj k drugimi avtorji iz te sredice.

Na sliki 3.1 so prikazane navadne sredice redov 20–47. Opazimo lahko zelo slabo stran neposredne uporabe navadnih sredic v analizi omrežja sodelovanj. Vsako delo s k avtorji prispeva v omrežju sodelovanj poln podgraf na k vozliščih. Za bibliografije del z



Slika 3.1

Navadne sredice redov 20–47 v omrežju **Co**(SN5).

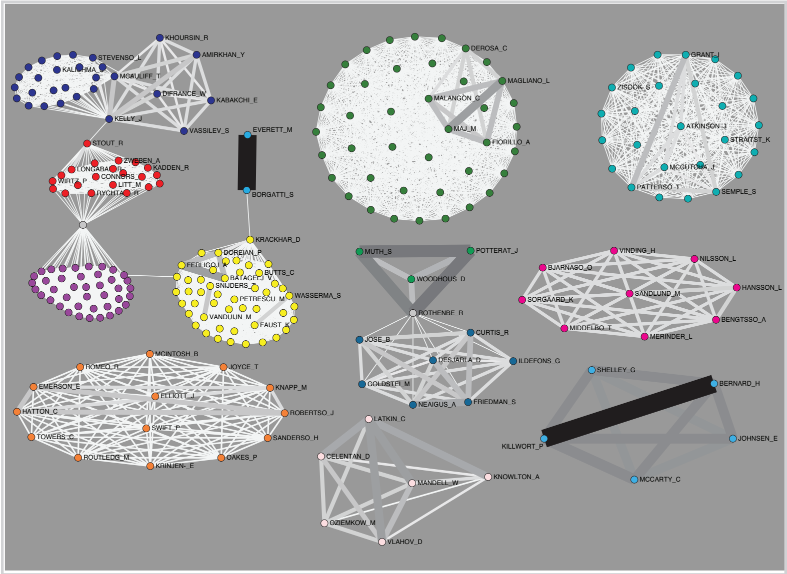
veliko avtorji metoda navadnih sredic najde za najvišji red sredice, ki predstavljajo tem delom pripadajoče polne podgrafe, in ne dejansko najbolj sodelovalne avtorje, kakor bi pričakovali.

Za bibliografije SN5 so navadne sredice redov 20–47 posledica člankov, ki pravzaprav predstavljajo dela z največ avtorji (21-48), kakor je prikazano v tabeli 3.2 in v dodatku A. Na sliki so prikazana imena le tistih avtorjev, ki so krajišča povezav z utežjo večjo od ena.

Tega problema bi se lahko lotili tako, da bi odstranili vse povezave z utežjo ena (ali do kakega drugega majhnega praga) in na dobljenem omrežju uporabili metodo navadnih sredic.

Boljša rešitev je identifikacija del s (pre)več avtorji – z visoko izhodno stopnjo v omrežju **WA** – in njihova odstranitev iz omrežja **WA** za namene te analize. Odstranjena dela lahko pregledamo ločeno.

Še en možen pristop pa je uporaba p_S-sredic [26] na omrežju sodelovanj **Co** – posplošitev navadnih sredic, v kateri stopnjo $\deg_{\mathscr{C}}(u)$ zamenjamo z vsoto uteži povezav



Slika 3.2

p_S-sredica reda 20 omrežja sodelovanj **Co**(SN5).

od vozlišča u do ostalih vozlišč v $\mathscr{U}$:

$$p_S(u, \mathscr{U}) = \sum_{v \in \mathscr{U}} w(u, v).$$

Podmnožica $\mathscr{U} \subseteq \mathscr{V}$ vozlišč določa p_S-*sredico* $\mathscr{C} = (\mathscr{U}, \mathscr{L}|\mathscr{U})$ reda t v omrežju $\mathscr{N} = (\mathscr{V}, \mathscr{L})$ natanko takrat, ko za vsako vozlišče $u \in \mathscr{U}$ velja $p_S(u, \mathscr{U}) \geq t$ in je množica $\mathscr{U}$ maksimalna takšna množica.

Na sliki 3.2 je prikazana p_S-sredica reda 20. Vsak avtor v njej ima vsaj 20 del z ostalimi avtorji v tej sredici.

Tudi v p_S-sredici omrežja SN5 se pojavijo klike, ki pripadajo člankom z velikim številom avtorjev. Poleg teh klik opazimo tudi močne sodelovalne skupine: { S. Borgatti, M. Everett }, { H. Bernard, P. Killworth, C. McCarty, E. Johnsen, G. Shelley }, { R. Rotenberg, S. Muth, J. Potterat, D. Woodhouse }, { L. Magliano, M. Maj, C. Malangon, A. Fiorillo } in ostale.

Tudi p_S-sredice ne dajo povsem zadovoljivega odgovora. Kadar želimo nevtralizirati prevlado prispevkov del z veliko avtorji, lahko to poskusimo rešiti z alternativnimi definicijami omrežja sodelovanj, pri čemer uporabimo normalizirano omrežje soavtorstev. Struktura (graf) omrežja sodelovanj ostane enaka, spremenijo pa se uteži povezav.

Drugo omrežje sodelovanj

je definirano kot

$$\mathbf{Cn} = \mathbf{AW} \star \mathbf{N}.$$

Vrednosti uteži cn_{ij} so definirane kot

$$cn_{ij} = \sum_{p \in \mathbf{W}} wa_{pi} n_{pj} = \sum_{p \in N(i) \cap N(j)} n_{pj}$$

in so enake prispevku avtorja j delom, pri katerih je sodeloval z avtorjem i.

V splošnem vrednosti matrike **Cn** niso vedno simetrične ($cn_{ij} = cn_{ji}$). V primeru, ko velja $n_{pi} = \frac{wa_{pi}}{\mathrm{outdeg}_{\mathbf{WA}}(p)}$, pa so

$$cn_{ij} = \sum_{p \in N(i) \cap N(j)} n_{pj} = \sum_{p \in N(i) \cap N(j)} \frac{1}{\mathrm{outdeg}_{\mathbf{WA}}(p)} = \sum_{p \in N(j) \cap N(i)} n_{pi} = cn_{ji}.$$

Skupen prispevek izrazov $wa_{pi} n_{pj}$ dela p iz definicije cn_{ij}

$$\sum_{j \in \mathbf{A}} \sum_{j \in \mathbf{A}} wa_{pi} n_{pj} = \sum_{i \in N(p)} \sum_{j \in \mathbf{A}} n_{pj} = \sum_{i \in N(p)} 1 = \mathrm{outdeg}_{\mathbf{WA}}(p)$$

je enak številu avtorjev dela p.

Podobno za avtorja i velja, da je skupni prispevek vrednosti cn_{ij}

$$\sum_{j \in \mathbf{A}} cn_{ij} = \sum_{j \in \mathbf{A}} \sum_{p \in \mathbf{W}} wa_{pi} n_{pj} = \sum_{p \in \mathbf{W}} wa_{pi} \sum_{j \in \mathbf{A}} n_{pj} = \sum_{p \in \mathbf{W}} wa_{pi} = \mathrm{indeg}_{\mathbf{WA}}(i)$$

enak številu del, pri katerih je avtor i sodeloval. Diagonalne vrednosti

$$cn_{ii} = \sum_{p \in N(i)} n_{pi}$$

so enake celotnemu prispevku avtorja i njegovim delom.

Torej lahko na podlagi elementov matrike **Cn** definiramo kazalnik *samozadostnosti*

$$S_i = \frac{cn_{ii}}{\mathrm{indeg}_{\mathbf{WA}}(i)}$$

Tabela 3.3

Seznam 'najboljših' avtorjev v SN5.

Zap. št.	Avtor	cn_{ii}	Skupaj	K_i
1	Burt, R.	43.83	53	0.173
2	Newman, M.	36.77	60	0.387
3	Doreian, P.	34.44	47	0.267
4	Bonacich, P.	30.17	41	0.264
5	Marsden, P.	29.42	37	0.205
6	Wellman, B.	26.87	41	0.345
7	Leydesdorf, L.	24.37	35	0.304
8	White, H.	23.50	33	0.288
9	Friedkin, N.	20.00	23	0.130
10	Borgatti, S.	19.20	41	0.532
11	Everett, M.	16.92	31	0.454
12	Litwin, H.	16.00	21	0.238
13	Freeman, L.	15.53	20	0.223
14	Barabasi, A.	14.99	35	0.572
15	Snijders, T.	14.99	30	0.500
16	Valente, T.	14.80	34	0.565
17	Breiger, R.	14.44	20	0.278
18	Skvoretz, J.	14.43	27	0.466
19	Krackhardt, D.	13.65	25	0.454
20	Carley, K.	12.93	28	0.538
21	Pattison, P.	12.10	27	0.552
22	Wasserman, S.	11.72	26	0.549
23	Berkman, L.	11.21	30	0.626
24	Moody, J.	10.83	15	0.278
25	Scott, J.	10.47	15	0.302
26	Latkin, C.	10.14	37	0.726
27	Morris, M.	9.98	20	0.501
28	Rothenberg, R.	9.82	28	0.649
29	Kadushin, C.	9.75	11	0.114
30	Faust, K.	9.72	18	0.460
31	Batagelj, V.	9.69	20	0.516
32	Mizruchi, M.	9.67	15	0.356
33	[Anon]	9.00	9	0.000
34	Johnson, J.	8.89	21	0.577
35	Fararo, T.	8.83	16	0.448
36	Lazega, E.	8.50	12	0.292
37	Knoke, D.	8.33	11	0.242
38	Ferligoj, A.	8.19	19	0.569
39	Brewer, D.	8.03	11	0.270
40	Klovdahl, A.	7.96	17	0.532
41	Hammer, M.	7.92	10	0.208
42	White, D.	7.83	15	0.478
43	Holme, P.	7.42	14	0.470
44	Boyd, J.	7.37	13	0.433
45	Kilduff, M.	7.25	16	0.547
46	Small, H.	7.00	7	0.000
47	Iacobucci, D.	7.00	12	0.417
48	Pappi, F.	6.83	10	0.317
49	Chen, C.	6.78	12	0.435
50	Seidman, S.	6.75	9	0.250

kot delež avtorjevega prispevka v skupnem številu del, pri katerih je sodeloval.

Kazalnik *sodelovalnosti* je komplement kazalniku samozadostnosti

$$K_i = 1 - S_i.$$

Vse vrednosti cn_{ij} se seštejejo v število vseh povezav v omrežju $\mathbf{WA}$:

$$\sum_{i \in \mathbf{A}} \sum_{j \in \mathbf{A}} cn_{ij} = \sum_{i \in \mathbf{A}} \mathrm{indeg}_{\mathbf{WA}}(i) = m_{\mathbf{WA}}.$$

V tabeli 3.3 je seznam 50 avtorjev z največjim prispevkom cn_{ii} na področju analize družbenih omrežij. Dodano je tudi število vseh del posameznega avtorja o tej temi in njegov kazalnik sodelovalnosti.

Tretje omrežje sodelovanj

je definirano kot

$$\mathbf{Ct} = \mathbf{N}^T \star \mathbf{N}.$$

Utež ct_{ij} je enaka skupnemu prispevku sodelovanj avtorjev i in j.

Skupen prispevek avtorjev v polnem podgrafu nekega dela je enak 1 :

$$\sum_{i \in \mathbf{A}} \sum_{j \in \mathbf{A}} n_{ip}^T n_{pj} = \sum_{i \in \mathbf{A}} n_{pi} \sum_{j \in \mathbf{A}} n_{pj} = \sum_{i \in \mathbf{A}} n_{pi} \cdot 1 = 1.$$

Uteži ct_{ij} so simetrične

$$ct_{ij} = \sum_{p \in \mathbf{W}} n_{ip}^T n_{pj} = \sum_{p \in \mathbf{W}} n_{jp}^T n_{pi} = ct_{ji}$$

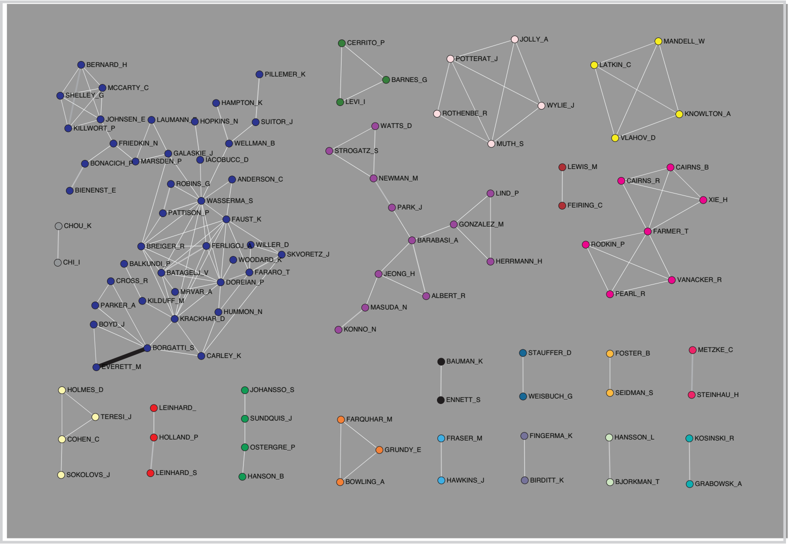
in vsota

$$\sum_{j \in \mathbf{A}} ct_{ij} = \sum_{j \in \mathbf{A}} \sum_{p \in \mathbf{W}} n_{pi} n_{pj} = \sum_{p \in \mathbf{W}} n_{pi} \sum_{j \in \mathbf{A}} n_{pj} = \sum_{p \in \mathbf{W}} n_{pi}$$

je enaka celotnemu prispevku avtorja i delom iz množice $\mathbf{W}$.

Vsota vseh uteži ct_{ij}

$$\sum_{i \in \mathbf{A}} \sum_{j \in \mathbf{A}} ct_{ij} = \sum_{i \in \mathbf{A}} \sum_{p \in \mathbf{W}} n_{pi} = \sum_{p \in \mathbf{W}} \sum_{i \in \mathbf{A}} n_{pi} = \sum_{p \in \mathbf{W}} 1 = |\mathbf{W}|$$



Slika 3.3

p_S-sredica reda 0.75 v tretjem omrežju sodelovanj **Ct**(SN5).

je enaka številu vseh del.

Avtorjev prispevek področju lahko definiramo kot

$$ac_i = \frac{|\mathbf{A}|}{|\mathbf{W}|} \sum_{p \in \mathbf{W}} n_{pi},$$

pri čemer velja

$$\sum_{i \in \mathbf{A}} ac_i = |\mathbf{A}|.$$

Torej je povprečna vrednost ac enaka 1.

Opazimo tudi, da velja

$$b(\mathbf{Co}) = b(\mathbf{Cn}) = b(\mathbf{Ct}).$$

Na sliki 3.3 je prikazana p_S-sredica reda 0.75 v tretjem omrežju sodelovanj **Ct**(SN5). V tej sredici ni več velikih klik. Največja komponenta sredice je sestavljena iz vodilnih raziskovalcev iz področja družbenih omrežij z naslednjimi zelo močnimi pari: Borgatti in Everett, Killworth in Bernard, Bonacich in Bienestock, Ferligoj in Batagelj, Pattison in Robins idr. Druga največja komponenta sestoji iz fizikov z naslednjimi zelo močnimi pari: Newman in Park, Barabasi in Albert ter Masuda in Konno. V manjši komponenti najdemo še naslednje tri pare: Lienhard in Holland (pri čemer je Lienhard predstavljen z dvema vozliščema), Metzke in Steinhaus ter Chou in Chi.

3.2.3 Izpeljana omrežja

Sklicevalni pari in skupna citiranja

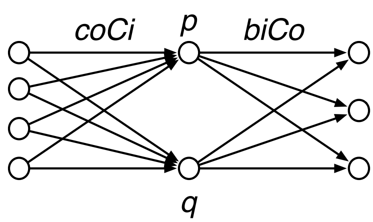
V programu `WoS2Pajek` je relacija citiranja definirana kot $p\,\mathbf{Ci}\,q \equiv p$ citira q. Torej lahko definiramo omrežje *sklicevalnih parov* **biCo** kot [74]

$$\mathbf{biCo} = \mathbf{Ci} \star \mathbf{Ci}^T.$$

Uteži v tem omrežju

$$bico_{pq} = \sum_{s \in \mathbf{W}} ci_{ps} ci_{qs} = \sum_{s \in N(p) \cap N(q)} 1$$

so enake številu del, na katera se sklicujeta deli p in q, kakor je prikazano na sliki 3.4. Uteži so simetrične: $bico_{pq} = bico_{qp}$.



Slika 3.4 Dela, na katera se sklicujeta deli p in q, so upoštevana v računanju uteži povezave pq v omrežju sklicevalnih parov. Dela, ki se sklicujejo na deli p in q, so upoštevana v računanju uteži povezave pq v omrežju sosklicevanj.

Zopet se pojavi problem del z veliko citati, še posebej problematični so pregledni članki. Njihov vpliv lahko nevtraliziramo z normalizacijo, recimo

$$\mathbf{biCon} = \frac{1}{2}(n(\mathbf{Ci}) \star \mathbf{Ci}^T + \mathbf{Ci} \star n(\mathbf{Ci})^T).$$

Hitro vidimo, da velja $bicon_{pq} \in [0, 1]$ in $bicon_{pq} = bicon_{qp}$ (simetrija). Velja tudi: $bicon_{pq} = 1$ natanko takrat, ko se deli p in q sklicujeta na isto delo.

Element cC_{pq} v prvem členu vsote te normalizacije predstavlja 'pomembnost navadnega' (p,q)-sklicevanja za delo p; element Cc_{pq} v drugem členu pa predstavlja 'pomembnost navadnega (p,q)-sklicevanja' za delo q.

$$bicon_{pq} = \frac{1}{2}(cC_{pq} + Cc_{pq})$$

Opazimo, da je prvi člen v definiciji omrežja **biCon** enak transponiranemu drugemu členu

$$(\mathbf{Ci} \star n(\mathbf{Ci})^T))^T = n(\mathbf{Ci}) \star \mathbf{Ci}^T$$

in zato velja $Cc_{pq} = cC_{qp}$. To dejstvo lahko uporabimo za učinkovitejši izračun omrežja **biCon**. Izračunati moramo namreč le prvi člen **cC**. Potem velja

$$bicon_{pq} = \frac{1}{2}(cC_{pq} + cC_{qp}).$$

V omrežju **biCon**(SN5) najdemo večjo komponento s povezavami z $bicon = 1$, katere vozlišča predstavljajo članke s sklicem na eno izmed knjig (Wasserman, S., Faust, K.: Social network analysis. Cambridge UP, 1994; Taylor, Howard F.: Balance in small groups. Van Nostrand Reinhold, 1970; Belle, D.: Childrens social networks and social supports. Wiley, 1989; Gottlieb, B. H.: Social networks and social support. Sage, 1981; Yan, Yunxiang: The flow of gifts. Stanford UP, 1996; Zhang, L.: Strangers in the City. Stanford UP, 2001). Najdemo tudi mnogo parov člankov z $bicon = 1$, ki jih je večinoma napisal isti avtor. Bolj zanimive skupine dobimo v obliki večjih otokov z vrednostmi pod ena.

Podobno lahko definiramo omrežje *sosklicevanj* **coCi** [75, 76] kot

$$\mathbf{coCi} = \mathbf{Ci}^T \star \mathbf{Ci}.$$

Pripadajoče uteži

$$coci_{pq} = \sum_{s \in \mathbf{W}} ci_{sp} ci_{sq} = \sum_{s \in N^-(p) \cap N^-(q)} 1$$

so enake številu del, ki se sklicujejo na deli p in q, kakor je prikazano na sliki 3.4. Z $N^-(p)$ označimo množico sosedov, od katerih lahko pridemo do vozlišča p.

Ostala izpeljana omrežja

Utež aci_{ip} v omrežju *sklicevanj avtorjev na dela*

$$\mathbf{ACi} = \mathbf{AW} \star \mathbf{Ci}$$

predstavlja število, kolikokrat se je avtor i skliceval na delo p.

Omrežje *sosklicevanj avtorjev* lahko dobimo kot

$$\mathbf{ACo} = b(\mathbf{ACi}) \star b(\mathbf{ACi})^T .$$

Utež aco_{ij} predstavlja število del, na katera sta se sklicevala avtorja i in j.

Utež ak_{ik} v omrežju *avtorjev in ključnih besed*

$$\mathbf{AK} = \mathbf{AW} \star \mathbf{WK}$$

pa predstavlja število del, v katerih je avtor i uporabil ključno besedo k.

Omrežje skupnih sklicevanj na dela

V [77] je predlagana naslednja definicija omrežja *skupnih sklicevanj na dela:*

$$\mathbf{ACA} = \mathbf{AW} \star \operatorname{diag}(\operatorname{indeg}_{\mathbf{Ci}}(p)) \star \mathbf{WA}.$$

Utež povezave med dvema avtorjema je enaka številu sklicevanj na dela, pri katerih sta sodelovala. Vrednost $\operatorname{indeg}_{\mathbf{Ci}}(p)$ je enaka številu citiranj delu p.

Definiramo lahko tudi normalizirano različico tega omrežja:

$$\mathbf{Cc} = \mathbf{AW} \star \operatorname{diag}\left(\frac{\operatorname{indeg}_{\mathbf{Ci}}(p)}{\operatorname{outdeg}_{\mathbf{Ci}}(p)^2}\right) \star \mathbf{WA},$$

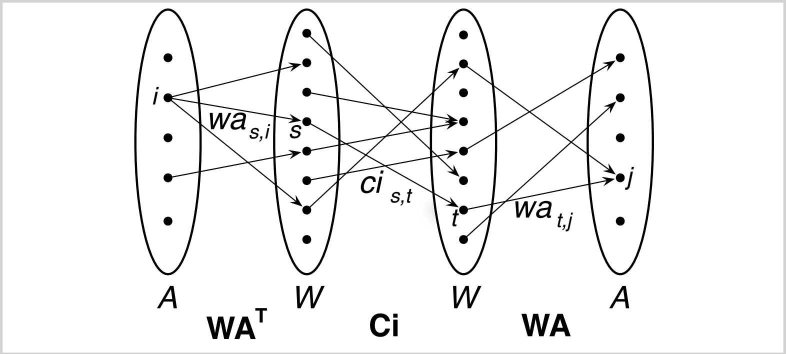
za katerega velja

$$\sum_{i\in\mathbf{A}} \sum_{j\in\mathbf{A}} wa_{ip} \frac{\operatorname{indeg}_{\mathbf{Ci}}(p)}{\operatorname{outdeg}_{\mathbf{Ci}}(p)^2} aw_{pj} = \operatorname{indeg}_{\mathbf{Ci}}(p)$$

in

$$\sum_{i\in\mathbf{A}} \sum_{j\in\mathbf{A}} cc_{ij} = \sum_{p\in\mathbf{W}} \operatorname{indeg}_{\mathbf{Ci}}(p) = |\mathscr{A}_{\mathbf{Ci}}|,$$

pri čemer je $|\mathscr{A}_{\mathbf{Ci}}|$ število povezav v omrežju sklicevanj $\mathbf{Ci}$.



Slika 3.5

Omrežje sklicevanj med avtorji.

Omrežje sklicevanj med avtorji

Omrežje sklicevanj med avtorji lahko dobimo kot

$$\mathbf{Ca} = \mathbf{AW} \star \mathbf{Ci} \star \mathbf{WA}.$$

Utež ca_{ij} predstavlja število del avtorja i, ki se sklicujejo na kakšno delo avtorja j (slika 3.5).

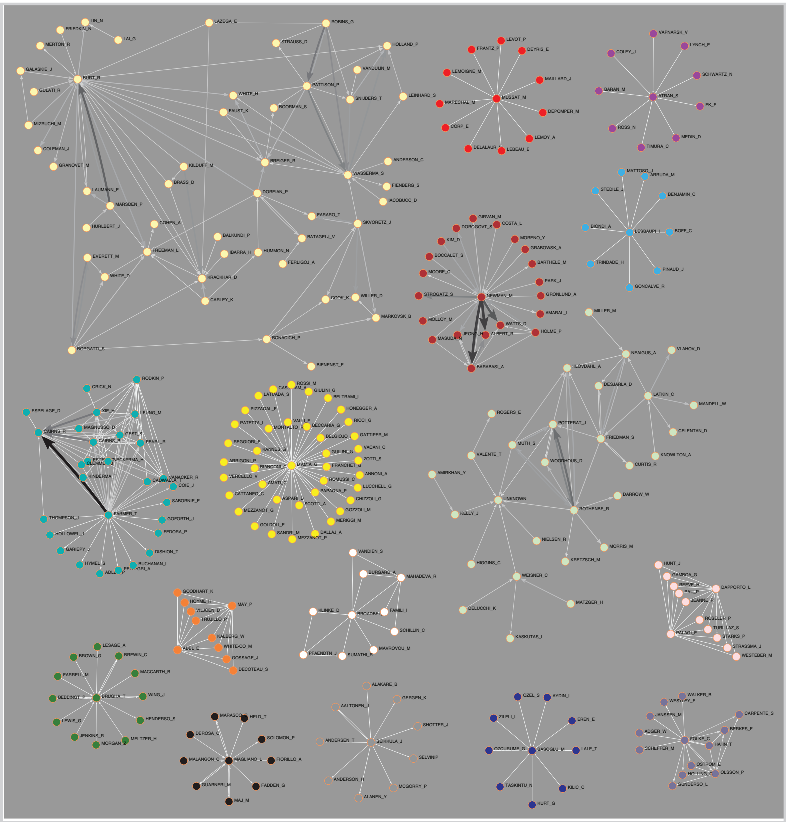
Na sliki 3.6 je prikazanih nekaj povezavnih otokov omrežja **Ca**(SN5). Največji otok je sestavljen iz glavnega otoka raziskovalcev družbenih omrežij z nekaj podskupinami: zvezda okoli R. Burt zgoraj levo, par S. Borgatti in M. Everett spodaj levo; verjetnostna skupina zgoraj desno s predstavniki G. Robins, P. Pattison, T. Snijders, S. Wasserman in P. Holland; in ostali: J. Skvoretz, D. Krackhardt, P. Doreian, R. Breiger, H. White, L. Freeman in P. Marsden.

'Brezlestvični' otok sestavljajo predvsem fiziki: M. Newman, A. Barabasi, D. Watts, R. Albert, P. Holme idr. V 'medicinskem' otoki so središčni avtorji J. Potterat, R. Rothenberg, D. Woodhouse, S. Muth, A. Klovdahl in S. Friedman. Najdemo tudi otok 'izobraževanja in psihologije' T. Farmer, R. Cairns, B. Cairns, H. Xie in P. Rodkin.

Večina otokov je v obliki zvezde, pri čemer je ponavadi v sredini profesor, krajna vozlišča pa so njegovi doktorski študenti.

3.2.4 Zaključek

V tem podpoglavju smo pokazali, da lahko z uporabo množenja omrežij dobimo različna zanimiva izpeljana omrežja. Predlagani pristop lahko uporabimo tudi na zbirki



Slika 3.6

Nekaj povezavnih otokov v omrežju **Ca**(SN5).

kakšnih drugačnih usklajenih omrežij [glej 78, str. 8260–8262].

Večina izpeljanih omrežij, ki smo jih predstavili, je enovrstnih omrežij. Na njih lahko uporabimo običajne pristope za analizo omrežij. Za analizo dvovrstnih omrežij pa lahko uporabimo neposredne metode, kot so (posplošene) dvovrstne sredice, dvovrstna kazala in opise in štirikotnike [27].

Omrežje sklicevanj (in ostala omrežja v tem podpoglavju) lahko pretvorimo v časovna omrežja z uporabo razbitja del glede na leto objave. S časovnimi rezinami lahko dobimo časovno zaporedje pripadajočih izpeljanih omrežij.

`Pajek` omogoča analizo na različnih nivojih, ki so definirani z razbitjem pripadajoče skupine enot in dobljeni s *skrčitvijo* razredov razbitja. Na primer razbitje avtorjev po ustanovah, razbitje ustanov po državah, razbitje avtorjev po disciplinah / področjih raziskave idr. Z uporabo *izreza* izbranih razredov razbitja lahko izberemo le del omrežja, ki nas zanima.

Opis uporabljenih metod v programu `Pajek` je dostopen na spletni strani
`http://pajek.imfm.si/doku.php?id=how_to:biblio`.

3.3 Analiza bibliografskih podatkov iz baze Zentralblatt MATH

Z bibliografskimi podatki lahko raziščemo razvoj nekega področja, najdemo najbolj sodelovalne avtorje, identificiramo področja raziskav, kjer avtorji močneje sodelujejo, identificiramo področja raziskav, kjer avtorji raje delajo sami ali v manjših skupinah idr. Analiza bibliografskih podatkov ne prispeva k razvoju raziskovalnih področij neposredno, ampak nam pomaga razumeti njihovo strukturo. Z analizo omrežij bibligrafskih podatkov se raziskovalci ukvarjajo že od E. Garfielda [79] dalje. V tem podpoglavju predstavljamo bibliografski vpogled v področje matematike v obdobju 1990-2010, kakor je zabeleženo v bibliografski bazi Zentralblatt MATH (ZB). Podatkovno bazo ZB vzdržujejo v uredniški pisarni FIZ Karlsruhe v Berlinu v sodelovanju z evropskimi akademijami in matematičnimi ustanovami.

V sodelovanju s prof. Berndom Wegnerjem in njegovimi sodelavci v FIZ Karlsruhe smo v januarju 2011 pridobili osnovne podatke o delih (člankih in knjigah) za obdobje 1990-2010, ki so shranjeni v bazi ZB. Analize bibliografskih podatkov smo se lotili z uporabo omrežij. Za izračune smo uporabili program `Pajek`[28, 30], orodje za analizo in prikaz velikih omrežij. V tem podpoglavju so predstavljeni rezultati osnovnih analiz omrežij (statistične informacije o podatkih) in identifikacija pomembnejših elementov (avtorjev, ključnih besed in revij).

V podpoglavju najprej opišemo podatke in opredelimo nekaj problemov, na katere smo naleteli pri pretvorbi podatkov v omrežja. Nato so opisane različne porazdelitve in analiza omrežja sodelovanj. V zadnjem delu tega podpoglavja se osredotočimo na izbrano področje matematike – teorijo grafov. Predstavimo analizo sodelovanj med grafovskimi teoretiki, značilne ključne besede in izstopajoče revije.

3.3.1 Podatki

V podatkovni bazi ZB je shranjenih več informacij o posameznem delu. Skupek informacij o nekem delu imenujemo *zapis*. Zapis je sestavljen iz različnih polj. Vsako polje ima dvoznakovni identifikator:

an – identifikacijska številka dela (določena v ZB),

ai – poenoteno imena avtorjev,

au – cela imena avtorjev,

py – leto objave,

cc – klasifikacija (Mathematical Subject Classification - MSC) [80],

ti – naslov dela,

ut – ključne besede,

is – identifikacijska številka ISSN revije (International Standard Serial Number),

so – naslov revije, strani dela v reviji, leto objave dela,

se – podatki o reviji (identifikacijska številka v ZB, celoten in kratek naslov, ISSN).

Primer zapisa:

```
an  01714102
ai  -; sastre-vazquez.patricia; -
is  ISSN 0368-492X
au  Us\'o-Dom\`enech, J.L.; Sastre-Vazquez, P.; Mateu, J.
py  2001
```

```
cc  *68U20
ti  Syntax and first entropic approximation of $L(M_T)$. A
      language for ecological modelling.
ut  modelling process; text-model based language
so  Kybernetes 30, No.9-10, 1304-1317 (2001).
se  00000540 Kybernetes Kybernetes 0368-492X
```

Problemi s podatki

Podatke o delih v podatkovno bazo ZB vnašajo uredniki. Najpogostejši problem so manjkajoči podatki – nekateri članki in knjige nimajo podanih vseh vrst informacij. Če manjkajočo informacijo potrebujemo za analize, to predstavlja dodaten problem.

Znaki, ki niso del tabele ASCII, so v besedilu predstavljeni z ukazi v TEXu, ki niso vedno enotni. Recimo

```
Must\u{a}\c{t}a, Costic\u{a}
Must\u a\c ta, Costic\u a
```

sta dva različna zapisa imena istega avtorja. Da bi rešili ta problem, potrebujemo skripto, ki prepozna vse različne zapise posameznega znaka in jih poenoti.

Imena avtorjev so v ZB le delno poenotena. Nekateri avtorji imajo poenoteno ime, drugi ne, spet tretji pa imajo več imen (sinonime). Ta problem ni enostavno rešljiv, saj bi morali pregledati vse avtorje in njihova poenotena imena ter jih popraviti. Obstajajo tudi avtorji, katerih imena so zapisana v več različnih oblikah. Recimo ime Mankoč Borštnik Norma Susana je zapisano kot

```
Bor\v stnik, N. S. Manko\v c,
Manko\v c Bor\v stnik, N.,
Manko\v c-Bor\v stnik, Norma,
Manko\v c Bor\v stnik, Norma Susana,
Mankoc-Borstnik, N.S. in
Manko\v c Bor\v stnik, N.S.
```

Poenotenje imen predstavlja problem še iz drugega razloga. Nekateri avtorji imajo zelo podobna ali celo enaka imena (homonimija). Zato se lahko zgodi, da imajo enaka tudi poenotena imena v ZB, kar predstavlja težavo pri razlikovanju med temi avtorji.

Poleg imen avtorjev smo imeli probleme tudi s ključnimi besedami. Ker vsa dela nimajo zapisanih ključnih besed, smo med ključne besede posameznega dela vključili tudi besede iz naslova dela. (Dejanske) ključne besede so pravzaprav fraze, ki so lahko sestavljene iz več besed. Te fraze smo ločili v besede in odstranili nepomenske besede (stop words). Vsebinsko povezane ključne besede smo poenotili z lematizacijo (MontyLingua). Tako smo recimo poenotili besedi `algebra` in `algebras`.

V bazi ZB imajo revije določena identifikacijska števila. To v osnovi reši problem poenotene identifikacije. Vendar smo tekom analiz našli revijo z dvema identifikatorjema:

```
00000552 Match Match 0340-6253
00003047 MATCH - Communications in Mathematical and in Computer
         Chemistry MATCH Commun. Math. Comput. Chem. 0340-6253
```

Oba zapisa smo obravnavali kot eno revijo.

Revije se spreminjajo skozi čas – lahko se pojavi nova revija, revija neha izhajati, nekaj revij se združi v eno revijo, revija se razdeli v več revij, reviji se spremeni naslov idr. Zaradi tovrstnih sprememb se več različic ene revije v bazi ZB pojavlja kot več različnih revij. Različne pojavitve posamezne revije, ki je spremenila naslov, smo združili.

Pripravljanje podatkov

S posebnim programom v Pythonu smo podatke pretvorili v format programa `Pajek` [28, 30]. Tako smo pridobili štiri usklajena dvovrstna omrežja in razbitje glede na leto objave.

Prva množica vozlišč v vseh štirih omrežjih je sestavljena iz identifikatorjev del, označimo jo z $\mathscr{W}$, vozlišča v drugi množici pa predstavljajo eno izmed sledečih množic:

- $\mathscr{A}$ – množica *avtorjev,*
- $\mathscr{J}$ – množica *revij,*
- $\mathscr{K}$ – množica *ključnih besed,*
- $\mathscr{M}$ – množica zapisov *MSC.*

Informacije o vozliščih smo pridobili iz zapisov o delih. Identifikator dela smo dobili iz polja `an`, revije iz polja `se` in zapise MSC iz polja `cc`. Ključne besede smo dobili iz polj `ut` in `ti` kot fraze in smo jih nato ločili v besede ter poenotili z lematizacijo. Imena avtorjev smo dobili iz polja `ai`. Če avtorjevo poenoteno ime v bazi ZB ne obstaja, smo ime dobili iz polja `au` in ga pretvorili v poenoteno obliko imena.

Pričakovali smo, da bo večina pomembnejših matematikov imela poenotena imena. Za ostala smo se odločili, da sinonime in homonime obravnavamo kot šum in jih posebej obravnavamo, kadar se pojavijo kot podvojitve v rezultatih analiz.

Povezave v vseh pridobljenih omrežjih so usmerjene in povezujejo posamezno delo z (nekaj) predstavniki iz druge množice. Omrežje soavtorstev $\mathbf{WA} = ((\mathscr{W}, \mathscr{A}), \mathscr{L}, w)$ je omrežje, v katerem je vsako delo povezano z vsemi avtorji tega dela, $(p, i) \in \mathscr{L} \Leftrightarrow i$ je avtor dela p. Ostala tri omrežja so definirana podobno – dela so povezana z revijami, ključnimi besedami in zapisi MSC v omrežjih del $\times$ revij **WJ**, del $\times$ ključnih besed **WK** in del $\times$ klasifikacij **WM**. V nadaljevanju bomo uporabljali poenostavljen zapis transponiranega omrežja: transponirano omrežje omrežja **WA** označimo z $\mathbf{AW} \equiv \mathbf{WA}^T$ in ga pridobimo iz **WA** z zamenjavo smeri vseh povezav. Velikosti vseh štirih omrežij so prikazane v tabeli 3.4.

Tabela 3.4

Velikosti dvovrstnih omrežij.

Matrika omrežja	**WA**	**WJ**	**WK**	**WM**
Velikost prve množice	1339201	1339201	1339201	1339201
Velikost druge množice	557104	3158	143513	12390
Število povezav	2550437	1331036	15062377	3370820

Kot smo že omenili, smo naleteli na problem z zapisi imen avtorjev. Nekateri se zaradi le delne poenotenosti v omrežju **WA** pojavijo dvakrat ali večkrat. Zato smo naredili razbitje množice avtorjev tako, da so v istem razredu imena istega avtorja. Recimo O'Regan Donal je enkrat zapisan kot `oregan.donal` in drugič kot `o'regan.d.` Enotno ime tega avtorja v bazi ZB je oregan.donal, pri nekaterih njegovih delih pa poenoteno ime ni zapisano in v takih primerih je naš program za pretvorbo podatkov ustvaril poenoteno ime iz celega imena avtorja – O'Regan Donal dobi poenoteno ime `o'regan.d.` Še en avtor, ki predstavlja podoben problem, je Pečarić Josip E. Njegovo poenoteno ime v bazi ZB je `pecaric.josip-e`, vendar pri nekaterih njegovih delih poenoteno ime ni zapisano in zato za tista dela dobimo poenotene oblike `pecaric.j` in `pecaric.j-e`, saj obstajata v bazi ZB dva različna zapisa njegovega imena: Pečarić J. in Pečarić J. E. Še en vir problemov z avtorji pa so vzhodnoevropski priimki: Krachkovskij A. P. in Krachkovskii A. P. verjetno predstavljata istega avtorja.

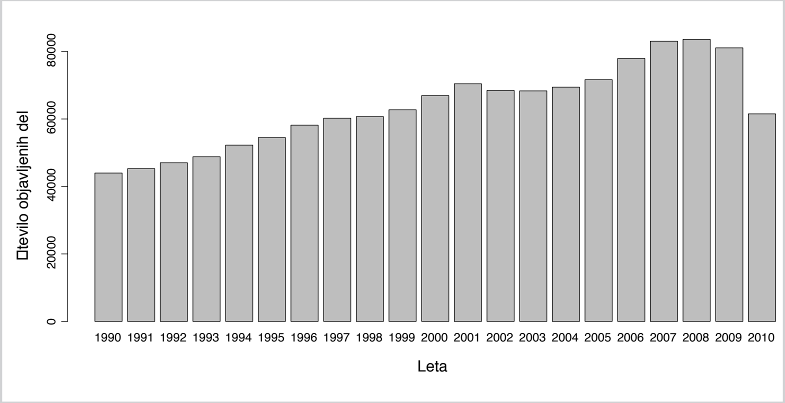
Razbitje enakovrednih imen avtorjev le delno reši problem poenotenja. Za dodatno pomoč smo uporabili tudi identifikacijo avtorjev, ki jo predlaga AMS [81]. Sledeče analize so bile narejene po dodatnemu poenotenju različnih pojavitev imena istega avtorja.

Rešili smo tudi problem z revijami. Različna imena iste revije smo nadomestili z enim imenom – iz 3158 imen revij smo pridobili 2665 poenotenih imen revij.

3.3.2 Porazdelitve lastnosti del

Porazdelitve različnih podatkov pridobimo s pomočjo stopenj vozlišč. Porazdelitev del glede na število avtorjev, ključnih besed in klasifikacij dobimo iz izhodnih stopenj vozlišč v množici del v omrežjih **WA**, **WK** in **WM**. Porazdelitev del glede na število revij nas ne zanima, saj bi naj bilo vsako delo objavljeno v kvečjemu eni reviji. Nobeno izmed del v bazi ZB ni bilo objavljeno v več kot eni reviji. Za 8165 del ni podatka o reviji.

Porazdelitev del, ki so bila objavljena v obdobju 1990-2010 in so shranjena v bazi ZB, glede na leto objave je prikazana na sliki 3.7. Opazimo rast števila shranjenih del – v 20-ih letih se je število shranjenih del skoraj podvojilo. Padec v letih 2009 in 2010 je odraz še ne shranjenih del.



Slika 3.7

Porazdelitev del glede na leto objave.

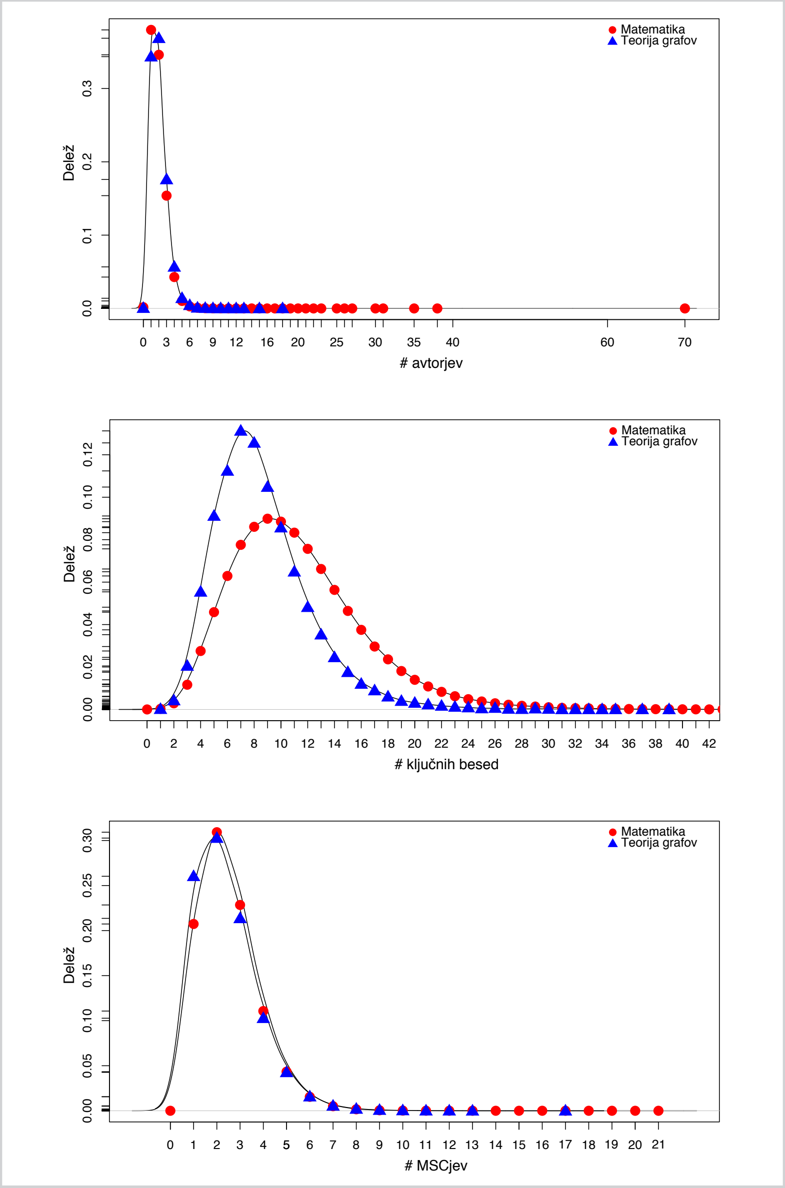
Število avtorjev dela $p \in \mathscr{W}$ v omrežju **WA** je enako njegovi izhodni stopnji $\mathrm{outdeg}(p)$. Porazdelitev del glede na število avtorjev, $\mathrm{F}(d) = |\{p \in \mathscr{W} : \mathrm{outdeg}(p) = d\}|$ = število del, od katerih ima vsako natanko d avtorjev, je prikazana v zgornjem delu slike 3.8. Več kot $\frac{1}{3}$ vseh del (37.99%) ima le enega avtorja in še ena $\frac{1}{3}$ vseh del (34.60%) ima dva avtorja. 2383 del nima podatkov o avtorjih in na drugi strani ima nekaj del veliko avtorjev – celo 70. Dela z največ avtorji so:

- 70 avtorjev – Aderholz, M. in drugi: Distributed applications monitoring at system and network level. Comput. Phys. Commun. 140, Št.1-2, 219-225 (2001).
- 38 avtorjev – Bridle, S. in drugi: Handbook for the GREAT08 challenge: an image analysis competition for cosmological lensing. Ann. Appl. Stat. 3, Št. 1, 6-37 (2009).
- 35 avtorjev – Regan, S.P. in drugi: Direct-drive inertial confinement fusion implosions on omega. Astrophys. Space Sci. 298, Št. 1-2, 227-233 (2005).

Porazdelitev del glede na število ključnih besed je prikazano na sredini slike 3.8. Pri interpretaciji tega grafa moramo upoštevati, da so bile ključne besede pridobljene iz ključnih besed dela in naslova dela. Ta porazdelitev je precej ploska. Približno 50% vseh del ima od 10 do 18 ključnih besed. Dela z največ ključnimi besedami so:

- 71 ključnih besed – Baianu, I.C. in drugi: Algebraic topology foundations of supersymmetry and symmetry breaking in quantum field theory and quantum gravity: a review. SIGMA, Symmetry Integrability Geom. Methods Appl. 5, Članek 051, 70 strani, le elektronsko (2009).
- 69 ključnih besed – Dutta, H.: On some sequence spaces generated by $\Delta(r)$ – and Δr – difference of infinite matrices. Int. J. Open Probl. Comput. Sci. Math., IJOPCM 2, Št. 4, 496-504 (2009).
- 68 ključnih besed – Cheng B. in Tong, H.: On consistent nonparametric order determination and chaos. J. R. Stat. Soc., Ser. B 54, Št.2, 427-449 (1992).

Porazdelitev del glede na število zapisov MSC je prikazana na dnu slike 3.8. Približno ena tretjina del (30.92%) ima določena dva zapisa MSC, približno 40% vseh del (43.57%) pa ima določen eden ali tri zapise MSC. Dela z največ klasifikacijami so:



Slika 3.8

Porazdelitve del glede na število avtorjev, ključnih besed in klasifikacij MSC. V slikah so vključene tudi porazdelitve za dela iz teorije grafov, ki so opisane v razdelku 3.3.4.

- 21 zapisov MSC – Auroux, D. in drugi: Report 35/2006: Four-dimensional Manifolds (August 6th – August 12th, 2006). Oberwolfach Rep. 3, Št. 3, 2059-2140 (2006).

- 21 zapisov MSC – Dechevsky, L.T.: Concluding remarks to paper 'Properties of function spaces generated by the averaged moduli of smoothness'. Int. J. Pure Appl. Math. 49, Št. 1, 147-152 (2008).

- 20 zapisov MSC – Aubin, J.-P.: A survey of viability theory. SIAM J. Control Optimization 28, Št. 4, 749-788 (1990).

Izkaže se, da lahko vse porazdelitve, prikazane na sliki 3.8, zelo dobro aproksimiramo z logaritemsko normalno porazdelitvijo, gamma porazdelitvijo in s posplošeno eksponentno krivuljo z obratno potenco ('reciprocal power exponential curve') $y = c \cdot (x+d)^{\frac{a}{b+x}}$. V vseh primerih dobimo najboljši približek z gamma porazdelitvijo. Rezultati so podani v tabeli 3.5. Tehnične podrobnosti so razložene v razdelku 2.5.2 *Fitting distributions* v [82].

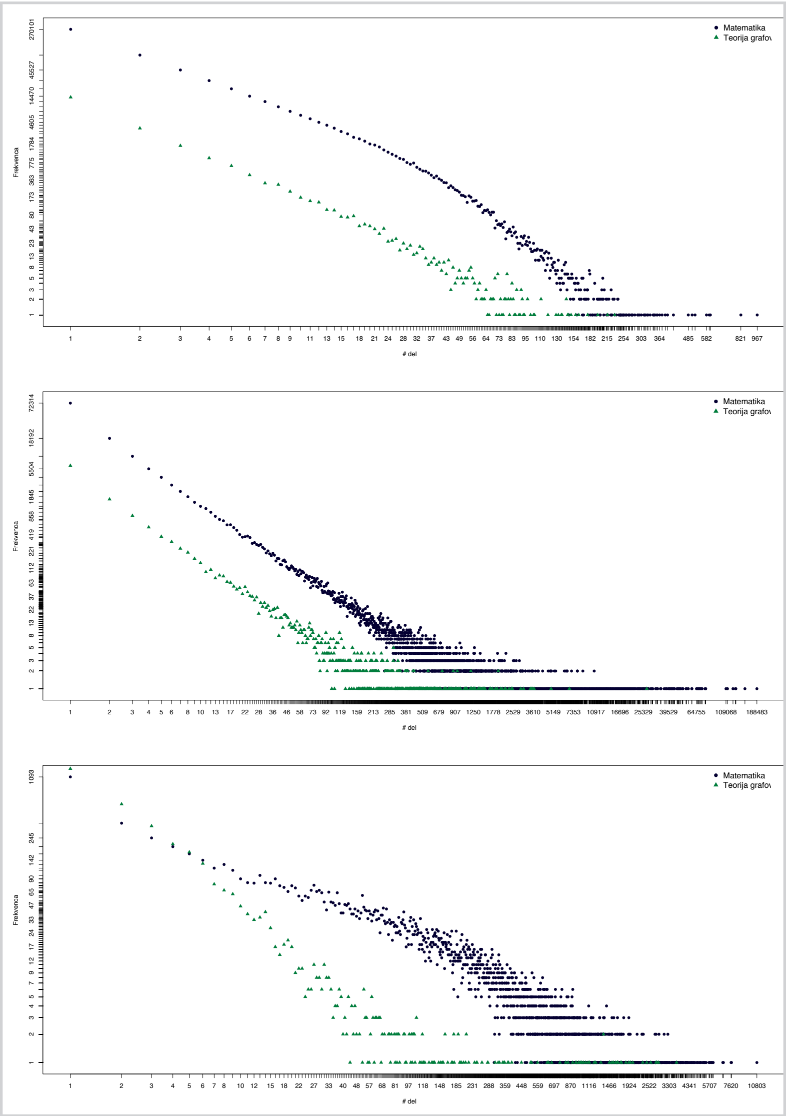
Tabela 3.5

Prileganje gamma porazdelitve $c \cdot \Gamma(x, a, b)$.

Porazdelitev	c	a	b	residual SS
Avtorji (vsi)	$1.342 \cdot 10^6$	3.621	1.905	99893627
Avtorji (teorija grafov)	$4.056 \cdot 10^4$	3.876	1.920	23743
Ključne besede (vsi)	$1.333 \cdot 10^6$	5.561	0.498	10079394
Ključne besede (teorija grafov)	$4.006 \cdot 10^4$	6.960	0.832	188488
zapisi MSC (vsi)	$1.345 \cdot 10^6$	3.718	1.457	273424172
zapisi MSC (teorija grafov)	$4.136 \cdot 10^4$	3.078	1.261	357960

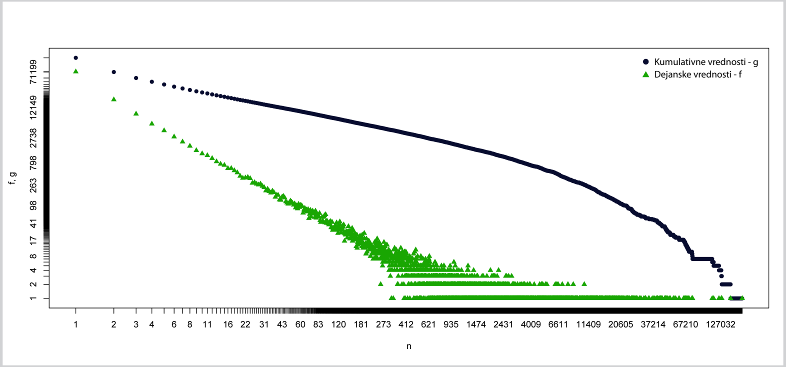
Poleg porazdelitev stopenj vozlišč v prvi množici dvovrstnih omrežij smo si ogledali tudi porazdelitve stopenj vozlišč v drugi množici. Porazdelitev avtorjev glede na število del, pri katerih so sodelovali, je prikazana na vrhu slike 3.9. Recimo točka v zgornjem levem kotu predstavlja 271013 avtorjev, ki so avtorji le enega dela v letih 1990-2010. Točke v spodnjem desnem kotu predstavljajo Ballico Edoarda z 967 soavtorskimi deli v danem časovnem obdobju, O'Regan Donala z 821 deli, Pečarić Josipa s 606 deli, Agarwal Ravija P. s 598 deli in Srivastava H.M. s 582 deli.

Porazdelitev ključnih besed glede na število del, v katerih so bile uporabljene, je prikazana na drugem diagramu na sliki 3.9. Točka v zgornjem levem kotu predstavlja



Slika 3.9

Od vrha proti dnu: Frekvenčna porazdelitev avtorjev glede na število del v dvojni logaritemski lestvici; Frekvenčna porazdelitev ključnih besed glede na število del v dvojni logaritemski lestvici; Frekvenčna porazdelitev zapisov MSC glede na število del v dvojni logaritemski lestvici. V teh diagramih so vključene tudi porazdelitve glede na dela iz teorije grafov, kar je opisano v razdelku 3.3.4.



Slika 3.10

Zaporedji f in g za porazdelitev ključnih besed po številu del.

72314 ključnih besed, ki so bile uporabljene le enkrat v časovnem obdobju 1990-2010. Točke spodaj desno predstavljajo najpogosteje uporabljene ključne besede: equation (uporabljena 188483-krat), problem (152514), function (129957), method (128740), model (123448), space (112000), solution (109068), linear (76241), theory (75873) in finite (75398). Te ključne besede se v matematiki uporabljajo zelo pogosto. Oblika porazdelitve ključnih besed po številu del v drugem diagramu slike 3.9 je značilna za empirične porazdelitve količin, ki sledijo potenčnemu zakonu $f_n = cn^{-\alpha}$. Z uporabo funkcije `power.law.fit` v R paketu `igraph`, s katero je implementiran M. Newmanov postopek, ki je opisan v [83], dobimo $\alpha = 1.85$. Sledenje porazdelitve potenčnemu zakonu lahko preverimo tudi vizualno: uporabimo lastnost, da za $\alpha > 1$ velja, da če zaporedje (f_n) sledi potenčnemu zakonu, potem tudi zaporedje (g_n) sledi potenčnemu zakonu. Pri tem je $g_n = \sum_{i=n}^{\infty} f_i \simeq Cn^{1-\alpha}$, kakor je predstavljeno z enačbo 4.38 v Barabási, A-L.: *Network Science,* 2014[33]. V skupni sliki obeh zaporedij v dvojni logaritemski lestvici bi dobili dve 'daljici', če bi obe zaporedji sledili potenčnemu zakonu. Za zaporedje ključnih besed po številu del to ne velja, kar je razvidno iz slike 3.10. Ta porazdelitev torej ne sledi potenčnemu zakonu.

Porazdelitev zapisov MSC glede na število del, ki so klasificirana z njimi, je prikazana v spodnjem diagramu na sliki 3.9. Vsako delo je klasificirano z enim primarnim zapisom MSC in mogoče s kakšnim sekundarnim zapisom MSC. Enaka primarni in sekundarni zapis MSC (recimo 74S05 in ★74S05) sta predstavljena z eno točko. Vsaka

[33] http://barabasilab.com/networksciencebook

točka predstavlja neko število zapisov MSC, ki je označeno na ordinatni osi, ki so bili uporabljeni za neko število del, ki je označeno na abscisni osi. Točka v zgornjem levem kotu predstavlja 1093 zapisov MSC, ki so bili uporabljeni za klasifikacijo le enega dela v časovnem obdobju 1990-2010. V tabeli 3.6 so našteti najpogosteje uporabljeni zapisi MSC. Ti zapisi MSC so na spodnjem diagramu slike 3.9 prikazani s točkami v spodnjem desnem kotu. Najpogosteje uporabljeni primarni zapisi MSC so našteti v tabeli 3.7.

Tabela 3.6

Tabela najpogosteje uporabljenih zapisov MSC.

Zapis MSC	*Ime dvoznakovnega zapisa MSC*	*Ime zapisa MSC*	*Število del*
80A20	Classical thermodynamics, heat transfer	Heat and mass transfer, heat flow	13279
74S05	Mechanics of deformable solids	Finite element methods	13271
68T05	Computer science	Learning and adaptive systems	11775
35B40	Biology and other natural sciences	Molecular structure	9338
62P10	Operations research, mathematical programming	Combinatorial optimization	8935
35Q53	Partial differential equations	KdV-like equations	8528
91B28	Game theory, economics, social and behavioral sciences	Finance, portfolios, investment	8366
76D05	Fluid mechanics	Navier-Stokes equations	8207
65N30	Numerical analysis	Finite elements, Rayleigh-Ritz and Galerkin methods, finite methods	7976
62M10	Statistics	Time series, auto-correlation, regression, etc.	7926

Tabela 3.7

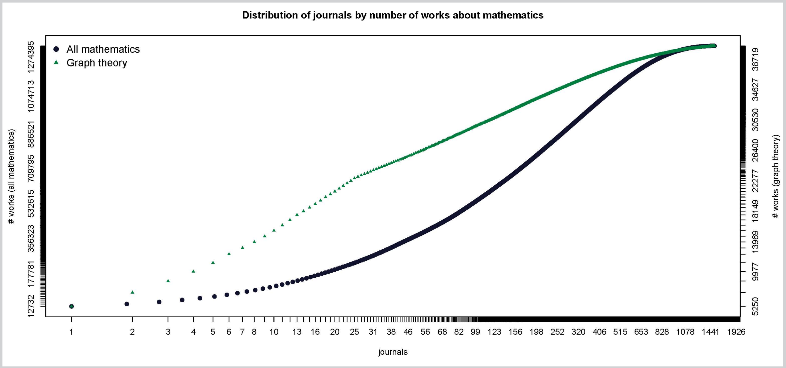
Tabela najpogosteje uporabljenih primarnih zapisov MSC.

Zapis MSC	*Ime dvoznakovnega zapisa MSC*	*Ime zapisa MSC*	*Število del*
74S05	Mechanics of deformable solids	Finite element methods	7620
01A70	History and biography	Biographies, obituaries, personalia, bibliographies	5983
68T05	Computer science	Learning and adaptive systems	5943
90B35	Operations research, mathematical programming	Scheduling theory, deterministic	5707
91B28	Game theory, economics, social and behavioral sciences	Finance, portfolios, investment	5386
62P10	Operations research, mathematical programming	Combinatorial optimization	5316
68U99	Computer science	None of the above, but in this section	5193
62-99	Statistics	Other applications	5073
35Q53	Partial differential equations	KdV-like equations	4792
90B30	Operations research, mathematical programming	Production models	4656

Na sliki 3.11 je prikazano zaporedje revij, v katerih so bila objavljena v bazi ZB shranjena dela v letih 1990-2010, v obliki Bradfordovega grafa. Vrednosti na x-osi so prikazane v logaritemski lestvici. Vsaka točka predstavlja eno revijo. Vrednosti na y-osi so kumulativne vsote teh del v revijah. Revije na skrajni levi strani so revije z največ objavljenimi deli. Te revije so:

- 32132 del – Journal of Physics A: Mathematical and General,
- 25464 del – Journal of Mathematical Analysis and Applications,
- 20564 del – Proceedings of the American Mathematical Society,
- 20322 del – Applied Mathematics and Computation in
- 18110 del – European Journal of Operational Research.

Revije v desnem kotu imajo v danih letih objavljeni le dve deli glede na bazo ZB. Nekatere izmed teh revij so: Journal of the History of Economic Thought, The Montana Mathematics Enthusiast, Journal of Mathematics Education, Vestnik Moskovskogo Universiteta. Seriya VI, International Journal of Energy, Environment and Economics idr.



Slika 3.11

Revije v formi Bradfordovega grafa. Dodana je tudi enaka porazdelitev za revije le iz področja teorije grafov, ki jo obravnavamo v razdelku 3.3.4.

3.3.3 Omrežje sodelovanj

Omrežje sodelovanj nam pomaga pri raziskavah sodelovanj med matematiki. Množica vozlišč v tem omrežju predstavlja množico avtorjev. Avtorja sta povezana, če sta soavtorja vsaj enega skupnega dela.

Omrežje sodelovanj definiramo kot $\mathbf{Co} = \mathbf{AW} \star \mathbf{WA}$ (stran 57). Vrednost oz. utež povezave med dvema avtorjema je enaka številu skupnih del. 20 avtorjev z največ soavtorji je naštetih v tabeli 3.8.

Kakor smo že omenili, lahko posamezno ime pripada več avtorjem. Imena avtorjev smo zato preverili v MathSciNet Authors Search [81]. Imena v tabeli 3.8 smo razdelili v dva stolpca – imena v levem stolpcu predstavljajo le enega avtorja, imena v desnem stolpcu pa lahko predstavljajo več avtorjev. Vrednosti v skrajno desnem stolpcu so števila znanih matematikov s posameznim imenom.

Tabela 3.8

Avtorji z največ soavtorji.

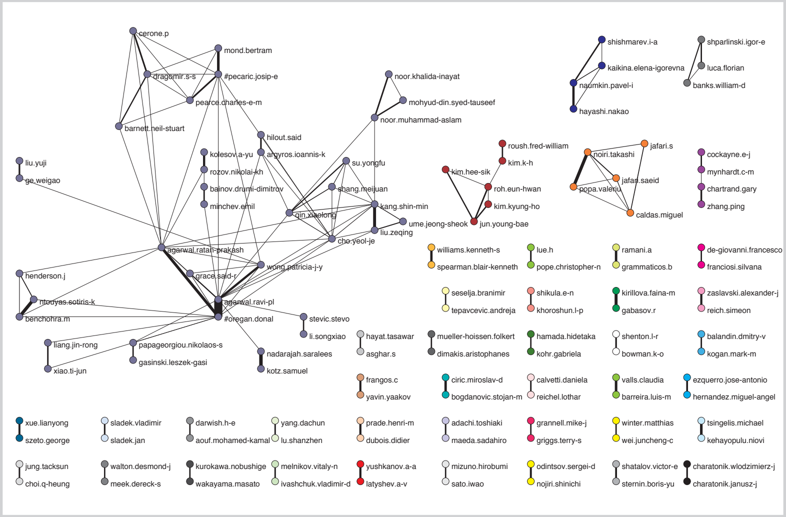
i	*Avtor*	*Število soavtorjev*	*Avtor*	*Število soavtorjev*	*Št. avtorjev s tem imenom v AMS*
1	Srivastava Hari Mohan	347	Wang Wei	463	282
2	Chen Guanrong	341	et al.	316	
3	Alon Noga	288	Zhang Wei	293	228
4	Pardalos Panos M.	212	Li Wei	277	193
5	Il'in V.A.	195	Li Jun	244	157
6			Wang Hui	232	132
7			Wang Yong	224	164
8			Wang Jun	223	166
9			Zhang Li	218	465
10			Li Li	217	324
11			Wang J.	208	1144
12			Li Gang	199	42
13			Zhang Jun	199	130
14			Li Ming	193	133
15			Wang Y.	192	1377

Podmnožico najbolj sodelovalnih avtorjev lahko določimo s p_S-sredicami, ki smo jo že opisali v razdelku 3.2.2.

p_S-sredica reda t je v omrežju sodelovanj takšno podomrežje, v katerem je vsak avtor prispeval vsaj t k skupnim delom z avtorji v tem podomrežju. Večja količina objavljenih del ne predstavlja nujno večje sodelovalnosti avtorja. V izračunu p_S-sredic seštevamo vrednosti povezav. Za nevtralizacijo prevelikega vpliva del z veliko soavtorji v omrežju sodelovanj uporabimo za izračun omrežja sodelovanj normalizirano omrežje soavtorstev $\mathbf{N}$ [35]: $\mathbf{N} = \mathrm{diag}(\frac{1}{\max(1,\deg(w))}) \cdot \mathbf{WA}$. V omrežju $\mathbf{N}$ so vrednosti povezav od dela do njegovih avtorjev enake in njihova vsota je enaka 1. V [35] smo za izračun prispevkov avtorjev njihovim delom uporabili normalizirano omrežje $\mathbf{Ct} = \mathbf{N}^T \star \mathbf{N}$. Za potrebe analiz podatkov ZB smo uporabili nekoliko drugačno normalizirano omrežje sodelovanj $\mathbf{Ct}' = \mathbf{N}^T \star \mathbf{N}'_{\mathbf{WA}}$, pri čemer je $\mathbf{N}'_{\mathbf{WA}} = \mathrm{diag}(\frac{1}{\max(1,\deg(w)-1)}) \cdot \mathbf{WA}$. Ker so vse (usmerjene) povezave v $\mathbf{Ct}'$ dvosmerne in z enako utežjo, jih nadomestimo z neusmerjenimi povezavami z dvojnimi utežmi. Tako nevtraliziramo vpliv del z veliko avtorji: k-klike avtorjev (nekega dela) bi prispevala utež $\frac{k\cdot(k-1)}{2}$, kar nevtraliziramo v

Ct′. Diagonalne vrednosti (uteži zank) nastavimo na nič.

Na sliki 3.12 je prikazana p_S-sredica reda $t = 30$ v normaliziranem omrežju sodelovanj podatkov ZB. V spodnji polovici opazimo predvsem pare avtorjev, ki mnogo bolj sodelujejo v parih kot v večjih skupinah. V večji skupini na levi opazimo močnejše povezave med določenimi avtorji – temnejše in debelejše povezave predstavljajo večji prispevek skupnim delom. Deset najmočnejših parov iz te p_S-sredice je naštetih v tabeli 3.9.



Slika 3.12

p_S-sredica reda $t = 30$ v omrežju sodelovanj **Ct′**.

Produktivnost posameznega avtorja lahko definiramo na več različnih načinov. Rečemo lahko, da je avtor bolj samozadosten, če je njegov prispevek k delom večji. To informacijo lahko pridobimo iz omrežja $\mathbf{Cn} = \mathbf{AW} \star \mathbf{N}$ [35]. Vrednost povezave cn_{ij} v tem omrežju sodelovanj je enaka prispevku avtorja i delom, pri katerih je sodeloval z avtorjem j. Utež zanke cn_{ii} je enaka prispevku avtorja i vsem delom, pri katerih je sodeloval.

Kazalnik samozadostnosti S_i definiramo kot delež avtorjevega prispevka v številu vseh njegovih del. Kazalnik sodelovalnosti K_i definiramo kot komplementarno vrednost kazalniku samozadostnosti $K_i = 1 - S_i$ [35].

'Najboljši' matematiki (najbolj produktivni) so našteti v tabeli 3.10, urejeni so po

Tabela 3.9

Seznam najbolj sodelovalnih parov.

i	*Prvi avtor*	*Drugi avtor*	*Vrednost*
1	Agarwal Ravi P.	O'Regan Donal	49.46
2	Kotz Samuel	Nadarajah Saralees	20.97
3	Ntouyas Sotiris K.	Benchohra Mouffak	19.69
4	Popa Valeriu	Noiri Takashi	19.00
5	Gabasov Rafail	Kirillova Faina Miha⊠lovna	17.88
6	Liu Zeqing	Kang Shin Min	17.72
7	O'Regan Donal	Agarwal Ratan Prakash	16.85
8	Pečarić Josip E.	Mond Bertram	15.57
9	Kehayopulu Niovi	Tsingelis Michael	15.53
10	Barreira Luis M.	Valls Claudia	15.28

vrednosti v drugem stolpcu – vrednosti prispevka cn_{ii}. Skupno število objavljenih del je navedeno v tretjem stolpcu, vrednost kazalnika sodelovalnosti pa je navedena v četrtem stolpcu. Ugotovili smo, da le tri imena v tej tabeli lahko predstavljajo več avtorjev (glede na podatke v AMS Authors Search): Evans D. J., Wang Wei in Zhou Yong.

Matematiki z največ deli niso nujno na vrhu seznama 'najboljših' matematikov. 'Najboljši' matematiki so napisali veliko del in imajo tudi visok prispevek k tem delom. Če je prispevek avtorja podoben številu njegovih del, ta avtor veliko dela sam oziroma v manjših skupinah. Avtorja z največ deli sta tudi 'najboljša' matematika – Edoardo Ballico and Donal O'Regan. Med njima pa je velika razlika – Edoardo Ballico več dela sam, Donald O'Regan pa v skupinah. Prispevek Ballica je skoraj enak številu njegovih del (89.51%), prispevek O'Regana pa je enak polovici števila njegovih del (52.72%). Po skupnem številu del jima sledijo Josip E. Pečarić, Srivastava Mohan Hari in Weigao Ge. Najbolj sodelovalni avtorji pa so Guanrong Chen, Ravi P. Agarwal, Lansun Chen, Jaume Llibre in Josip E. Pečarić, saj imajo največje vrednosti kazalnika sodelovalnosti. V seznamu 'najboljših' matematikov najdemo tudi bivšega urednika kontroverzne[34] revije Chaos, Solitons & Fractals El Naschie Mohameda Saladina.

[34]http://poynder.blogspot.com/2010/05/elsevier-chaos-solitons-fractals-risen.html

Tabela 3.10

Seznam 'najbojših' matematikov v letih 1990-2010.

i	*Avtor*	$\mathbf{cn_{ii}}$	*Št. del*	$\mathbf{K_i}$
1	Ballico Edoardo	865.58	967	0.105
2	O'Regan Donal	432.80	821	0.470
3	Argyros Ioannis Konstantinos	353.17	373	0.053
4	Shelah Saharon	302.49	519	0.417
5	Verma Ram U.	297.58	314	0.052
6	Srivastava Hari Mohan	267.17	582	0.541
7	Pečarić Josip E.	265.28	608	0.564
8	Papageorgiou Nikolaos S.	263.92	418	0.369
9	Pachpatte Baburao G.	255.00	261	0.023
10	Maslov Victor P.	247.08	324	0.237
11	Agarwal Ravi P.	244.12	598	0.592
12	Wazwaz Abdul-Majid	242.67	254	0.045
13	Noor Muhammad Aslam	241.00	351	0.313
14	Jun Young Bae	239.40	480	0.501
15	Dragomir S. S.	232.73	364	0.361
16	Le Maohua	226.17	242	0.065
17	Ge Weigao	225.78	504	0.552
18	Nadarajah Saralees	213.70	315	0.322
19	Ramm Alexander G.	211.92	270	0.215
20	Stević Stevo	206.20	257	0.198
21	Gamkrelidze R.V.	190.65	268	0.289
22	Zaslavski Alexander J.	187.50	236	0.206
23	El Naschie Mohamed Saladin	183.08	186	0.016
24	Evans D. J.	182.88	356	0.486
25	Wang Wei	178.79	394	0.546
26	Nazarov Serguei A.	176.95	266	0.335
27	Chen Huanyin	175.50	206	0.148
28	Alzer Horst	171.92	198	0.132
29	Luca Florian	167.62	292	0.426
30	Danchev Peter Vassilev	166.00	170	0.024
31	Guo Boling	165.08	341	0.516
32	Nishimoto Katsuyuki	163.42	213	0.233
33	Chajda Ivan	162.00	244	0.336
34	Shparlinski Igor E.	161.50	292	0.447
35	Owa Shigeyoshi	161.06	331	0.513
36	Anastassiou George A.	157.08	200	0.215
37	Noiri Takashi	152.25	321	0.526
38	Ikramov Kh.D.	151.08	198	0.237
39	Jakubík Ján	150.83	158	0.045
40	Zhou Yong	150.03	245	0.388
41	Chen Guanrong	149.54	385	0.612
42	Biswas Indranil	148.83	224	0.336
43	Llibre Jaume	148.77	345	0.569
44	Khrennikov Andrei Yu.	148.52	192	0.226
45	Hall Peter G.	145.45	294	0.505
46	Chen Lansun	141.03	344	0.590
47	Aouf Mohamed Kamal	139.99	251	0.442
48	Chen Bang-Yen	138.00	177	0.220
49	Park Sehie	137.25	166	0.173
50	Alon Noga	136.37	298	0.542

3.3.4 05Cxx Teorija grafov

Teorija grafov je poddisciplina kombinatorike z oznako 05 in njegov triznakovni zapisa MSC je 05C. Na teorijo grafov lahko gledamo kot takšno ali pa kot teorijo grafov z uporabo – vključimo zapise MSC iz drugih matematičnih disciplin, katerih vsebina je povezana s teorijo grafov.

V nadaljnih analizah uporabljamo omrežje $\mathbf{WM}_3$, ki je skrčena verzija omrežja $\mathbf{WM}$: množica petznakovnih zapisov MSC je skrčena v triznakovne zapise MSC. Kombinacija tega omrežja z ostalimi omrežji omogoča vpogled v teorijo grafov skozi objavljena dela.

Za identifikacijo revij, ki so objavile največ v bazi ZB shranjenih del o teoriji grafov, potrebujemo omrežji $\mathbf{WJ}$ in $\mathbf{WM}_3$. Vrednosti povezav v drugem omrežju so lahko večje od ena – delo je lahko klasificirano z več kot enim zapisom MSC z istimi prvimi tremi znaki. Te vrednosti pretvorimo v enice in zmnožimo omrežji $\mathbf{WJ}$ in $\mathbf{WM}_3$, s čimer dobimo omrežje $\mathbf{JM}_3 = \mathbf{JW} \star b(\mathbf{WM}_3)$, kjer je $b(\mathcal{N})$ binarna verzija omrežja $\mathcal{N}$. Vrednost povezave $jm3_{ic}$ v tem omrežju je enaka številu shranjenih del, ki so bila objavljena v reviji i in so klasificirana z zapisom MSC c. To omrežje smo normalizirali podobno kot $\mathbf{WA}$, da smo dobili normalizirano omrežje sodelovanj: $n(\mathbf{JM}_3) = \mathrm{diag}\frac{1}{\text{utežena stopnja}(j)}\mathbf{JM}_3$. Utežena stopnja vozlišča je enaka vsoti uteži incidenčnih povezav. Vsota uteži incidenčnih povezav je za vsako revijo v omrežju $n(\mathbf{JM}_3)$ enaka 1.

Ogledali smo si vrednosti povezav med revijami in klasifikacijami teorije grafov. Vrednost povezave predstavlja delež v bazi ZB shranjenih del iz izbranih revij, ki so klasificirana z zapisom MSC o teoriji grafov. V levem stolpcu tabele 3.11 so naštete revije z najvišjimi deleži takšnih del. V desnem stolpcu tabele 3.11 pa so naštete revije, ki imajo najvišje deleže del o teoriji grafov in njeni uporabi. Zapisi MSC, ki predstavljajo uporabo teorije grafov v drugih disciplinah matematike, so 68R10, 81Q30, 81T15, 82B20, 82C20, 90C35, 92E10, 94C15, 05E30, 57M15, 57M25, 20F65, 90B10, 05B30, 05D10, 91A43, 91A46, 60B20, 91D30, 68R10, 68W05, 81Q30, 81T15, 82B20, 82C20, 90C35, 92E10, 94C15 in vsi zapisi MSC, ki se začnejo z 90B.

Razlike med obema seznamoma so zlahka opazne. V eni reviji (The European Physical Journal B. Condensed Matter) so imela vsa objavljena dela, ki so shranjena v bazi ZB, vsaj eno klasifikacijo o teoriji grafov ali njenih uporab.

Identifikacijo revij z največ objavljenimi deli o teoriji grafov lahko opravimo tudi

Tabela 3.11

Revije z najvišjimi deleži shranjenih del, ki so klasificirana s teorijo grafov v letih 1990-2010: samo teorija grafov levo in teorija grafov z uporabo desno.

Teorija grafov		*Teorija grafov z uporabo*	
Journal of Graph Theory (0364-9024, 1097-0118)	89.15%	The European Physical Journal B. Condensed Matter (1434-6028)	100.00%
AKCE International Journal of Graphs and Combinatorics (0972-8600)	82.55%	Journal of Graph Theory (0364-9024, 1097-0118)	95.87%
Journal of Combinatorial Theory. Series B (0095-8956)	68.85%	AKCE International Journal of Graphs and Combinatorics (0972-8600)	88.89%
Graphs and Combinatorics (0911-0119, 1435-5914)	59.45%	Journal of Combinatorial Theory. Series B (0095-8956)	83.20%
Ars Combinatoria (0381-7032)	50.04%	ITS Journal (1024-8072)	80.77%
The Australasian Journal of Combinatorics (1034-4942)	46.96%	International Journal of Flexible Manufacturing Systems (0920-6299, 1572-9370)	79.44%
JCMCC. The Journal of Combinatorial Mathematics and Combinatorial Computing (0835-3026)	43.81%	Graphs and Combinatorics (0911-0119, 1435-5914)	76.49%
Ars Mathematica Contemporanea (1855-3966, 1855-3974)	42.86%	International Journal of Production Research (0020-7543, 1366-588X)	76.48%
Congressus Numerantium (0384-9864)	42.81%	Match (0340-6253)	75.05%
Match (0340-6253)	42.80%	Journal of Graph Algorithms and Applications (1526-1719)	74.72%
Discrete Mathematics (0012-365X)	42.39%	Location Science (0966-8349)	74.29%
Bulletin of the Institute of Combinatorics and its Applications (1183-1278)	42.28%	Journal of Scheduling (1094-6136, 1099-1425)	68.84%
Advances and Applications in Discrete Mathematics (0974-1658)	42.27%	Journal of Interconnection Networks (0219-2659)	66.67%
Combinatorica (0209-9683)	33.99%	Studies in Locational Analysis (1105-5162)	66.48%
Combinatorics, Probability and Computing (0963-5483, 1469-2163)	33.76%	Networks (0028-3045, 1097-0037)	66.04%
College Mathematics Journal (0746-8342)	33.33%	Transportation Science (0041-1655)	63.77%
International Journal of Mathematical Combinatorics (1937-1055)	29.29%	Networks and Spatial Economics (1566-113X, 1572-9427)	62.00%
Random Structures & Algorithms (1042-9832, 1098-2418)	29.06%	The Australasian Journal of Combinatorics (1034-4942)	60.31%
Discussiones Mathematicae. Graph Theory (1234-3099)	28.95%	Ars Combinatoria (0381-7032)	59.96%
Journal of Combinatorics, Information & System Sciences (0250-9628)	28.24%	JCMCC. The Journal of Combinatorial Mathematics and Combinatorial Computing (0835-3026)	59.87%

s pomočjo *mere pristranskosti* ali biasov [84]. Pristranskost revije za ali proti neki veji matematike definiramo kot

$$\text{bias} = \log_2 \frac{\text{delež del o tej veji matematike v reviji}}{\text{delež del o tej veji v vseh revijah skupaj}}.$$

Ta vrednost nam pove, ali revija favorizira izbrano vejo matematike (pozitivna vrednost) ali ji ni naklonjena (negativna vrednost). Če je vrednost pristranskosti enaka nič, je v izbrani reviji objavljen tolikšen delež del o izbrani veji matematike, kolikšen je v vseh revijah skupaj.

V tabeli 3.12 so naštete revije z največjimi pozitivnimi vrednostmi pristranskosti za teorijo grafov, v tabeli 3.13 pa so naštete revije z najbolj negativnimi vrednostmi pristranskosti o teoriji grafov. V izračun pristranskosti smo vključili le revije, za katere je v bazi ZB shranjenih vsaj 50 del. Avtor lahko uporabi pristranskost o temi svojega dela kot mero primernosti izbire revije za objavo tega dela. Pozitivna vrednost pristranskosti pomeni, da je bolj verjetno, da bo izbrana revija to delo objavila, negativna vrednost pristranskosti pa, da je bolj verjetno, da bo izbrana revija to delo zavrnila.

Tabela 3.12

Revije z najbolj pozitivnimi vrednostmi pristranskosti za teorijo grafov v letih 1990-2010.

Revija	*Pristranskost*
Journal of Graph Theory (0364-9024)	6.035
Discussiones Mathematicae. Graph Theory (1234-3099)	6.023
AKCE International Journal of Graphs and Combinatorics (0972-8600)	5.928
Journal of Combinatorial Theory. Series B (0095-8956)	5.774
Graphs and Combinatorics (0911-0119)	5.618
Applicable Analysis and Discrete Mathematics (1452-8630)	5.604
Ars Combinatoria. The Canadian Journal of Combinatorics (0381-7032)	5.297
The Australasian Journal of Combinatorics (1034-4942)	5.265
JCMCC. The Journal of Combinatorial Mathematics and Combinatorial Computing (1983-0823)	5.238
MATCH - Communications in Mathematical and in Computer Chemistry (0340-6253)	5.233

Tabela 3.13

Revije z največjimi negativnimi vrednostmi pristranskosti za teorijo grafov med leti 1990-2010.

Revija	*Pristranskost*
International Journal of Solids and Structures (0020-7683)	−7.765
Journal of Differential Equations (0022-0396)	−7.427
International Journal of Modern Physics A. Particles and Fields, Gravitation and Cosmology (0217-751X)	−7.127
Classical and Quantum Gravity. An International Journal of Gravitational Physics, Cosmology, Geometry and Field Theory (0264-9381)	−6.982
Modern Physics Letters A. Particles and Fields, Gravitation, Cosmology, Nuclear Physics (0217-7323)	−6.797
Systems & Control Letters (0167-6911)	−6.516
Applicable Analysis. An International Journal (0003-6811)	−6.472
Acta Arithmetica (0065-1036)	−6.357
Nonlinear Analysis. Theory, Methods & Applications. Series A: Theory and Methods. An International Multidisciplinary Journal (0362-546X)	−6.124
Annals of Physics (0003-4916)	−6.070

Za nadaljne analize smo uporabili omrežje $\mathbf{WM}^{[05C]}$, ki je dejansko omrežje **WM** z drugo množico vozlišč omejeno tako, da so v njej le zapisi MSC o teoriji grafov. Omrežje $\mathbf{WM}^{[05C]}$ dobimo tako, da najprej naredimo razbitje klasifikacij σ, v kateri so vse klasifikacije 05C v enem razredu, ostale klasifikacije pa v drugem razredu. Z razbitjem σ dobimo podomrežje $\mathbf{WM}_\sigma$ omrežja **WM**. Omrežje $\mathbf{WM}_\sigma$ vsebuje kot vozlišča vsa dela in vse klasifikacije 05C. Nato smo določili razbitje izhodnih stopenj del τ in odstranili iz omrežja $\mathbf{WM}_\sigma$ vsa vozlišča (dela) izhodne stopnje nič. Dela z višjimi stopnjami so klasificirana z vsaj enim zapisom MSC o teoriji grafov. Dobljeno omrežje je $\mathbf{WM}^{[05C]}$.

Tudi na omrežjih **WA**, **WJ** in **WK** smo uporabili razbitje τ in tako pridobili omrežja $\mathbf{WA}^{[05C]}$, $\mathbf{WJ}^{[05C]}$ in $\mathbf{WK}^{[05C]}$, v katerih so vozlišča le dela, ki so klasificirana z vsaj enim zapisom MSC o teoriji grafov, in njihovi avtorji, revije in ključne besede.

Z analizo stopenj vozlišč v pridobljenih omrežjih smo dobili porazdelitve različnih podatkov, kakor smo to že naredili na celotni množici del v razdelku 3.3.2.

Porazdelitev del o teoriji grafov glede na število avtorjev je prikazana v zgornjem diagramu slike 3.8 v razdelku 3.3.2. Porazdelitev je prikazana v istem diagramu kot porazdelitev vseh analiziranih del glede na število avtorjev. Porazdelitvi sta si podobni. Več kot $\frac{1}{3}$ vseh del o teoriji grafov (34.34%) ima samo enega avtorja, še več (36.86%) pa jih ima le dva avtorja. 19 del nima shranjenega podatka o avtorjih.

Porazdelitev del o teoriji grafov glede na število ključnih besed je prikazana v srednjem diagramu slike 3.8. Ta porazdelitev ima višji vrh pri nižji vrednosti (7) v primerjavi s porazdelitvijo vseh del. Približno 65% del o teoriji grafov ima med 5 in 10 ključnih besed.

Porazdelitev del o teoriji grafov glede na število zapisov MSC je prikazana v spodnjem diagramu slike 3.8. Tudi ta porazdelitev je zelo podobna tovrstni porazdelitvi vseh analiziranih del. Približno ena tretjina (30.27%) del o teoriji grafov ima določena dva zapisa MSC in 47.38% del ima določenega enega ali tri zapise MSC.

Porazdelitev avtorjev glede na število del o teoriji grafov je prikazana v zgornjem diagramu slike 3.9 v razdelku 3.3.2 skupaj s tovrstno porazdelitvijo avtorjev glede na število vseh del. Svetlejša točka zgoraj levo predstavlja 13801 avtorjev, ki so v letih 1990-2010 objavili le en članek. Obe porazdelitvi sta si podobni.

Porazdelitev ključnih besed glede na število del o teoriji grafov je prikazana na srednjem diagramu slike 3.9 skupaj s porazdelitvijo ključnih besed glede na število vseh del. Svetlejša točka zgoraj levo predstavlja 6231 ključnih besed, ki so bile v letih 1990-2010 uporabljene v opisu le enega dela.

Porazdelitev zapisov MSC glede na število del o teoriji grafov je prikazana v spodnjem diagramu slike 3.9 skupaj s porazdelitvijo zapisov MSC glede na število vseh analiziranih del. Svetlejša točka zgoraj levo predstavlja 1336 zapisov MSC, ki so klasificirali le po eno delo o teoriji grafov v letih 1990-2010. V primerjavi s porazdelitvijo zapisov MSC glede na število vseh del ima ta porazdelitev višjo vrednost na začetku (pri vrednosti 1) in nato pada hitreje.

Revije, urejene v padajočem zaporedju glede na število objavljenih del o teoriji grafov v letih 1990-2010, so prikazane na sliki 3.11 skupaj z zaporedjem revij, urejenih v padajočem zaporedju vseh objavljenih del v letih 1990-2010. Vsaka točka predstavlja eno revijo. Revije na levi strani so objavile veliko število del. Forma Bradfordovega grafa tega zaporedja revij se ujema z formo Bradfordovega grafa zaporedja revij z objavljenimi vsemi deli le na desnem koncu.

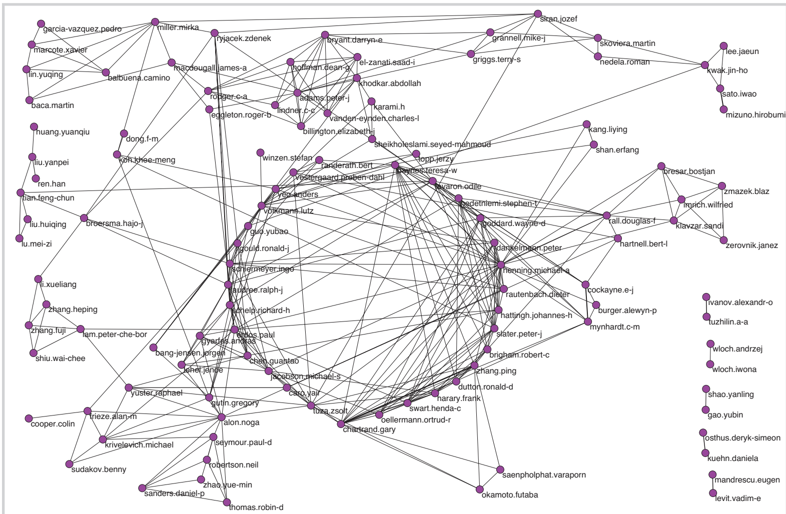
Razbitje del τ smo uporabili na omrežju $\mathbf{WA}$ in tako dobili omrežje $\mathbf{WA}^{[05C]}$, v katerem so vozlišča le dela o teoriji grafov in njihovi avtorji. S pomočjo razbitja vhodnih stopenj druge množice vozlišč v omrežju $\mathbf{WA}^{[05C]}$ dobimo seznam avtorjev z največjimi števili del o teoriji grafov. Te vrednosti dobimo tudi, če pogledamo vrednosti zank v normaliziranem omrežju sodelovanj (tabela 3.14). Preverili smo enoličnost imen v AMS Authors Search in našli le dve imeni, ki lahko predstavljata več avtorjev: Liu Guizhen (dva avtorja) in Zhang Ping (58 avtorjev).

Tabela 3.14

Seznam 20 avtorjev z največjimi prispevki k svojim delom iz teorije grafov v letih 1990-2010.

i	*Avtor*	$\mathbf{cn}_{ii}$	*Skupaj*	$\mathbf{K}_i$
1	Volkmann Lutz	123.55	216	0.428
2	Henning Michael A.	110.87	232	0.522
3	Liu Yanpei	102.42	196	0.478
4	Alon Noga	85.39	177	0.518
5	Tuza Zsolt	77.05	150	0.486
6	Zhu Xuding	76.85	132	0.418
7	Gutman Ivan	70.13	143	0.510
8	Thomassen Carsten	68.83	82	0.161
9	Mohar Bojan	67.85	111	0.389
10	Liu Guizhen	67.28	137	0.509
11	Liu Bolian	63.67	119	0.465
12	Klavžar Sandi	62.74	129	0.514
13	Bollobás Béla	62.53	143	0.563
14	Zhang Ping	60.47	157	0.615
15	Li Xueliang	60.40	136	0.556
16	Rödl Vojtěch	57.98	146	0.603
17	Zhang Zhongfu	57.07	162	0.648
18	McKee Terry A.	54.98	64	0.141
19	Zelinka Bohdan	54.50	57	0.044
20	Yuster Raphael	52.56	79	0.335

Za grafovske teoretike poiščemo močna sodelovanja v omrežju sodelovanj tako, kot smo to naredili za matematike. Normalizirano omrežje sodelovanj $\mathbf{Ct}^{[05C]}$ za teorijo grafov izračunamo z normaliziranim $\mathbf{WA}^{[05C]}$. p_S-sredica reda $t = 3.5$ v tem omrežju je prikazana na sliki 3.13. Sestavljena je iz nekaj parov avtorjev in ene večje skupine. Opazimo močnejše podskupine (debelejše in temnejše povezave) te skupine, ki so med sabo povezane s šibkejšimi povezavami (tanjše in svetlejše povezave).



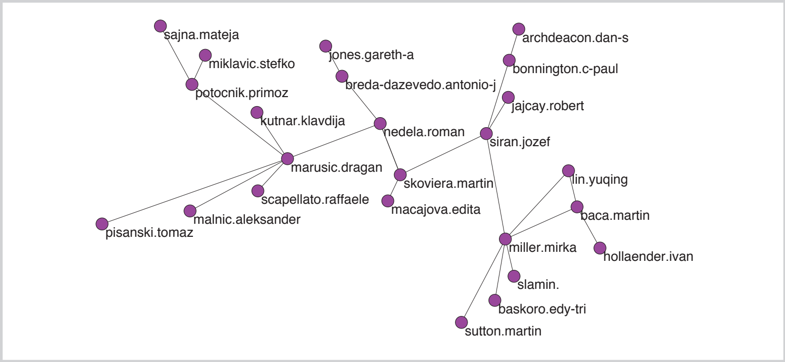
Slika 3.13

p_S-sredica reda $t = 3.50$ v omrežju sodelovanj **Ct** grafovskih teoretikov.

Drugi način za identifikacijo močnih skupin sodelovanj med grafovskimi teoretiki je uporaba povezavnih otokov [71].

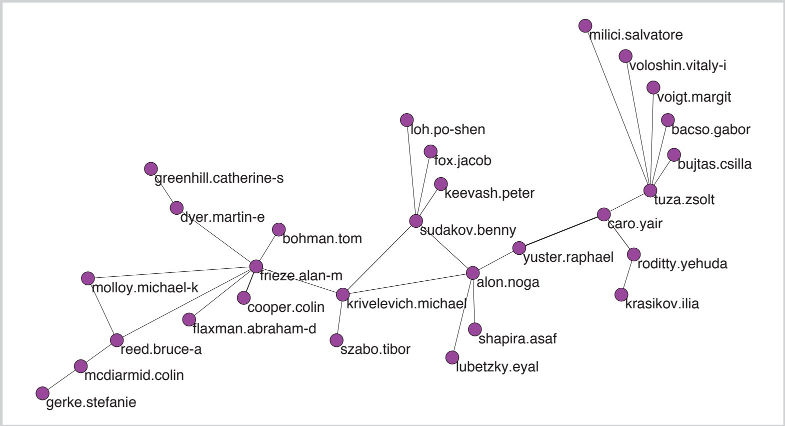
Povezavni otok v omrežju $\mathcal{N} = (\mathcal{V}, \mathcal{E}, w)$ je takšno podomrežje $\mathcal{M} = (\mathcal{U}, \mathcal{F}, w)$, da v njem obstaja vpeto drevo $\mathcal{T}$, da so vrednosti povezav z natanko enim krajiščem v $\mathcal{U}$ manjše ali enake vrednosti povezav v vpetem drevesu $\mathcal{T}$. Povezavni otoki določajo lokalno pomembna podomrežja. Na slikah 3.14, 3.15 in 3.16 so prikazani trije povezavni otoki grafovskih teoretikov v normaliziranem omrežju sodelovanj **Ct**, katerih velikosti so v intervalu med 10 in 30.

Veliko del, klasificiranih z zapisi MSC teorije grafov, je klasificiranih tudi z drugimi zapisi MSC. Več različnih klasifikacij dela pomeni večjo interdisciplinarnost vsebine dela. Omrežje $\mathbf{WM}^{[05C]}$ smo uporabili, da smo dobili seznam klasifikacij, ki se najpo-



Slika 3.14

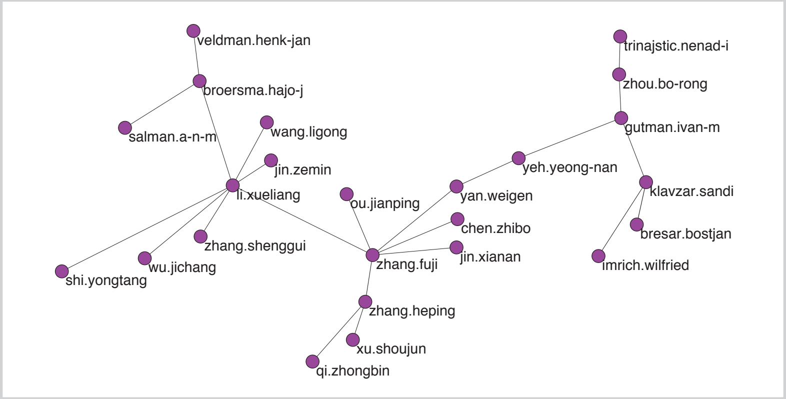
Povezavni otok grafovskih teoretikov v normaliziranem omrežju sodelovanj **Ct** s podskupino slovenskih in slovaških matematikov.



Slika 3.15

Povezavni otok grafovskih teoretikov v normaliziranem omrežju sodelovanj **Ct** z Noga Alon v sredini.

gosteje pojavljajo skupaj s klasifikacijami o teoriji grafov. Za pridobitev te informacije smo v omrežju **WM** uporabili razbitje del τ in tako dobili podomrežje del o teoriji grafov in vse klasifikacije $\mathbf{W}_\tau\mathbf{M}$. Množico del smo skrčili v eno vozlišče. Klasifikacije z največjimi uteženimi vhodnimi stopnjami predstavljajo področja matematike, ki se najpogosteje prepletajo s področjem teorije grafov. Te klasifikacije so naštete v tabeli 3.15. Vsaka izmed teh klasifikacij je definirana s kodo MSC v prvem stolpcu, imenom v tretjem stolpcu in dvoznakovnim zapisom MSC v drugem stolpcu. Klasifikacije so urejene glede na vrednosti v zadnjem stolpcu – po številu del, v katerih se prepletajo s teorijo grafov.



Slika 3.16

Povezavni otok grafovskih teoretikov v normaliziranem omrežju sodelovanj **Ct** z večinoma azijskimi imeni.

Tabela 3.15

Seznam zapisov MSC, ki so se v letih 1990-2010 največkrat sopojavili s teorijo grafov.

Zapis MSC	*Ime dvoznakovnega zapisa MSC*	*Podpodročje zapisa MSC*	*Št. del*
68R10	Computer science	Graph Theory	3528
68Q25	Computer science	Analysis of algorithms and problem complexity	1140
90C35	Operations research, mathematical programming	Programming involving graphs or networks	868
92E10	Biology and other natural sciences	Molecular structure	591
90C27	Operations research, mathematical programming	Combinatorial optimization	572
60C05	Probability theory and stochastic processes	Combinatorial probability	521
05A15	Combinatorics	Exact enumeration problems, generating functions	492
15A18	Linear and multilinear algebra; matrix theory	Eigenvalues, singular values, and eigenvectors	392
57M15	Manifolds and cell complexes	Relations with graph theory	381
05B35	Combinatorics	Matroids, geometric lattices	334
94C15	Information and communication, circuits	Applications of graph theory	317
68W25	Computer science	Approximation algorithms	315
05E30	Combinatorics	Association schemes, strongly regular graphs	312
06A07	Order, lattices, ordered algebraic structures	Combinatorics of partially ordered sets	291
90B10	Operations research, mathematical programming	Network models, deterministic	283
20B25	Group theory and generalizations	Finite automorphism groups of algebraic, geometric, or combinatorial structures	280
20D60	Group theory and generalizations	Arithmetic and combinatorial problems	275
68M10	Computer science	Network design and communication	274
91A43	Game theory, economics, social and behavioral sciences	Games involving graphs	237
05B20	Combinatorics	Matrices	222

Zapisi za veliko večino del vsebujejo informacije o ključnih besedah. Nekatere ključne besede so pogoste v vseh področjih matematike, spet druge pa so specifične za samo določena področja. S pravo utežitvijo ključnih besed lahko določimo njihovo pomembnost za posamezna področja matematike. Uporabili smo utežitev TF-IDF [85].

Področja matematike lahko predstavimo z zapisi MSC. Z množenjem omrežij **MW** in **WK** smo dobili omrežje **MK**. Zapise MSC smo skrčili v manjše število vozlišč glede na njihove troznakovne kode MSC.

Ključne besede, ki se pojavljajo v vseh področjih matematike, dobijo z utežitvijo TF-IDF vrednost nič. Ostale dobijo vrednost:

$$\text{TF-IDF}(\textit{ključna beseda, MSC}) = \text{TF}(\textit{ključna beseda, MSC}) \times \text{IDF}(\textit{ključna beseda})$$

$$\text{TF}(\textit{ključna beseda, MSC}) = \frac{\text{Utež povezave od } \textit{ključne besede do zapisa MSC}}{\text{Vsota vseh povezav do zapisa } \textit{MSC}}$$

$$\text{IDF}(\textit{ključna beseda}) = \log \frac{\text{Št. zapisov MSC}}{\text{Št. zapisov MSC, ki so povezani s } \textit{ključno besedo}}$$

Ključnih besed, ki so z zapisi MSC teorije grafov povezane z neničelno utežitvijo TF-IDF, je 12460. V tabeli 3.16 so naštete tiste z največjo utežitvijo. V tej tabeli so dodana tudi števila pojavitev posamezne ključne besede v delih o teoriji grafov in število pojavitev v vseh delih skupaj.

Tabela 3.16

Seznam ključnih besed z največjimi utežitvami TF-IDF za triznakovni zapis MSC 05C v letih 1990-2010.

Ključna beseda	*Št. pojavitev znotraj teorije grafov*	*Št. vseh pojavitev*	*Utežitev TF-IDF* $(\cdot 10^{-4})$
Coloring	4133	6676	45.24
Digraph	4049	5695	44.32
Chromatic	3958	5138	43.32
Subgraph	2298	3473	35.31
Domination	2500	3883	27.36
Clique	1788	3169	25.29
Vertex	4611	12981	25.23
Hypergraph	2258	3808	24.71
Bipartite	2776	5045	24.08
Tournament	1335	2341	21.92
Matching	1038	1891	17.04
Label	1960	4656	17.00
Ramsey	1305	3148	16.58
Claw	625	715	15.02
Colour	1278	2385	13.99
Girth	736	954	13.93
Connectivity	2528	5462	13.83
Hamiltonicity	465	551	11.96
Match	2159	15140	11.82
Chordal	738	1384	11.34

3.3.5 Zaključek

Bibliografske podatke lahko analiziramo na več različnih načinov. V tem podpoglavju smo predstavili nekaj pristopov z analizo omrežij, ki smo jih uporabili na podatkih iz podatkovne baze Zentralblatt MATH, kjer so shranjene informacije o matematičnih publikacijah. S pomočjo naših analiz podatkov o publikacijah iz obdobja 1990-2010 smo ugotovili, da matematiki raje delajo sami ali v manjših skupinah. Večinoma se tudi omejujejo na specifična področja matematike. To sklepamo iz majhnega števila zapisov MSC, ki klasificirajo dela, in iz majhnega števila ključnih besed za posamezno delo.

Ker so vhodni podatki v podatkovni bazi le delno poenoteni, se pojavljajo problemi s podatki. Ti problemi lahko povzročijo nepravilnosti rezultatov. Nekatere izmed problemov smo vsaj delno rešili v namene naših analiz (poenotenje imen avtorjev in imen revij).

Podrobneje smo si ogledali podatke o delih o teoriji grafov in določili revije, ki so bolj naklonjene temu področju matematike, najboljše grafovske teoretike glede na njihove prispevke k delom, področja matematike, ki se močneje prepletajo s teorijo grafov skozi dela, in ključne besede, ki so značilne za teorijo grafov.

Množenje usklajenih dvovrstnih omrežij omogoča izračun različnih izpeljanih omrežij. V omrežju $\mathbf{AJ} = \mathbf{AW} \star \mathbf{WJ}$ so shranjene informacije o številu shranjenih del, katerih soavtor je izbrani avtor in so bila objavljena v izbrani reviji. To omrežje lahko posebej analiziramo ali pa iz njega izpeljemo še kakšno omrežje. Tako ga lahko pomnožimo z njegovim transponiranim omrežjem in dobimo omrežje $\mathbf{JJ} = b(\mathbf{JA} \star \mathbf{AJ})$. Dve reviji sta v tem omrežju povezani, če obstaja avtor, ki je objavil kakšno delo v vsaki izmed obeh revij. Druga možnost pa je uporaba binarnih omrežij: $\mathbf{JJ}_{\mathbf{A}} = b(\mathbf{JA}) \star b(\mathbf{AJ})$. V dobljenem omrežju je utež povezave med dvema revijama enaka številu avtorjev, ki so objavili vsaj kakšno delo v vsaki izmed teh dveh revij. S pristopom, ki smo ga predstavili v tem podpoglavju, lahko analiziramo podobnosti med shranjenimi revijami.

To je le en primer, kaj vse bi še lahko analizirali s podatki iz baze ZB. V naših analizah nismo upoštevali informacije o letu objave. S tem lahko analize še dodatno razširimo na časovna oziroma dinamična omrežja.

4

Posplošene dvovrstne sredice

Analiza omrežij je eden izmed pristopov k analizi relacijskih podatkov. V tem poglavju se ukvarjamo z analizo dvovrstnih omrežij [86, 87].

Klasičen pristop k analizi dvovrstnih omrežij je sestavljen iz dveh korakov: pretvorbe dvovrstnega omrežja v pripadajoče navadno (enovrstno) omrežje in analiza pridobljenega omrežja s standardnimi metodami analize omrežij [41]. Zaenkrat obstaja malo metod za neposredno analizo dvovrstnih omrežij [88–90]. Uporabimo lahko dvodelne statistike na stopnjah, posplošeno bločno modeliranje, (p, q)-sredice, dvovrstne opise in kazala, uteži števila štirikotnikov in še nekaj drugih. Veliko metod za identifikacijo pomembnih podomrežij je na voljo za navadna omrežja (posplošene sredice, povezavni otoki, vozliščni otoki, razvrščanje, bločno modeliranje, mere središčnosti in pomembnosti idr.). V tem poglavju bomo predstavili novo metodo za neposredno analizo dvovrstnih omrežij, s katero lahko identificiramo pomembna podomrežja glede na izbrane lastnosti vozlišč.

Ideji posplošenih sredic v navadnih omrežjih in (p, q)-sredic v dvovrstnih omrežjih smo združili v pojem *posplošenih dvovrstnih sredic.* Razvili in implementirali smo algoritem za določitev posplošenih dvovrstnih sredic za izbrane lastnosti vozlišč in praga za obe podmnožici vozlišč. Predlagamo tudi algoritem za iskanje gnezdenih posplošenih dvovrstnih sredic za eno pribito (fiksno) pražno vrednost.

V naslednjem podpoglavju bomo pogledali dela, ki opisujejo ideje, ki smo jih uporabili za razvoj naše metode. V podpoglavju 4.2 predstavimo algoritem za določitev posplošenih dvovrstnih sredic. Naštejemo tudi nekaj zanimivih lastnosti vozlišč, ki določajo vrsto sredic. Algoritem, dokaz njegove pravilnosti in preprosta analiza zahtevnosti algoritma so predstavljeni v podpoglavju 4.2.4. V podpoglavju 4.3 so predstavljeni rezultati uporabe naše metode na realnih podatkih.

4.1 Sorodna dela

Pojem navadne ali k-sredice je uvedel Seidman leta 1983 [25]. Naj bo $\mathscr{G} = (\mathscr{V}, \mathscr{L})$ graf. Naj bo k dano naravno število in naj bo $\deg(v)$ stopnja vozlišča $v \in \mathscr{V}$. Podgraf $\mathscr{H}_k = (\mathscr{C}_k, \mathscr{L}|\mathscr{C}_k)$, porojen s podmnožico $\mathscr{C}_k \subseteq \mathscr{V}$, imenujemo *k-sredica* ali navadna sredica natanko takrat, ko je $\deg_{\mathscr{H}_k}(v) \geq k$ za vse $v \in \mathscr{C}_k$ in je $\mathscr{H}_k$ maksimalen tak podgraf. Če stopnjo nadomestimo s kako drugo lastnostjo vozlišč, dobimo pojem *posplošene sredice,* kot je podana v [26]. Lastnost vozlišč je lahko stopnja, maksimum uteži incidenčnih povezav, vsota uteži incidenčnih povezav idr. Podrobneje so lastnosti vozlišč opisane v podpoglavju 4.2.

Druga posplošitev k-sredic je posplošitev na dvovrstna omrežja. Pojem (p,q)-*sredic* je predstavljen v [27]. Podmnožica $\mathscr{C} \subseteq \mathscr{V}$ določa (p,q)-sredico v dvovrstnem omrežju $\mathscr{N} = ((\mathscr{V}_1,\mathscr{V}_2),\mathscr{L}), \mathscr{V} = \mathscr{V}_1 \cup \mathscr{V}_2$ natanko takrat, ko v podomrežju $\mathscr{K} = ((\mathscr{C}_1,\mathscr{C}_2),\mathscr{L}|\mathscr{C}), \mathscr{C}_1 = \mathscr{C} \cap \mathscr{V}_1, \mathscr{C}_2 = \mathscr{C} \cap \mathscr{V}_2$, porojenem s $\mathscr{C}$, velja za vse $v \in \mathscr{C}_1 : \deg_{\mathscr{K}}(v) \geq p$ in velja za vse $v \in \mathscr{C}_2 : \deg_{\mathscr{K}}(v) \geq q$ ter je $\mathscr{C}$ maksimalna taka podmnožica v $\mathscr{V}$.

Ideji iz posplošenih sredic in (p,q)-sredic smo združili v pojem posplošenih dvovrstnih sredic. Posplošene dvovrstne sredice so definirane podobno kot (p,q)-sredice z eno izjemo – namesto stopnje vozlišč dovolimo uporabo tudi drugih lastnosti vozlišč. Uporabljeni lastnosti vozlišč za podmnožici $\mathscr{V}_1$ in $\mathscr{V}_2$ sta lahko različni.

V določeni literaturi, recimo [91], so dvodelne sredice definirane kot polna dvovrstna podomrežja danega omrežja. Določene so z velikostjo obeh podmnožic vozlišč. Tako je dvodelna sredica ekvivalentna (p,q)-sredici z maksimalnimi stopnjami.

Posplošenim dvovrstnim sredicam podobne metode so še različice skupnostih v dvovrstnih omrežjih: z maksimizacijo monotone funkcije [92], z dualno projekcijo [93], z lastnostmi prostora lastnih vrednosti matrike omrežja [94], z izmenjavo oznak in rekurzivnim deljenjem obeh vrst vozlišč [95], s stohastičnim bločnim modeliranjem [96] in mnogimi drugimi. Še ena zelo podobna metoda je dvodelno razvrščanje [97, 98]. Glavna razlika med temi metodami in našo metodo je to, da je rezultat skupnostih in dvodelnega razvrščanja razvrstitev celotne množice vozlišč, z metodo posplošenih dvovrstnih sredic pa smo se posvetili določanju pomembnega podomrežja. Nobena izmed teh metod tudi ne omogoča tolikšne različnosti – uporabe različnih mer in pragov mer za obe množici vozlišč.

Metode, kot sta (p,q)-sredice in uteži štirikotnikov, so pomembne za iskanje gostejših delov dvovrstnih omrežij. Posplošene dvovrstne sredice pa so odvisne od izbire lastnosti vozlišč in lahko zato izrazijo kakšen drug vidik zgradbe omrežja (recimo moč povezav). Ta metoda torej predstavlja nov pristop k analizi dvovrstnih omrežij. Ne predstavlja le izboljšave neke obstoječe metode ampak posplošitev (p,q)-sredic.

4.2 *Algoritem za posplošene dvovrstne sredice*

Kakor smo že omenili v uvodu tega poglavja, so bili algoritmi za določitev k-sredic, posplošenih sredic in (p,q)-sredic že razviti [25–27]. V tem delu predlagamo nov algoritem, ki je združitev in razširitev ideje posplošenih sredic v navadnih omrežjih in

(p, q)-sredic v dvovrstnih omrežjih. Poleg programske izvedbe novega algoritma tudi dokažemo njegovo pravilnost in analiziramo njegovo zahtevnost. Za preverjanje uporabnosti metode smo algoritem uporabili na realnih omrežjih.

4.2.1 Lastnosti vozlišč

Nekatere lastnosti vozlišč so že bile predlagane v [26]. Za omrežje $\mathscr{N} = (\mathscr{V}, \mathscr{L}, w)$ in utežitveno funkcijo $w : \mathscr{L} \to \mathbb{R}$ predpostavimo, da je *funkcija lastnosti* $f(v, \mathscr{C}) \in \mathbb{R}, \mathscr{C} \subseteq \mathscr{V}$, definirana za vse $v \in \mathscr{V}$. Podmnožica $\mathscr{C}$ poraja podomrežje, na katero je funkcija lastnosti omejena. V neusmerjenem omrežju velja $w(u, v) = w(v, u)$ za vse pare vozlišč $u, v \in \mathscr{V}$.

Označimo soseščino vozlišča v kot $N(v)$ in soseščino vozlišča v znotraj podmnožice $\mathscr{C}$ kot $N(v, \mathscr{C}) = N(v) \cap \mathscr{C}$.

Rečemo, da je funkcija lastnosti $f(v, \mathscr{C})$ *lokalna* natanko takrat, ko velja

$$f(v, \mathscr{C}) = f(v, N(v, \mathscr{C})) \text{ za vse } v \in \mathscr{V}.$$

Funkcija lastnosti $f(v, \mathscr{C})$ je *monotona* natanko takrat, ko velja

$$\mathscr{C}_1 \subset \mathscr{C}_2 \Rightarrow \forall v \in \mathscr{V} : f(v, \mathscr{C}_1) \leq f(v, \mathscr{C}_2).$$

Primeri funkcij lastnosti so:

1. $f_1(v, \mathscr{C}) = \mathrm{deg}_{\mathscr{C}}(v)$: Stopnja vozlišča v znotraj $\mathscr{C}$.
2. $f_2(v, \mathscr{C}) = \mathrm{indeg}_{\mathscr{C}}(v)$: Vhodna stopnja vozlišča v znotraj $\mathscr{C}$.
3. $f_3(v, \mathscr{C}) = \mathrm{outdeg}_{\mathscr{C}}(v)$: Izhodna stopnja vozlišča v znotraj $\mathscr{C}$.
4. $f_4(v, \mathscr{C}) = \mathrm{indeg}_{\mathscr{C}}(v) + \mathrm{outdeg}_{\mathscr{C}}(v)$: Za usmerjeno omrežje je $f_4 = f_1$.
5. $f_5(v, \mathscr{C}) = \sum_{u \in N(v, \mathscr{C})} w(v, u), w : \mathscr{L} \to \mathbb{R}_0^+$: Vsota uteži tistih povezav znotraj $\mathscr{C}$, ki imajo vozlišče v za krajišče.
6. $f_6(v, \mathscr{C}) = \max_{u \in N(v, \mathscr{C})} w(v, u), w : \mathscr{L} \to \mathbb{R}$. Maksimum uteži vseh povezav znotraj $\mathscr{C}$, ki imajo vozlišče v za krajišče.
7. $f_7(v, \mathscr{C}) = \frac{\mathrm{deg}_{\mathscr{C}}(v)}{\mathrm{deg}(v)}$, če je $\mathrm{deg}(v) > 0$ in sicer $f_7(v, \mathscr{C}) = 0$: Delež $N(v, \mathscr{C})$ v $N(v)$.

8. $f_8(v,\mathscr{C}) = \frac{\deg_{\mathscr{C}}(v)}{\max_{u\in N(v)}\deg(u)}$, če je $\deg(v) > 0$ in sicer $f_8(v,\mathscr{C}) = 0$: Relativna gostota soseščine vozlišča v znotraj $\mathscr{C}$.

9. $f_9(v,\mathscr{C}) = \max_{u\in N(v,\mathscr{C})}\deg(u) - \min_{u\in N(v,\mathscr{C})}\deg(u)$: Različnost sosedov vozlišča v znotraj $\mathscr{C}$ glede na njihove stopnje.

10. $f_{10}(v,\mathscr{C}) = \max_{u\in N^+(v,\mathscr{C})}\deg(u) - \min_{u\in N^+(v,\mathscr{C})}\deg(u)$: Različnost polne soseščine vozlišča v znotraj $\mathscr{C}$ glede na stopnje vozlišč v polni soseščini; $N^+(v,\mathscr{C}) = N(v,\mathscr{C}) \cup \{v\}$.

11. $f_{11}(v,\mathscr{C}) = \frac{\sum_{u\in N(v,\mathscr{C})} w(v,u)}{\sum_{u\in N(v)} w(v,u)}, w : \mathscr{L} \to \mathbb{R}_0^+$, če je $\sum_{u\in N(v)} w(v,u) > 0$ in sicer $f_{11}(v,\mathscr{C}) = 0$: Delež uteži povezav z vozliščem v za krajišče, ki imajo tudi drugo krajišče znotraj $\mathscr{C}$.

12. $f_{12}(v,\mathscr{C})$ = število trikotnikov, ki gredo skozi vozlišče v in imajo vsa tri vozlišča znotraj $\mathscr{C}$.

Hitro opazimo, da so vse naštete funkcije lastnosti lokalne in monotone. Primer nemonotone funkcije bi bilo povprečje uteži incidenčnih povezav

$$f(v,\mathscr{C}) = \frac{1}{\deg_{\mathscr{C}}(v)} \sum_{u\in N(v,\mathscr{C})} w(v,u),$$

če je $N(v,\mathscr{C}) \neq \emptyset$ in sicer $f(v,\mathscr{C}) = 0$. Primer nelokalne funkcije bi bilo število ciklov dolžine k, $k \geq 4$, ki gredo skozi posamezno vozlišče.

Za vse naštete funkcije tudi velja $f(v,\emptyset) = 0$ za vsak $v \in \mathscr{V}$.

4.2.2 Posplošene dvovrstne sredice

Definicija 4.1: Naj bo $\mathscr{N} = ((\mathscr{V}_1,\mathscr{V}_2),\mathscr{L},(f,g),w), \mathscr{V} = \mathscr{V}_1 \cup \mathscr{V}_2$ dvovrstno omrežje in naj bo $\mathscr{P}(\mathscr{V})$ potenčna množica množice $\mathscr{V}$. Naj bosta funkciji f in g definirani na omrežju $\mathscr{N} : f,g : \mathscr{V} \times \mathscr{P}(\mathscr{V}) \longrightarrow \mathbb{R}_0^+$.

Posplošena dvovrstna sredica $\mathscr{C} = \mathrm{Core}(p,q;f,g)$, $p,q \in \mathbb{R}_0^+$, je maksimalna množica $\mathscr{C} \subseteq \mathscr{V}$ v dvovrstnem omrežju $\mathscr{N}$, za katero velja, da v podomrežju $\mathscr{K} = ((\mathscr{C}_1,\mathscr{C}_2),\mathscr{L}|\mathscr{C}), \mathscr{C}_1 = \mathscr{C} \cap \mathscr{V}_1, \mathscr{C}_2 = \mathscr{C} \cap \mathscr{V}_2$, porojenim s $\mathscr{C}$, za vsak $v \in \mathscr{C}_1$ velja $f(v,\mathscr{C}) \geq p$ in za vsak $v \in \mathscr{C}_2$ velja $g(v,\mathscr{C}) \geq q$.

V nadaljevanju privzamemo, da je omrežje $\mathcal{N}$ končno – množici $\mathcal{V}$ in $\mathcal{L}$ sta končni. Kadar funkciji f in g poznamo in sta parametra p in q pribita, uporabimo okrajšavo $\mathrm{Core}(p,q) \equiv \mathrm{Core}(p,q;f,g)$.

Množica posplošenih dvovrstnih sredic v omrežju $\mathcal{N} = ((\mathcal{V}_1,\mathcal{V}_2),\mathcal{L},(f,g),w)$ je definirana kot

$$\mathrm{Cores}(\mathcal{N}) = \{\mathrm{Core}(p,q;f,g); p \in \mathbb{R}_0^+ \wedge q \in \mathbb{R}_0^+\}.$$

V tej množici lahko definiramo relacijo

$$\mathrm{Core}(p,q;f,g) \sqsubseteq \mathrm{Core}(p',q';f,g)$$

kot

$$\mathcal{C}_1(p,q;f,g) \subseteq \mathcal{C}_1(p',q';f,g) \ \wedge \ \mathcal{C}_2(p,q;f,g) \subseteq \mathcal{C}_2(p',q';f,g).$$

Ker je $\mathrm{Cores}(\mathcal{N}) \subseteq \mathcal{P}(\mathcal{V})$ in je $(\mathcal{P}(\mathcal{V}),\subseteq)$ delno urejena množica, sta tudi $(\mathrm{Cores}(\mathcal{N}),\subseteq)$ in $(\mathrm{Cores}(\mathcal{N}),\sqsubseteq)$ delno urejeni.

Za posplošene dvovrstne sredice velja:

- $\mathrm{Core}(0,0;f,g) = \mathcal{V}$,
- podomrežje $\mathcal{K} = ((\mathcal{C}_1,\mathcal{C}_2),\mathcal{L}|\mathcal{C})$, $\mathcal{C} = \mathrm{Core}(p,q;f,g)$ ni nujno povezano.

Lema 4.1: Naj bosta $\mathcal{N} = ((\mathcal{V}_1,\mathcal{V}_2),\mathcal{L},\{f,g\},w)$ in njemu zrcaljeno omrežje $\tilde{\mathcal{N}} = ((\mathcal{V}_2,\mathcal{V}_1),\mathcal{L},\{f,g\},w)$ dvovrstni omrežji. Potem velja

$$\mathrm{Core}_{\tilde{\mathcal{N}}}(p,q;f,g) = \mathrm{Core}_{\mathcal{N}}(q,p;g,f).$$

Dokaz:

Po definiciji za $\mathcal{C} = \mathrm{Core}_{\tilde{\mathcal{N}}}(p,q;f,g)$ veljata pogoja

$$\forall v \in \mathcal{C}_2 : f(v,\mathcal{C}) \geq p \quad \text{in} \quad \forall v \in \mathcal{C}_1 : g(v,\mathcal{C}) \geq q$$

in enaka pogoja veljata tudi za $\mathcal{C}' = \mathrm{Core}_{\mathcal{N}}(q,p;g,f)$. Zato sta ti množici enaki. ∎

Lema 4.2: Naj bo $F \subseteq \mathbb{R}$ množica vrednosti funkcije lastnosti f (zaloga vrednosti funkcije lastnosti f) in $\varphi : F \longrightarrow \mathbb{R}_0^+$ strogo naraščajoča funkcija. Potem je

$$\mathrm{Core}(p,q;f,g) = \mathrm{Core}(\varphi(p),q;\varphi \circ f,g).$$

Dokaz: Po definiciji velja $\varphi \circ f(v,\mathscr{C}) = \varphi(f(v,\mathscr{C}))$. $\mathscr{C}$ določa posplošeno dvovrstno sredico $\mathrm{Core}(\varphi(p), q; \varphi \circ f, g)$ natanko takrat, ko velja

$$\forall v \in \mathscr{C}_1 : \varphi \circ f(v,\mathscr{C}) \geq \varphi(p) \quad \text{in} \quad \forall v \in \mathscr{C}_2 : g(v,\mathscr{C}) \geq q$$

$$\Updownarrow \varphi \text{ naraščajoča}$$

$$\forall v \in \mathscr{C}_1 : f(v,\mathscr{C}) \geq p \quad \text{in} \quad \forall v \in \mathscr{C}_2 : g(v,\mathscr{C}) \geq q.$$

To velja natanko takrat, ko $\mathscr{C}$ določa posplošeno dvovrstno sredico $\mathrm{Core}(p, q; f, g)$. ■

Posledica 4.1: Naj bo $F \subseteq \mathbb{R}$ množica vrednosti funkcije lastnosti f in $\varphi : F \longrightarrow \mathbb{R}_0^+$ strogo naraščajoča funkcija. Potem je

$$\mathrm{Core}(q, p; g, f) = \mathrm{Core}(q, \varphi(p); g, \varphi \circ f).$$

Dokaz: Naj bo $\mathscr{N} = ((\mathscr{V}_1, \mathscr{V}_2), \mathscr{L}, (f, g), w)$ omrežje, v katerem so posplošene dvovrstne sredice iz posledice 4.1. Naj bo $\tilde{\mathscr{N}} = ((\mathscr{V}_2, \mathscr{V}_1), \mathscr{L}, (f, g), w)$ njemu zrcalno omrežje. Potem je:

$$\begin{aligned}
\mathrm{Core}_{\mathscr{N}}(q, \varphi(p); g, \varphi \circ f) &\overset{\text{Lema 4.1}}{=} \mathrm{Core}_{\tilde{\mathscr{N}}}(\varphi(p), q; \varphi \circ f, g) \\
&\overset{\text{Lema 4.2}}{=} \mathrm{Core}_{\tilde{\mathscr{N}}}(p, q; f, g) \\
&\overset{\text{Lema 4.1}}{=} \mathrm{Core}_{\mathscr{N}}(q, p; g, f)
\end{aligned}$$

■

Lema 4.2 in posledica 4.1 nam povesta, da z enakovrednimi lastnostmi v urejenostni lestvici dobimo enake posplošene dvovrstne sredice.

4.2.3 Meja za vrednosti pragov

Izrek 4.1: V dvovrstnem omrežju $\mathscr{N} = ((\mathscr{V}_1, \mathscr{V}_2), \mathscr{L}, (f, g), w)$ velja:

$$p_1 \leq p_2 \wedge q_1 \leq q_2 \Rightarrow \mathrm{Core}(p_2, q_2) \subseteq \mathrm{Core}(p_1, q_1).$$

Dokaz: Po definiciji posplošene dvovrstne sredice $\mathscr{C} = \mathrm{Core}(p_2, q_2)$ velja:

$$\forall v \in \mathscr{C}_1 : f(v, \mathscr{C}) \geq p_2 \quad \text{in} \quad \forall v \in \mathscr{C}_2 : g(v, \mathscr{C}) \geq q_2.$$

Zato velja tudi:

$$\forall v \in \mathscr{C}_1 : f(v, \mathscr{C}) \geq p_1 \quad \text{in} \quad \forall v \in \mathscr{C}_2 : g(v, \mathscr{C}) \geq q_1,$$

iz česar sledi

$$\mathrm{Core}(p_2, q_2) = \mathscr{C} \subseteq \mathrm{Core}(p_1, q_1).$$

■

Za dano podmnožico $\mathscr{C} \subseteq \mathscr{V}$ naj bo $p(\mathscr{C}) = \min_{v \in \mathscr{C}_1} f(v, \mathscr{C})$ minimalna vrednost funkcije lastnosti v množici $\mathscr{C}_1 = \mathscr{C} \cap \mathscr{V}_1$ in $q(\mathscr{C}) = \min_{v \in \mathscr{C}_2} g(v, \mathscr{C})$ minimalna vrednost funkcije lastnosti v množici $\mathscr{C}_2 = \mathscr{C} \cap \mathscr{V}_2$. Velja $\mathscr{C} \subseteq \mathrm{Core}(p(\mathscr{C}), q(\mathscr{C}))$.

Za dano dvovrstno omrežje $\mathscr{N} = (\mathscr{V}, \mathscr{L}, (f, g), w)$ naj bo $P = \{p(\mathscr{C}) : \mathscr{C} \subseteq \mathscr{V}\}$ množica vseh možnih vrednosti $p(\mathscr{C})$ in $Q = \{q(\mathscr{C}) : \mathscr{C} \subseteq \mathscr{V}\}$ množica vseh možnih vrednosti $q(\mathscr{C})$. P in Q sta končni množici. Torej lahko njune elemente uredimo v seznam:

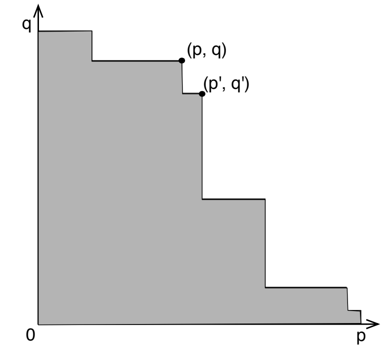
$$P = \{p_1, p_2, \dots, p_r\}; \quad p_i < p_{i+1},$$
$$Q = \{q_1, q_2, \dots, q_s\}; \quad q_i < q_{i+1}.$$

Za izbrani funkciji lastnosti f in g nas zanimajo le tiste vrednosti para (p, q), ki določajo različne neničelne posplošene dvovrstne sredice.

V kartezičnem koordinatnem sistemu s koordinatama (p, q) (slika 4.1) je območje vseh možnih parov pragov p in q omejeno na prvi kvadrant. Manjkajoča meja tega območja je zlomljenka v obliki stopnic.

Naj bo za $p, q \in \mathbb{R}_0^+$ $\Pi(p) = \{\mathscr{C} : \mathscr{C} \subseteq \mathscr{V} \wedge p(\mathscr{C}) \geq p\}$ množica množic, za katere je vrednost funkcije lastnosti prve podmnožice enaka vsaj p, in $\Gamma(q) = \{\mathscr{C} : \mathscr{C} \subseteq \mathscr{V} \wedge q(\mathscr{C}) \geq q\}$ množica množic, za katere je vrednost funkcije lastnosti druge podmnožice enaka vsaj q. Naj bo $G(p) = \{q(\mathscr{C}) : \mathscr{C} \in \Pi(p)\}$ množica vseh možnih vrednosti $q(\mathscr{C})$, za katere množica $\mathscr{C}$ pripada množici $\Pi(p)$ in je $q_{\Pi(p)} = \max G(p)$ maksimalna takšna vrednost $q(\mathscr{C})$. Naj bo $H(q) = \{p(\mathscr{C}) : \mathscr{C} \in \Gamma(q)\}$ množica vseh možnih vrednosti $p(\mathscr{C})$, za katere množica $\mathscr{C}$ pripada množici $\Gamma(q)$ in je $p_{\Gamma(q)} = \max H(q)$ maksimalna takšna vrednost $p(\mathscr{C})$.

Lema 4.3: Za $p, p' \in P$ in $q, q' \in Q$ veljajo naslednje lastnosti:



Slika 4.1

Območje vseh možnih parov pragov (p, q), ki določajo posplošene dvovrstne sredice.

1. $\Gamma(q_{\Pi(p)}) \neq \emptyset$ in $\Pi(p_{\Gamma(q)}) \neq \emptyset$,
2. $\mathscr{C} \in \Gamma(q_{\Pi(p)}) \Rightarrow \mathscr{C} \subseteq \mathrm{Core}(p, q_{\Pi(p)})$ in
$\mathscr{C} \in \Pi(p_{\Gamma(q)}) \Rightarrow \mathscr{C} \subseteq \mathrm{Core}(p_{\Gamma(q)}, q)$,
3. $q > q_{\Pi(p)} \Rightarrow \mathrm{Core}(p, q) = \emptyset$ in
$p > p_{\Gamma(q)} \Rightarrow \mathrm{Core}(p, q) = \emptyset$,
4. $p < p' \Rightarrow q_{\Pi(p')} \leq q_{\Pi(p)}$ in
$q < q' \Rightarrow p_{\Gamma(q')} \leq p_{\Gamma(q)}$.

Dokaz: Dokazi za lastnosti so preprosti, zato pokažimo le dokaz za prvi del četrte lastnosti.

$$\begin{aligned} \Pi(p') &= \{\mathscr{C} : \mathscr{C} \subseteq \mathscr{V} \wedge p(\mathscr{C}) \geq p'\} \\ &\subseteq \{\mathscr{C} : \mathscr{C} \subseteq \mathscr{V} \wedge p(\mathscr{C}) \geq p\} = \Pi(p) \end{aligned}$$

Iz tega sledi $G(p') \subseteq G(p)$ in zato $q_{\Pi(p')} \leq q_{\Pi(p)}$. ■

Iz teh lastnosti izhaja, da za vsak $p \in P$ obstaja maksimalen $q \in Q$: $q = q_{\Pi(p)}$; in za vsak $q \in Q$ obstaja maksimalen $p \in P : p = p_{\Gamma(q)}$. To je formalni dokaz stopničaste oblike meje. Na osnovi tega lahko razvijemo algoritem za določitev meje množice (P, Q) v koordinatnem sistemu (p, q) – glej algoritem 4.1.

V splošnem ima algoritem 4.1 le teoretično vrednost, saj sta lahko množici P in Q zelo veliki. Uporabimo ga lahko za posebne funkcije lastnosti – recimo f_1, f_2, f_3 in f_4, za katere sta množici P in Q relativno majhni.

Algoritem 4.1

Algoritem za določanje meje množice (P, Q) v koordinatnem sistemu (p, q).

Vhod: $P = \{p_1, p_2, \dots, p_r\}; p_i < p_{i+1}$.
Izhod: množica $meja \subseteq P \times Q$.
Algoritem:
$q_{\max} \leftarrow 0$
$meja \leftarrow \emptyset$
for $p \in reverse(P)$ *do*
  $q \leftarrow q_\Pi(p)$
  if $q > q_{max}$ *then*
    $q_{max} \leftarrow q$
    $meja \leftarrow meja \cup \{(p, q)\}$
  end if
end for

4.2.4 Algoritem

Predlagamo algoritem za določanje posplošenih dvovrstnih sredic v dvovrstnih omrežjih za dani vrednosti pragov p in q in za dani funkciji lastnosti f in g.

Osnovna ideja algoritma za posplošene dvovrstne sredice je v ponavljanju odstranjevanja vozlišč, ki sredici ne pripadajo. Ker je omrežje dvovrstno, je pogoj odstranjevanja vozlišč odvisen od tega, kateri množici vozlišče pripada, kakor je razvidno iz algoritma 4.2.

Algoritem 4.2

Osnovni algoritem za določitev posplošene dvovrstne sredice..

$\mathscr{C} \leftarrow \mathscr{V}$
while $\exists v \in \mathscr{C} : (v \in \mathscr{C}_1 \wedge f(v, \mathscr{C}) < p) \vee (v \in \mathscr{C}_2 \wedge g(v, \mathscr{C}) < q)$ *do*
  $\mathscr{C} \leftarrow \mathscr{C} \setminus \{v\}$
end while

S preprosto predelavo dokaza izreka 1 iz [26] lahko dokažemo naslednji izrek.

Izrek 4.2: Algoritem 4.2 določi posplošeno dvovrstno sredico za dani vrednosti pragov

Algoritem 4.3

Algoritem za določitev posplošene dvovrstne sredice za vrednosti pragov (p, q) in za monotoni in lokalni funkciji lastnosti vozlišč f in g.

Vhod: omrežje $\mathscr{N} = (\mathscr{V}_1 \cup \mathscr{V}_2, \mathscr{L}, (f, g), w)$, $\mathscr{V}_1 \cap \mathscr{V}_2 = \emptyset$; vrednosti pragov p, q.
Izhod: posplošena dvovrstna sredica $\mathscr{C} = \mathscr{C}_1 \cup \mathscr{C}_2$.

Procedura:
function ODSTRANJUJ(h, t, $\mathscr{C}_{trenutna}$, $\mathscr{C}_{druga}$, kopica$_{trenutna}$, kopica$_{druga}$)
    while ¬ Prazna(kopica$_{trenutna}$) *do*
        if Koren(kopica$_{trenutna}$).vrednost ≥ t *then return*
        end if
        u ← Koren(kopica$_{trenutna}$).ključ
        OdstraniKoren(kopica$_{trenutna}$)
        $\mathscr{C}_{trenutna} \leftarrow \mathscr{C}_{trenutna} \setminus \{u\}$
        for $v \in$ N(u, $\mathscr{C}_{druga}$) *do*
            Posodobi kopico$_{druga}$(v, h(v, $\mathscr{C}$))
        end for
    end while
end function

Algoritem:
$\mathscr{C}_1 \leftarrow \mathscr{V}_1$, $\mathscr{C}_2 \leftarrow \mathscr{V}_2$
for $v \in \mathscr{V}_1$ *do*
    vrednost[v] ← f(v, $\mathscr{V}$)
end for
for $v \in \mathscr{V}_2$ *do*
    vrednost[v] ← g(v, $\mathscr{V}$)
end for
zgradi kopico$_1$(vrednost, $\mathscr{C}_1$), kopico$_2$(vrednost, $\mathscr{C}_2$)
repeat
    ODSTRANJUJ(f, p, $\mathscr{C}_1$, $\mathscr{C}_2$, kopica$_1$, kopica$_2$)
    ODSTRANJUJ(g, q, $\mathscr{C}_2$, $\mathscr{C}_1$, kopica$_2$, kopica$_1$)
until še odstraniš kakšno vozlišče

(p, q) za katerikoli monotoni funkciji lastnosti vozlišč $f(v, \mathscr{C})$ in $g(v, \mathscr{C})$. Rezultat algoritma ni odvisen od vrstnega reda brisanja vozlišč.

Ker vrstni red brisanja ne vpliva na končni rezultat, lahko pri oblikovanju algoritma začnemo z brisanjem iz prve množice, nadaljujemo v drugi, zopet v prvi in to nadaljujemo, dokler lahko še kaj brišemo.

Za shranjevanje vozlišč smo uporabili kopice (vrste s prednostjo) [99], v katerih smo vozlišča uredili tako, da lahko do vozlišča z najmanjšo vrednostjo funkcije lastnosti učinkovito dostopamo kot korena kopice. Vrednost funkcije lastnosti izračunamo za vsako vozlišče posebej glede na množico povezav $\mathscr{L}$ in uteži w, kakor je opisano v razdelku 4.2.1.

Potrebujemo dve kopici – za vsako podmnožico vozlišč eno. Element `E` v kopici je par `E = (ključ, vrednost)`, pri čemer je identifikator vozlišča `E.ključ` in vrednost funkcije lastnosti `E.vrednost`. Elementi so urejeni po njihovih vrednostih funkcije lastnosti. Vemo, da so vsi sosedi vozlišč iz prve podmnožice v drugi podmnožici in obratno.

Za omejitev posodabljanja predpostavimo, da sta obe funkciji lastnosti f in g lokalni. V tem primeru je dovolj posodobiti le kopico, v kateri so sosedje izbrisanega vozlišča.

Te omejitve smo uporabili v algoritmu 4.3 za določanje posplošene dvovrstne sredice za monotoni in lokalni funkciji lastnosti f in g.

V algoritmu 4.3 sta oba praga fiksirana. Zato nas ne zanima, ali odstranimo iz posamezne podmnožice trenutne posplošene dvovrstne sredice vozlišče z najmanjšo vrednostjo funkcije lastnosti. Vedeti moramo le, ali je vrednost funkcije lastnosti vozlišča manjša od vrednosti praga. Torej lahko algoritem poenostavimo – namesto dvojiških kopic uporabimo vrste: za vsako izmed množic vozlišč omrežja uporabimo eno vrsto za vozlišča, katerih vrednost funkcije lastnosti ustreza pragu (so enake ali nad pragom), in eno vrsto za vozlišča, katerih vrednosti funkcije lastnosti ne ustrezajo pragu (so pod pragom). V algoritmu 4.4 je prikazan postopek z uporabo vrst namesto dvojiških kopic.

4.2.5 Časovna zahtevnost algoritma

Predpostavimo, da je časovna zahtevnost računanja vrednosti lokalne funkcije lastnosti $O(\deg(v))$ za vsako vozlišče v v omrežju. Ker je uporaba vrst časovno manj zahtevna,

Algoritem 4.4

Algoritem za določitev posplošene dvovrstne sredice za vrednosti pragov (p, q) in za monotoni in lokalni funkciji lastnosti vozlišč f in g – z vrstami.

Vhod: omrežje $\mathcal{N} = ((\mathcal{V}_1, \mathcal{V}_2), \mathcal{L}, (f, g), w)$, vrednosti pragov p, q.
Izhod: posplošena dvovrstna sredica $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2$.

Procedura:

```
function ODSTRANJUJ(h, t, 𝒞_trenutna, 𝒞_druga, pod_trenutna, nad_druga, pod_druga)
    while pod_trenutna ≠ ∅ do
        u ← odstraniElement iz pod_trenutna
        𝒞_trenutna ← 𝒞_trenutna \ {u}
        for v ∈ N(u, 𝒞_druga) do
            if (v, k) ∈ nad_druga then
                k_nov ← posodobiVrednost(v, h(v, 𝒞))
                if k_nov < t then
                    nad_druga ← nad_druga \ {(v, k)}
                    pod_druga ← pod_druga ∪ {v}
                else
                    posodobiElement((v, k), k_nov) v nad_druga
                end if
            end if
        end for
    end while
end function
```

Algoritem 4.4

Algoritem za določitev posplošene dvovrstne sredice za vrednosti pragov (p, q) in za monotoni in lokalni funkciji lastnosti vozlišč f in g – z vrstami (nadaljevanje).

Algoritem:
$\mathscr{C}_1 \leftarrow \mathscr{V}_1, \mathscr{C}_2 \leftarrow \mathscr{V}_2$
$\mathrm{nad}_1, \mathrm{pod}_1, \mathrm{nad}_2, \mathrm{pod}_2 \leftarrow \emptyset$
for $v \in \mathscr{V}_1$ *do*
    if $\mathrm{f}(v, \mathscr{V}) < \mathrm{p}$ *then*
        dodaj v v pod_1
    else
        dodaj $(v, \mathrm{f}(v, \mathscr{V}))$ v nad_1
    end if
end for
for $v \in \mathscr{V}_2$ *do*
    if $\mathrm{g}(v, \mathscr{V}) < \mathrm{q}$ *then*
        dodaj v v pod_2
    else
        dodaj $(v, \mathrm{g}(v, \mathscr{V}))$ v nad_2
    end if
end for
repeat
    Odstranjuj$(\mathrm{f}, \mathrm{p}, \mathscr{C}_1, \mathscr{C}_2, \mathrm{pod}_1, \mathrm{nad}_2, \mathrm{pod}_2)$
    Odstranjuj$(\mathrm{g}, \mathrm{q}, \mathscr{C}_2, \mathscr{C}_1, \mathrm{pod}_2, \mathrm{nad}_1, \mathrm{pod}_1)$
until še odstraniš kakšno vozlišče

bomo računali časovno zahtevnost za postopek z uporabo teh podatkovnih struktur in ne dvojiških kopic. Gradnja podatkovnih struktur zahteva $O(n)$ časa. Ker velja $\sum_{v \in \mathscr{V}} \deg(v) \leq 2m$, je časovna zahtevnost inicializacije enaka

$$O(\sum_{v \in \mathscr{V}} \deg(v) + n) = O(m + n).$$

Na vsakem koraku zanke algoritma odstranimo kako vozlišče, njegovim sosedom pa se spremeni vrednost funkcije lastnosti. Glede na nove vrednosti jih lahko iz slovarja prestavimo v množico.

Za posodobitev vrednosti funkcije lastnosti lahko porabimo manj časa kot za izračun. Recimo vrednost funkcije $f_1(v, \mathscr{C}) = \deg_{\mathscr{C}}(v)$ popravimo tako, da jo zmanjšamo za ena vsakič, ko odstranimo kakega soseda vozlišča v. Posodobitev vrednosti funkcij $f_1, f_2, f_3, f_4, f_5, f_7, f_8$ in f_{11} traja $O(1)$ časa; za funkcije lastnosti f_6, f_9, f_{10} in f_{12} pa potrebujemo $O(\deg(v))$ časa za vsako vozlišče v.

Množici in slovarja so implementirani tako, da sprememba lege elementa zahteva $O(1)$ časa.

Naj bo $s = (v_1, v_2, \dots, v_d)$ zaporedje vozlišč, ki so bila odstranjena tekom algoritma. Odstranitev vozlišča v_i traja $O(1)$ časa. Posodobitev vrednosti funkcije lastnosti vsakega njegovega soseda in sprememba lege tega soseda traja $O(1)$ oziroma $O(\deg(v_i))$ časa, odvisno od uporabljene funkcije lastnosti; za celotno zaporedje s posodobitev traja

$$\sum_{v_i \in s} \deg(v_i) \cdot O(1) \leq 2m \cdot O(1) = O(m)$$

za funkcije lastnosti $f_1, f_2, f_3, f_4, f_5, f_7, f_8$ in f_{11} ter

$$\begin{aligned} & \sum_{v_i \in s} \deg(v_i) \cdot O(\deg(v_i)) \\ \leq \quad & O(\Delta) \cdot \sum_{v_i \in s} \deg(v_i) \\ \leq \quad & 2m \cdot O(\Delta) = O(m \cdot \Delta), \end{aligned}$$

kjer je $\Delta = \max_{v \in \mathscr{V}}(\deg(v))$, za funkcije lastnosti f_6, f_9, f_{10} in f_{12}.

Časovna zahtevnost inicializacije algoritma je manjša od časovne zahtevnosti glavne zanke algoritma. Torej je časovna zahtevnost celotnega algoritma za določanje posplošene dvovrstne sredice enaka

$$O(m)$$

za funkcije lastnosti $f_1, f_2, f_3, f_4, f_5, f_7, f_8$ in f_{11} ter

$$O(m \cdot \Delta)$$

za funkcije lastnosti f_6, f_9, f_{10} in f_{12}. Če bi namesto vrst uporabili dvojiške kopice, bi se časovna zahtevnost v obeh primerih povečala za faktor $O(\log n)$.

4.2.6 Algoritem za eno dano pražno vrednost

Algoritem 4.3 smo prilagodili tako, da smo pribili samo en prag za podmnožici vozlišč. Rezultat novega algoritma je vektor, v katerem ima vsako vozlišče vrednost, ki je enaka maksimalni vrednosti nepribitega praga, za katerega je vozlišče še v posplošeni dvovrstni sredici. Ta algoritem lahko uporabimo za lažje določanje pragov, za katere dobimo najbolj zanimive posplošene dvovrstne sredice.

Algoritem 4.5 predstavlja tovrstno prilagoditev algoritma 4.3. V tem algoritmu pribijemo prag p za prvo podmnožico vozlišč. Za določitev praga za drugo podmnožico vozlišč uporabimo lemo 4.1.

Algoritem temelji na ideji brisanja vozlišč, ki ne zadoščajo pragu. V implementaciji tega algoritma uporabimo kopice, saj potrebujemo informacijo o vozlišču z najmanjšo vrednostjo funkcije lastnosti. Vrednost funkcije lastnosti izračunamo glede na množico povezav $\mathscr{L}$ in uteži w, kakor je opisano v razdelku 4.2.1. Vozlišča so v kopicah obeh podmnožic vozlišč glede na vrednost funkcije lastnosti.

Posamezen korak glavne zanke algoritma začnemo z brisanjem vseh vozlišč v kopici za prvo podmnožico vozlišč, ki ne zadoščajo pribitemu pragu. Odstranjenim vozliščem damo v končnem vektorju vrednost praga za drugo podmnožico iz prejšnje ponovitve zanke, torej vrednost drugega praga, za katero je to vozlišče še v posplošeni dvovrstni sredici. Vrednosti funkcije lastnosti sosednjih vozlišč se lahko spremenijo, zato nastavimo drugi prag q na trenutno najmanjšo vrednost v kopici druge podmnožici vozlišč po prvem delu koraka. Po tem odstranimo vsa vozlišča iz druge kopice, ki imajo vrednost funkcije lastnosti enako trenutni vrednosti q. Odstranjenim vozliščem v končnem vektorju nastavimo vrednosti na q. Na koncu vsake ponovitve zanke nastavimo prejšnjo vrednost praga q na njegovo trenutno vrednost, njegovo trenutno vrednost pa na vrednost v korenu druge kopice.

Algoritem 4.5

Algoritem za določanje množice posplošenih dvovrstnih sredic za pribit prag p in za monotoni in lokalni funkciji lastnosti vozlišč f in g.

Vhod: omrežje $\mathcal{N} = (\mathcal{V}_1 \cup \mathcal{V}_2; \mathcal{L}, \{f, g\}, w)$, $\mathcal{V}_1 \cap \mathcal{V}_2 = \emptyset$; prag p.
Izhod: vektor T, $T[v]$ = taka maksimalna vrednost praga q, da je $v \in \mathrm{Core}(p, q)$.

Proceduri:
function OdstranjujPribitega(f, p, q, $\mathcal{C}_1$, $\mathcal{C}_2$, $kopica_1$, $kopica_2$)
    while ¬Prazna($kopica_1$) *do*
        if Koren($kopica_1$).vrednost ≥ p *then return*
        end if
        u ← Koren($kopica_1$).ključ
        OdstraniKoren($kopica_1$)
        T[u] ← q
        $\mathcal{C}_1 \leftarrow \mathcal{C}_1 \setminus \{u\}$
        for $v \in$ N(u, $\mathcal{C}_2$) *do*
            posodobi $kopico_2$(v, f(v, $\mathcal{C}_1$))
        end for
    end while
end function
function OdstranjujSpreminjajočega(g, q, $\mathcal{C}_2$, $\mathcal{C}_1$, $kopica_2$, $kopica_1$)
    while ¬Prazna($kopica_2$) *do*
        if Koren($kopica_2$).vrednost > q *then return*
        end if
        u ← Koren($kopica_2$).ključ
        OdstraniKoren($kopica_2$)
        T[u] ← q
        $\mathcal{C}_2 \leftarrow \mathcal{C}_2 \setminus \{u\}$
        for $v \in$ N(u, $\mathcal{C}_1$) *do*
            posodobi $kopico_1$(v, g(v, $\mathcal{C}_2$))
        end for
    end while
end function

Algoritem 4.5

Algoritem za določanje množice posplošenih dvovrstnih sredic za pribit prag p in za monotoni funkciji lastnosti vozlišč f in g (nadaljevanje).

Algoritem:
$\mathscr{C}_1 \leftarrow \mathscr{V}_1, \mathscr{C}_2 \leftarrow \mathscr{V}_2$
for $v \in \mathscr{V}_1$ *do*
  vrednost[v] $\leftarrow$ f(v, $\mathscr{V}_2$), $T[v] \leftarrow -1$
end for
for $v \in \mathscr{V}_2$ *do*
  vrednost[v] $\leftarrow$ g(v, $\mathscr{V}_1$), $T[v] \leftarrow -1$
end for
q ← -1
zgradi kopico$_1$(vrednost, $\mathscr{V}_1$), kopico$_2$(vrednost, $\mathscr{V}_2$)
repeat
  OdstranjujPribitega(f, p, q, $\mathscr{C}_1$, $\mathscr{C}_2$, kopica$_1$, kopica$_2$)
  if ¬Prazna(kopica$_2$) *then*
    q ← Koren(kopica$_2$).vrednost
    OdstranjujSpreminjajočega(g, q, $\mathscr{C}_2$, $\mathscr{C}_1$, kopica$_2$, kopica$_1$)
  end if
until ¬(Prazna(kopica$_1$) ∧ Prazna(kopica$_2$))

4.3 Uporaba metode

Ker lahko izberemo različne lastnosti vozlišč za določanje pomembnih dvovrstnih podomrežij, lahko z analizo pridobimo informacije o različnih vrstah skupin z eno samo metodo. Nabor funkcij lastnosti je odprt – nadaljujemo z iskanjem zanimivih lastnosti vozlišč, ki jih lahko vključimo v naš seznam in v implementacijo ter jih uporabimo na realnih podatkih.

Metodo posplošenih dvovrstnih sredic lahko uporabimo na različnih realnih podatkih. Vhodno omrežje ni nujno dvovrstno. Metodo lahko uporabimo tudi na navadnih enovrstnih omrežjih, ki jih obravnavamo kot dvovrstna omrežja. Recimo v omrežju citiranj lahko 'vrstične' avtorje obravnavamo kot uporabnike in 'stolpične' avtorje kot ponudnike znanja. Uporaba posplošenih dvovrstnih sredic analitiku omogoča iskanje podomrežij, ki so določena z različnima funkcijama lastnosti.

4.3.1 Social Networks – področje družbenih omrežij

Oglejmo si uporabo posplošenih dvovrstnih sredic na primeru. Web of Science je bibliografska baza. Uporabili bomo podatke, ki so bili iz te baze pridobljeni leta 2008 z iskalnim nizom 'social network*'. Te podatke smo dopolnili z opisi najpogostejših referenc in z bibliografijami okoli 100 analitikov družabnih omrežij. Iz teh podatkov smo zgradili nekaj dvovrstnih omrežij. Dve izmed njih sta omrežje del × revij ter omrežje del × avtorjev (193376 del, 14651 revij in 75930 avtorjev). Pomnožili smo transponirano omrežje prvega omrežja z drugim omrežjem in dobili omrežje revij × avtorjev. Revija in avtor sta v tem omrežju povezana, če je avtor objavil vsaj eno delo v reviji. Utež povezave je enaka številu teh del.

Enostavne posplošene dvovrstne sredice dobimo z uporabo enakih funkcij lastnosti za obe podmnožici vozlišč. Če izberemo

$$f_A(v, \mathscr{C}) = g_A(v, \mathscr{C}) = \deg_{\mathscr{C}}(v),$$

dobimo (p, q)-sredice. V (p, q)-sredici so revije, ki so objavile dela vsaj p različnih avtorjev v tej sredici, in v tej sredici so avtorji, ki so objavili dela v vsaj q različnih revijah v tej sredici.

Določili smo posplošeno dvovrstno sredico za funkciji f_A in g_A in izbrani vrednosti pragov $p = 85$ in $q = 3$, za kateri dobimo najmanjšo posplošeno dvovrstno sredico. Ta sredica je ena izmed posplošenih dvovrstnih sredic na robu (p, q) območja. S tem dobimo podomrežje del, v katerih je svoja dela objavilo vsaj 85 avtorjev (v tem podomrežju), in avtorji, ki so svoja dela objavili v vsaj 3 različnih revijah (v tem podomrežju). V dobljenem podomrežju so 4 revije in 128 avtorjev. V tabeli 4.3.1 so našteti avtorji, ki so povezani z vsemi štirimi revijami znotraj dobljene posplošene dvovrstne sredice. Revije so American Sociological Review (stopnje 122), American Journal of Sociology (112), Social Forces (90) in Annual Review of Sociology (85).

Če želimo upoštevati število objavljenih del (uteži povezav), lahko uporabimo

$$f_B(v, \mathscr{C}) = \sum_{u \in N(v, \mathscr{C})} w(v, u),$$

$g_B(v, \mathscr{C}) = \deg_{\mathscr{C}}(v)$ pa ostane enaka. V tej posplošeni dvovrstni sredici s pragoma p in q so revije, v katerih so avtorji iz te sredice objavili vsaj p del, in avtorji, ki so objavili svoja dela v vsaj q različnih revijah iz te sredice.

Tabela 4.1

Avtorji v Core(85, 3; $\deg_{\mathscr{C}}(v), \deg_{\mathscr{C}}(v)$), ki so povezani z vsemi revijami v tej posplošeni dvovrstni sredici.

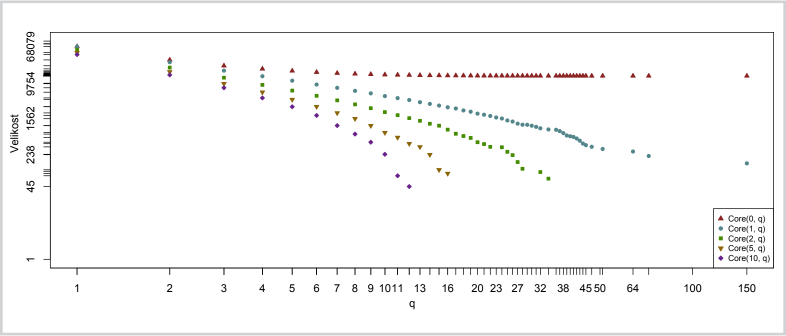
1	Breiger R.	10	Kandel D.	19	Olzak S.
2	Burt R.	11	Keister L.	20	Portes A.
3	DiMaggio P.	12	Knoke D.	21	Reskin B.
4	Fischer C.	13	Lieberson S.	22	Ridgeway C.
5	Friedkin N.	14	Lin N.	23	Sampson R.
6	Galaskiewicz J.	15	Marsden P.	24	Skvoretz J.
7	Glass J.	16	McPherson J.	25	Thoits P.
8	Kalleberg A.	17	Mizruchi M.		
9	Kalmijn M.	18	Nee V.		

Določili smo nekaj posplošenih dvovrstnih sredic za ti dve funkciji. Uporabili smo algoritem za en pribit prag. Izbrali smo $p \in \{0, 1, 2, 5, 10\}$ in določili posplošene dvovrstne sredice za vseh pet vrednosti praga p in spreminjajočega praga q. Na sliki 4.2 je prikazan diagram povezanosti med vrednostjo praga q in velikostjo posplošene dvovrstne sredice. Opazimo, da se velikost posplošene dvovrstne sredice spreminja glede na vrednost praga p za vse vrednosti praga q. Ker sta koordinatni osi v logaritmični lestvici, točka za Core$(0, 0; f_B, g_B)$ ni prikazana. Njena velikost je enaka velikosti množice vozlišč, saj je ta posplošena dvovrstna sredica enaka celotnemu omrežju. Na sliki 4.3 je prikazana Core$(10, 12; f_B, g_B)$ za izbrani funkciji lastnosti. To je najmanjša posplošena dvovrstna sredica za $p = 10$. V Core$(10, 12; f_B, g_B)$ so vse revije, v katerih so avtorji iz te sredice objavili vsaj 10 del, in vsi avtorji, ki so objavili svoja dela v vsaj 12 različnih revijah. Debelina povezav predstavlja število objavljenih del – debelejša povezava pomeni več del. Na sliki 4.3 tudi hitro opazimo še vedno prisoten problem identifikacije imen – pari identifikatorjev revij

- `amer sociol rev`, `am sociol rev`: American Sociological Review,
- `amer j sociol`, `am j sociol`: American Journal of Sociology,
- `adm sci q`, `admin sci quart`: Administrative Science Quarterly

predstavljajo isto revijo. Zapis `sociol method` predstavlja eno izmed treh revij: Sociological Methodology, Sociological Methods & Research in International Journal of

Social Research Methodology. To predstavlja še dodaten problem, saj se reviji Sociological Methodology in Sociological Methods & Research pojavita na sliki 4.3 še s svojima identifikatorjema (`sociol methodol` in `sociological methods`).



Slika 4.2

Relacija med pragom q in velikostjo posplošene dvovrstne sredice za fiksno vrednost praga p in funkcij lastnosti f_B in g_B.

Izberemo lahko tudi zapletenejši funkciji lastnosti:

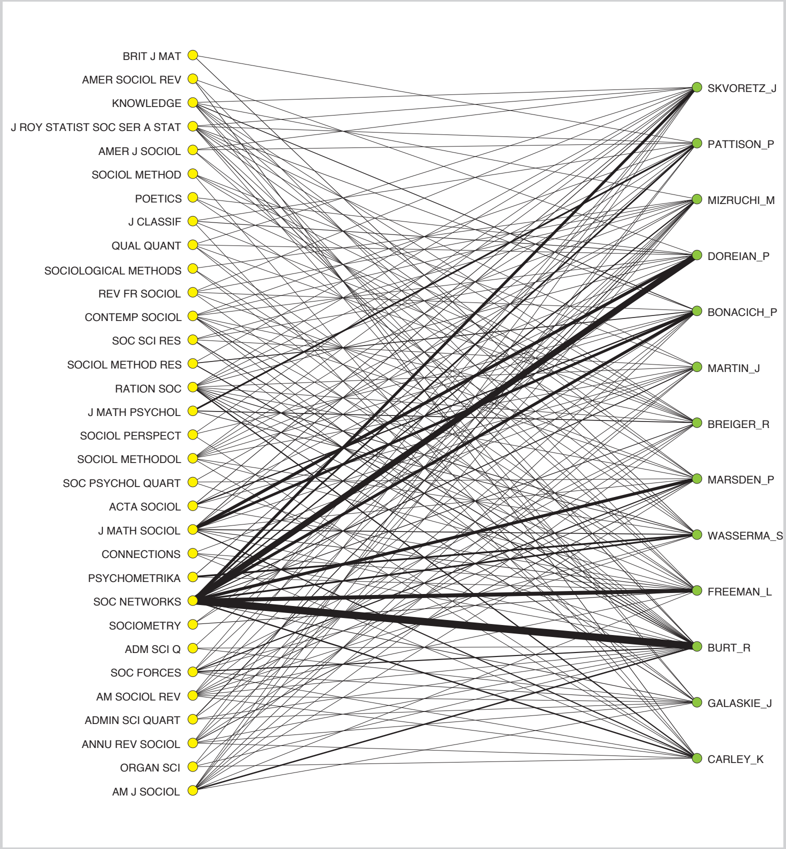
$$f_C(v, \mathscr{C}) = \max_{u \in N(v,\mathscr{C})} \deg(u) - \min_{u \in N(v,\mathscr{C})} \deg(u)$$

in

$$g_C(v, \mathscr{C}) = \frac{\sum_{u \in N(v,\mathscr{C})} w(v,u)}{\sum_{u \in N(v)} w(v,u)}.$$

V Core($p, q; f_C, g_C$) so revije, v katerih je objavljeno približno (glede na vrednost praga p) enako število del vsakega avtorja iz te sredice, ki so povezani s temi revijami. V tej sredici so tisti avtorji, ki so objavili vsaj q% svojih del v revijah iz te sredice. Uporaba posplošenih dvovrstnih sredic s tema dvema funkcijama lastnosti omogoča iskanje revij in avtorjev, ki so tesno povezani.

Določili smo mejo (p, q) območja posplošenih dvovrstnih sredic za izbrani funkciji lastnosti. Meja je prikazana na sliki 4.4. Poleg vsakega zunanjega kota meje je zapisan par velikosti obeh podmnožic vozlišč v posplošeni dvovrstni sredici. Tako je recimo Core($70, 0.25; f_C, g_C$) sestavljena iz 26 vozlišč v prvi množici in 118 vozlišč v drugi množici.



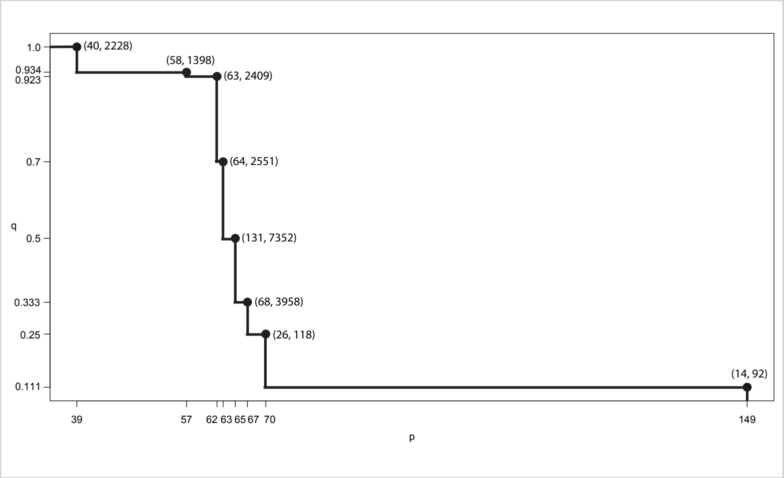
Slika 4.3

Posplošena dvovrstna sredica za $p = 10$ in $q = 12$ ter za funkciji lastnosti f_B in g_B.

4.4 Zaključek

V tem poglavju predlagamo novo metodo za neposredno analizo dvovrstnih omrežij. Predlagamo tudi algoritem za določanje posplošenih dvovrstnih sredic in predstavimo nekaj primerov uporabe metode na realnih podatkih. Predstavljeni pristop lahko razširimo na večvrstna omrežja.

Algoritem 4.3 smo že razširili v algoritem za iskanje gnezdenih posplošenih dvovrstnih sredic za en fiksen prag, ki je prikazana v algoritmu 4.5. Algoritem želimo razširiti



Slika 4.4

Meja (p, q) območja za f_C in g_C.

še dlje – za izdelavo vseh (p, q) parov, ki definirajo različne posplošene dvovrstne sredice in za določitev meje (p, q) parov, kot je prikazano na sliki 4.1.

5

Zaključek

5.1 Glavni znanstveni dosežki

V tem podpoglavju so povzeti glavni znanstveni dosežki. Poleg vsakega dosežka je zapisano poglavje disertacije, kjer je ta dosežek podrobneje opisan. Dodan je tudi seznam mojih prispevkov, ki obravnavajo posamezen dosežek. Ti prispevki so bili predstavljeni na mednarodnih konferencah, objavljeni v priznanih znanstvenih revijah ali objavljeni v priznani znanstveni enciklopediji.

- *Narejen pregled polkolobarjev za analizo omrežij*
  V poglavju 2 je opisan algoritem za hitro množenje velikih redkih omrežij, katerega uporabnost v analizi omrežij lahko razširimo z navedenimi polkolobarji, kar omogoča nove poglede na lastnosti produkta dveh omrežij. Uporaba polkolobarjev v analizi omrežij je opisana v prispevku [100].

- *Pregled pristopov k problemom pri pridobivanju omrežnih podatkov*
  V poglavju 3 so opisani glavni pristopi k pridobivanju podatkov in tudi delne rešitve problemov, ki se pojavljajo pri tem. Rezultati so bili objavljeni v prispevku [101] in v članku [102].

- *Množenje omrežij v analizi omrežij*
  V poglavju 3 sta poleg virov podatkov navedena dva primera analize omrežij. V teh primerih je poudarjen pomen uporabe množenja omrežij in normalizacije omrežja, ki omogočata izpeljavo novih omrežij, ki zaradi ustrezne normalizacije dajo smiselne rezultate. Ta dva primera sta opisana v člankih [35] in [102].

- *Posplošene dvovrstne sredice*
  Metoda posplošenih dvovrstnih sredic je opisana v poglavju 4 in je nova metoda za neposredno analizo dvovrstnih omrežij. Predstavljena je bila v prispevku na konferenci [103], članek o tej metodi pa je v recenzijskem postopku v reviji Social Networks.

5.2 Nadaljne raziskovalno delo

Uporaba različnih polkolobarjev v analizi omrežij je nakazana le teoretično. To nameravamo premostiti z vključitvijo polkolobarjev v nadaljne analize omrežij.

Ostalo je še nekaj odprtih vprašanj pri razvoju metode posplošenih dvovrstnih sredic. Izboljšati nameravamo učinkovitost algoritma in eksperimentalno analizirati zahtevnost algoritma na naključnih dvovrstnih omrežjih. Algoritem želim razširiti v iskanje vseh neničelnih različnih posplošenih dvovrstnih sredic danega dvovrstnega omrežja, torej v algoritem za nepribita praga, in v algoritem za določanje mejnih parov vrednosti pragov. Posvetili se bomo tudi razširitvi funkcij lastnosti vozlišč v imenske funkcije lastnosti in v vektorje funkcij lastnosti.

Algoritma za množenje omrežij z uporabo različnih polkolobarjev in posplošene dvovrstne sredice sta v programskem jeziku C++ sprogramirana le v osnovni različici, kar bomo še nadgradili. Ta nadgradnja vsebuje uporabo vseh navedenih polkolobarjev pri množenju omrežij. Pri posplošenih dvovrstnih sredicah bomo dopolnili seznam dovoljenih lastnosti vozlišč in algoritem za dva fiksna praga posplošili v algoritem za samo en pribit prag.

Lotili se bomo tudi razvrščanja dvovrstnih omrežij, saj pričakujemo, da nam bo uspelo postopke razvrščanja z relacijsko omejitvijo razširiti na dvovrstna omrežja. Možnost hierarhičnega razvrščanja v dvovrstnih omrežjih bo mogočila nov pogled na razvrščanje enot, saj bomo takšno razvrščanje lahko uporabili na dveh množicah enot, ki sta med sabo povezani z dano relacijo.

Primera analiz omrežij, ki sta opisana v poglavju 3, bomo razširili s časovno komponento - poleg omrežij bomo uporabili razbitje del glede na leto objave. Analizo bibliografskih podatkov iz baze Zentralblatt MATH bomo dopolnili z analizami več izpeljanih omrežij, ki jih bomo pridobili iz obstoječih s kombinacijo množenja omrežij in različnih normalizacij.

Menimo, da oba pogleda na analizo omrežij, ki ju obravnavamo v tej disertaciji, hodita z roko v roki – metode niso uporabne brez realnih podatkov, analize realnih podatkov ne morem izvesti brez primernih metod. Zato je pomembno, da se poleg razvoja metod posvečamo tudi analizi realnih podatkov. Področje analize dvovrstnih omrežij pa je zelo perspektivno, saj še vedno ostaja veliko odprtih možnosti za razvoj metod, ki omogočajo neposredno analizo dvovrstnih omrežij.

Članki z največ avtorji

A

V razdelku 3.2.2 je v tabeli 3.2 navedena porazdelitev del glede na število avtorjev v podatkih SN5. V tem dodatku so navedeni podatki o sedmih delih z največ avtorji.

[1] Allen, J., Anton, R. F., Babor, T. F., Carbonari, J., Carroll, K. M., Connors, G. J., Cooney, N. L., Del Boca, F. K., Di-Clemente, C. C., Donovan, D., Kadden, R. M., Litt, M., Longabaugh, R., Mattson, M., Miller, W. R., Randall, C. L., Rounsaville, B. J., Rychtarik, R. G., Stout, R. L., Tonigan, J. S., Wirtz, P. W., & Zweben, A. (1998). Matching Alcoholism Treatments to Client Heterogeneity: Project Match Three-Year Drinking Outcomes. *Alcoholism-Clinical and Experimental Research, 22(6),* 1300–1311.

[2] Doll, L.S., Petersen, L.R., White, C.R., Johnson, E.S., Ward, J.W., Williams, A., Altman, R., Becker, G., Bernarducci, J., Busch, M., Clary, N., Davis, J., Darr, F., Grindon, A., Kleinman, S., Lamberson, H., Lenes, B., Menitove, J., Molinaris, J., Ness, P., Raevsky, C., Holland, P., Shafer, A.W., Sherwood, W., Stevens, C., Vaughan, H., Arnold, E., Donovan, D.E., Kessler, D., Harper, M., Hernandez, J., Ksell, T., Mcelfresh, S., Myers, M.K., Montgomery, A., Nason, M., Sanchez, J., Schulze, G., Shahan, M., Stephenson, S., Theobald, J., & Wilke, D. (1992). Homosexually and nonhomosexually identified men who have sex with men – a behavioral comparison. *Journal of Sex Research, 29(1)*, 1–14.

[3] Kelly, J. A., Murphy, D. A., Sikkema, K. J., McAuliffe, T. L., Roffman, R. A., Solomon, L. J., Winett, R. A., Kalichman, S. C., Heckman, T. G., Perry, M. J., Stevenson, L. Y., Hauth, A. C., Koob, J. J., Morgan, M. G., Norman, A., Lemke, A., Steiner, S., Trenary, B., Flynn, B., Ayotte, D. R., Desiderato, L. L., Lombard, D. N., & Yaffe, D. M. (1997). Randomised, controlled, community-level HIV-prevention intervention for sexual-risk behaviour among homosexual men in US cities. *Lancet, 350(9090)*, 1500–1505. doi: 10.1016/S0140-6736(97)07439-4.

[4] Magliano, L., Fiorillo, A., Malangone, C., De Rosa, C., Favata, G., Sasso, A., Prezioso, M., Pezzenati, L., Gentile, F., Casale, L., Bondi, E., Test, G. R., Di Lella, M., Biscussi, E., Degl'Innocent, F., Bellini, R., Di Nunzio, R., Matrella, L., Salmeri, R., Cantone, R., Gargiulo, L., Esposito, A., Delcuratolo, V., Giannini, M., Maresca, L., Cavaliere, G., Scandone, B., Folla, M., Raffaeli, M., Innocente, P., Dagianti, F., Lucania, S., Scorsino, A., Bardicchia, F., Cerullo, G., Curreli, R., Miscali, S., Scordato, M., Campo, G., Maj, M.; & Gruppo

Di Lavoro (2006). Family psychoeducational interventions for schizophrenia in routine settings: impact on patients' clinical status and social functioning and on relatives' burden and resources. *Epidemiologia e Psichiatria Sociale, 15(3),* 219–27.

[5] Pierce, J. P., Natarajan, L., Caan, B. J., Parker, B. A., Greenberg, E. R., Flatt, S. W., Rock, C. L., Kealey, S., Al-Delaimy, W. K., Bardwell, W. A., Carlson, R. W., Emond, J. A., Faerber, S., Gold, E. B., Hajek, R. A., Hollenbach, K., Jones, L. A., Karanja, N., Madlensky, L., Marshall, J., Newman, V. A., Ritenbaugh, C., Thomson, C. A., Wasserman, L., & Stefanick, M. L. (2007). Influence of a Diet Very High in Vegetables, Fruit, and Fiber and Low in Fat on Prognosis Following Treatment for Breast Cancer. *JAMA, 298(3),* 289–298. doi: 10.1001/jama.298.3.289.

[6] Semple, S. J., Patterson, T. L., Temoshok, L. R., McCutchan, J. A., Straits-Troster, K. A., Chandler, J. L., Grant, I., Atkinson, J. H., Velin, R. A., Oldfield, E. C., Wallace, M.R., Malone, J., Spector, S. A., Thal, L., Heaton, R. K., Hesselink, J., Jernigan, T., Wiley, C. A., Olshen, R., Abramson, I., Butters, N., Dupont, R., Zisook, S., Jeste, D., Sieburg, H., & Weinrich, J. D. (1993). Identification of psychological stressors among HIV-positive women. *Women & Health, 20(4),* 15–36. doi: 10.1300/J013v20n04_02

[7] Snijders, T. A. B., Robinson, T., Atkinson, A. C., Riani, M., Gormley, I. C., Murphy, T. B., Sweeting, T., Leslie, D. S., Longford, N. T., Kent, J. T., Lawrance, T., Airoldi, E. M., Besag, J., Blei, D., Fienberg, S. E., Breiger, R., Butts, C. T., Doreian, P., Batagelj, V., Ferligoj, A., Draper, D., van Duijn, M. A. J., Faust, K., Petrescu-Prahova, M., Forster, J. J., Gelman, A., Goodreau, S. M., Greenwood, P. E., Gruenberg, K., Francis, B., Hennig, C., Hoff, P. D., Hunter, D. R., Husmeier, D., Glasbey, C., Krackhardt, D., Kuha, J., Skrondal, A., Lawson, A., Liao, T. F., Mendes, B., Reinert, G., Richardson, S., Lewin, A., Titterington, D. M., Wasserman, S., Werhli, A. V., & Ghazal, P. (2007). Discussion on the paper by Handcock, Raftery and Tantrum. *Journal of the Royal Statistical Society: Series A – Statistics in Society, 170(2),* 322–354. doi: 10.1111/j.1467-985X.2007.00471.x.

LITERATURA

[1] V. Batagelj. *Analiza omrežij. 7. Aciklična in dvovrstna omrežja*. Ljubljana, december 2006. Podiplomski študij statistike UL.

[2] S. K. Abdali in B. D. Saunders. Transitive closure and related semiring properties via eliminants. *Theoretical Computer Science*, 40:257–274, 1985.

[3] F. Baccelli, G. Cohen, G. J. Olsder in J.-P. Quadrat. *Synchronization and linearity. An algebra for discrete event systems*. John Wiley & Sons Ltd, West Sussex, Velika Britanija, 1992. Elektronska izdaja 2001.

[4] J. S. Baras in G. Theodorakopoulos. *Path Problems in Networks*. Morgan & Claypool, Berkeley, 2010.

[5] V. Batagelj. Semirings for social networks analysis. *The Journal of Mathematical Sociology*, 19(1):53–68, 1994.

[6] V. Batagelj. Analiza omrežij. 6. Matrike in omrežja, november 2006. Podiplomski študij statistike UL.

[7] V. Batagelj in A. Mrvar. Analysis of kinship relations with Pajek. *Social Science Computer Review*, 26(2): 224–246, 2008.

[8] P. F. Felzenszwalb in J. J. McAuley. Fast inference with Min-Sum matrix product. *IEEE Transactions on pattern analysis and machine intelligence*, 33(12): 2549–2554, 2011.

[9] S. Gaubert in Max Plus Group. Methods and applications of $(\max, +)$ linear algebra. V *14th Symposium on Theoretical Aspects of Computer Science*, Hansestadt Luebeck, Nemčija, februar – marec 1997.

[10] J. S. Golan. Some recent applications of semiring theory. V *International Conference on Algebra in Memory of Kostia Beidar*, National Cheng Kung University, Tajvan, marec 2005.

[11] M. Gondran in M. Minoux. *Graph, dioids and semirings*. Springer, New York, ZDA, 2008.

[12] M. Ojsteršek. Spreminjanje označenega omrežja skozi čas – primer: odnosi v študentskem naselju. Diplomsko delo, Fakulteta za družbene vede, UL, Ljubljana, 2010. mentor: A. Mrvar.

[13] R. W. Schvaneveldt, D. W. Dearholt in F. T. Durso. Graph theoretic foundations of pathfinder networks. *Comput. Math. Applic.*, 15(4):337–345, 1988.

[14] U. Zimmermann. *Annals of Discrete Mathematics: Linear and Combinatorial Optimization in Ordered Algebraic Structures*. Mathematisches Institut, Universitaet zu Koeln, Nemčija, 1981.

[15] J. Zumbraegel. Semirings. Spletna stran. http://shannoninstitute.ucd.ie/~jzumbr/semirings.html.

[16] V. Batagelj. *Network Analysis: Two mode Networks*. Fakulteta za družbene vede, Univerza v Ljubljani, Ljubljana, avgust 2006. ECPR Summer School.

[17] A. L. Barabási, H. Jeong, Z. Néda, E. Ravasi, A. Schubert in T. Vicsek. Evolution of the social network of scientific collaborations. *Physica A.*, 331(3–4): 590–614, 2002.

[18] V. Batagelj. Efficient algorithms for citation network analysis. *IMFM Preprint Series*, 41(897):1–29, 2003.

[19] J. Ferlež, C. Faloutsos, J. Leskovec, D. Mladenić in M. Grobelnik. Monitoring network evolution using MDL. V *IEEE International Conference on Data Engineering (ICDE)*, Cancun, Mehika, 2008.

[20] N. Kejžar, S. Korenjak Černe in V. Batagelj. Network analysis of works on clustering and classification from Web of Science. V H. Locarek-Junge in C. Weihs, editors, *Proceedings of the 11th IFCS Biennial Conference and 33rd Annual Conference of the Gesellschaft fuer Klassifikation e.V.*, str. 525–536. Springer, 2010.

[21] M. E. J. Newman. The structure of scientific collaboration networks. *Proc. of National Academy of Science USA*, 98(2):404–409, 2001.

[22] M. Seshadri, S. Machiraju, A. Sridharan, J. Bolot, C. Faloutsos in J. Leskovec. Mobile call graphs: Beyond power-law and lognormal distributions. V *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, Las Vegas, ZDA, 2008.

[23] X. Shi, J. Leskovec in D. A. McFarland. Citing for high impact. V *Joint Conference on Digital Libraries (JCDL)*, Surfer's Paradise, Avstralija, 2010.

[24] V. Batagelj. WoS2Pajek – program za pretvorbo datotek WoS v omrežja, 9 2013. http://pajek.imfm.si/doku.php?id=wos2pajek.

[25] S. B. Seidman. Network structure and minimum degree. *Social Networks*, 5:269–287, 1983.

[26] V. Batagelj in M. Zaveršnik. Fast algorithms for determining (generalized) core groups in social networks. *Advances in Data Analysis and Classification*, 5(2): 129–145, 2011.

[27] A. Ahmed, V. Batagelj, X. Fu, S.-H. Hong, D. Merrick in A. Mrvar. Visualisation and analysis of the internet movie database. V *APVIS*, Sydney, Avstralija, 2007.

[28] V. Batagelj in A. Mrvar. *Pajek and Pajek-XXL – Program for analysis and visualization of large networks*, marec 2014. http://mrvar.fdv.uni-lj.si/pajek/pajekman.pdf.

[29] V. Batagelj in A. Mrvar. Pajek wiki, 3 2014. URL http://pajek.imfm.si. http://pajek.imfm.si, verzija 2.05.

[30] V. Batagelj, A. Mrvar in W. de Nooy. *Exploratory Social Network Analysis with Pajek*. Cambridge University Press, 2 edition, 2012.

[31] D. Lehmann. Algebraic structures for transitive closure. *Theoretical Computer Science*, 4(1):59–76, 1977.

[32] R. Melhem. Parallel solution of linear systems with striped sparse matrices. *Parallel Computing*, 6(2): 165–184, 1988.

[33] R. Yuster in U. Zwick. Fast sparse matrix multiplication. *ACM Transactions on Algorithms*, 1(1):2–13, 2005.

[34] R. Pagh in M. Stöckel. The input/output complexity of sparse matrix multiplication. *ArXiv e-prints*, 2014.

[35] V. Batagelj in M. Cerinšek. On bibliographic networks. *Scientometrics*, 96(3):845–864, 2013.

[36] J. G. Fletcher. A more general algorithm for computing closed semiring costs between vertices of a directed graph. *Communications of the ACM*, 23(6): 350–351, 1980.

[37] B. Carré. *Graphs and Networks*. Clarendon Press, 1979.

[38] Algebraic and combinatorial methods in operations research. V R. E. Burkard, R. A. Cuninghame-Green in U. Zimmermann, editors, *Annals of Discrete Mathematics*, zvezek 19. Elsevier, 1984.

[39] J. Kepner in J. Gilbert. *Graph algorithms in the language of linear algebra*. SIAM, 2011.

[40] A. Quirin, O. Cordón, J. Santamaria, B. Vargas-Quesada in F. Moya-Anegón. A new variant of the pathfinder algorithm to generate large visual science maps in cubic time. *Information Processing & Management*, 44(4):1611–1623, 2008.

[41] S. Wasserman in K. Faust. *Social network analysis: methods and applications*. Cambridge University Press, 1994.

[42] P. V. Marsden. *The SAGE Handbook of Social Network Analysis*, chapter Survey methods for network data. Sage Publications, 2011.

[43] L. A. Goodman. Snowball sampling. *Annals of mathematical statistics*, 32(1):148–170, 1961.

[44] T. Martin, B. Karrer in M. E. J. Newman. Co-authorship and citation in scientific publishing. http://arxiv.org/abs/1304.0473v1, 2013.

[45] K. Lozar Mafreda, V. Vehovar in V. Hlebec. Collecting ego-centred network data via the web. *Metodološki zvezki*, 1(2):295–321, 2004.

[46] S. P. Borgatti in J. L. Molina. Ethical and strategic issues in organizational network analysis. *Journal of applied behavioral science*, 39(3):337–350, 2003.

[47] R. Enyon, I. Fry in R. Schroeder. *The SAGE Handbook of Online Research Methods*, chapter The ethics of internet research. Sage Publications, 2008.

[48] A. Charlesworth. *The SAGE Handbook of Online Research Methods*, chapter Understanding and managing legal issues in internet research. Sage Publications, 2008.

[49] R. L. Breiger. Ethical dilemmas in social network research: Introduction to special issue. *Social Networks*, 27(2):88–93, 2005.

[50] J. C. Mitchell. *Social networks in urban situations*, chapter The concept and use of social networks. Manchester University Press, 1969.

[51] J. Shipman, J. D. Wilson in A. Todd. *Introduction to physical science*. Cengage Learning, 12 edition, 2009.

[52] S. F. Sampson. *A novitiate in a period of change. An experimental and case study of social relationships*. PhD thesis, Cornell University, 1968.

[53] P. V. Marsden. Network data and measurement. *Annual review of sociology*, 16:435–463, 1990.

[54] M. S. Mizruchi in J. Galaskiewicz. Networks of interorganizational relations. *Sociological methods & research*, 22(1):46–70, 1993.

[55] L. Schmidt. *Using archives. A guide to effective research*. Society of American Archivists, 2011.

[56] J. Ullman in J. Widom. *First course in database systems*. Prentice-Hall Inc., 3 edition, 2008.

[57] W. Voorsluys, J. Broberg in R. Buyya. *Cloud computing: Principles and Paradigms*, chapter Introduction to cloud computing. Wiley Press, 2011.

[58] R. Angles in C. Gutierrez. Survey of graph database models. *ACM Computing Surveys*, 40(1):1–39, 2008.

[59] T. White. *Hadoop: The definite guide*. O'Reilly Media, 3 edition, 2012.

[60] F. B. Abdesslem, I. Parris in T. Henderson. *Computational Social Networks*, chapter Reliable Online Social Network Data Collection. Springer, 2012.

[61] R. Popping. *Computer-assisted text analysis*. Sage Publications, 2000.

[62] R. Franzosi. *From words to numbers: Narrative, data, and social science*. Cambridge University Press, 2004.

[63] T. Berners-Lee, J. Hendler in O. Lassila. The semantic web. *Scientific American*, 284(5):28–37, 2001.

[64] V. Batagelj in U. Brandes. Efficient generation of large random networks. *Physical Review E*, 71(3): 036113, 2005.

[65] Random graphs and complex networks, 2013. URL http://www.win.tue.nl/~rhofstad/NotesRGCN.pdf.

[66] P. Erdös in A. Rényi. On random graphs. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.

[67] E. N. Gilbert. Random graphs. *Annals of mathematical statistics*, 30:1141–1144, 1959.

[68] D. J. Watts in S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

[69] A. L. Barabási in R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[70] N. Kejžar, Z. Nikolovski in V. Batagelj. Probabilistic inductive classes of graphs. *The Journal of Mathematical Sociology*, 32(2):85–109, 2008.

[71] M. Zaveršnik in V. Batagelj. Islands. V *XXIV. International Sunbelt Social Network Conference*, 2004.

[72] M. E. J. Newman. Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality. *Physical Review E*, 64(1):0161321–0161327, 2001.

[73] G. Salton in M. J. McGill. *Introduction to modern Information Retrieval*. McGraw-Hill, 1987.

[74] M. M. Kessler. Bibliographic coupling between scientific papers. *American Documentation*, 14(1):10–25, 1963.

[75] K. E. Rosengren. *Sociological Aspects of the Literary System*. Natur och Kultur, 1968.

[76] H. G. Small. Co-citation in the scientific literature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24(4):265–269, 2011.

[77] W. Quattrociocchi in W. Amblard. Selection in scientific networks. *Social Network Analysis and Mining*, 2 (3):229–237, 2011.

[78] V. Batagelj. Social network analysis, large-scale. V *Encyclopedia of complexity and systems science*, str. 8245–8265. Springer, 2009.

[79] E. Garfield. *Citation Indexing: Its Theory and Application in Science, Technology, and Humanities*. Wiley Press, New York, ZDA, 1979.

[80] B. Wegner in D. Werner. Zentralblatt math, mathematics subject classification 2010, 2010. URL http://www.mathem.pub.ro/dept/MSC-2010_ZBL.pdf.

[81] P. TePaske King in N. Richert. Database reviews and reports. the identification of authors in the mathematical reviews database, 2001. URL http://www.istl.org/01-summer/databases.html.

[82] V. Batagelj, P. Dorean, A. Ferligoj in N. Kejžar. *Understanding Large Temporal Networks and Spatial Networks: Exploration, Pattern Searching, Visualization and Network Evolution*. Wiley Press, New York, ZDA, 2014.

[83] A. Clauset, C. R. Shalizi in M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51 (4):661–703, 2009.

[84] J. F. Grcar. Topical bias in generalist mathematics journals. *Notices of the AMS*, 57(11):1421–1424, 2010.

[85] S. Robertson. Understanding inverse document frequency: On theoretical arguments for IDF. *Journal of Documentation*, 60(5):503–520, 2004.

[86] S. P. Borgatti in D. Halgin. *Analyzing affiliation networks*. Sage Publications, Los Angeles, ZDA, 2011.

[87] M. Latapy, C. Magnien in N. Del Vecchio. Basic notions for the analysis of large two-mode networks. *Social Networks*, 30(1):31–48, 2008.

[88] F. Agneesseens in M. G. Everett. Special issue on advances in two-mode social networks. *Social Networks*, 35(2):145–278, 2013.

[89] Encyclopedia of complexity in systems science. *Social network analysis, large-scale*. Springer, New York, ZDA, 2009.

[90] J. Scott in P. J. Carrington, editors. *The SAGE Handbook of Social Network Analysis*. SAGE Publications, Los Angeles, ZDA, 2011.

[91] G. Xu, Y. Zhang in L. Li. *Web mining and social networking: Techniques and applications*. Springer, New York, 2011.

[92] M. Sozio in A. Gionis. The community-search problem and how to plan a successful cocktail party. V *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, str. 939–948, 2010.

[93] D. Melamed. Community structures in bipartite networks: A dual-projection approach. *PLoS ONE*, 9, 2014.

[94] M. J. Barber. Modularity and community detection in bipartite networks. *Physical review. E*, 76:066102, 2007.

[95] L. Xin in T. Murata. Community detection in large-scale bipartite networks. V *IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technologies*, str. 50–57, 2009.

[96] D. B. Larremore, A. Clauset in A. Z. Jacobs. Efficiently inferring community structure in bipartite networks. *Physical Review E*, 90:012805, 2014.

[97] D. Baier in K.-D. Wernecke. An exchange algorithm for two-mode cluster analysis. V *Innovations in Classification, Data Science, and Information Systems: Proceedings of the 27th Annual Conference of the Gesellschaft für Klassifikation e.V.*, str. 62–68, 2003.

[98] I. Van Mechelen, H.-H. Bock in P. De Boeck. Two-mode clustering methods: a structured overview. *Statistical Methods and Medical Research*, 13:363–394, 2004.

[99] T. H. Cormen, C. E. Leierson, R. L. Rivest in C. Stein. *Introduction to Algorithms*. MIT Press and McGraw-Hill, 3 edition, 2009.

[100] M. Cerinšek in V. Batagelj. Semirings and matrix analysis of networks. V R. Alhajj in J. Rokne, editors, *Encyclopedia of Social Network Analysis and Mining*, zvezek 3, str. 1681–1687. Springer, 2014.

[101] M. Cerinšek in V. Batagelj. Sources of network data. V R. Alhajj in J. Rokne, editors, *Encyclopedia of Social Network Analysis and Mining*, zvezek 3, str. 1946–1954. Springer, 2014.

[102] M. Cerinšek in V. Batagelj. Network analysis of Zentralblatt MATH data. *Scientometrics*, 102(1): 977–1001, 2015.

[103] M. Cerinšek in V. Batagelj. Generalized two-mode cores. V *Eurogiga Midterm Conference*, 2012.

STVARNO KAZALO