

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Matjaž Hribernik

**Napovedovanje izidov tekem badmintona na osnovi
analize omrežij**

DIPLOMSKO DELO

UNIVERZITETNI ŠTUDIJSKI PROGRAM PRVE STOPNJE
RAČUNALNIŠTVO IN INFORMATIKA

Ljubljana, 2014

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Matjaž Hribernik

**Napovedovanje izidov tekem badmintona na osnovi
analize omrežij**

DIPLOMSKO DELO

UNIVERZITETNI ŠTUDIJSKI PROGRAM PRVE STOPNJE
RAČUNALNIŠTVO IN INFORMATIKA

MENTOR:izr. prof. dr. Marko Bajec

Ljubljana, 2014

Rezultati diplomskega dela so intelektualna lastnina avtorja in Fakultete za računalništvo in informatiko Univerze v Ljubljani. Za objavljanje ali izkoriščanje rezultatov diplomskega dela je potrebno pisno soglasje avtorja, Fakultete za računalništvo in informatiko ter mentorja.



Št. naloge: 00140 / 2013
Datum: 5.4.2013

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko izdaja naslednjo nalogo:

Kandidat: **MATJAŽ HRIBERNIK**

Naslov: **NAPOVEDOVANJE IZIDOV TEKEM BADMINTONA NA OSNOVI
ANALIZE OMREŽIJ
PREDICTING THE OUTCOME OF BADMINTON MATCHES BASED ON
NETWORK ANALYSIS**

Vrsta naloge: Diplomsko delo univerzitetnega študija prve stopnje

Tematika naloge:

Napovedovanje rezultatov in dosežkov tekmovalcev v različnih športih je med privrženci zelo pogosto. Kandidat naj v diplomskem delu preuči, kako uspešno je moč napovedati razvrstitev igralcev badmintona na osnovi izidov odigranih tekem v Sloveniji v preteklih letih. Pri tem naj podatke predstavi v obliki omrežja ter preizkusi izbrane metode in tehnike analize omrežij. Osredotoči naj se predvsem na mere središčnosti vozlišč ter algoritme analize povezav. Kandidat naj predstavi uspešnost napovedovanja in na podlagi pridobljenega znanja predlaga možne izboljšave.

Mentor:
izr. prof. dr. Marko Bajec



Dekan:
prof. dr. Nikolaj Zimic

IZJAVA O AVTORSTVU
diplomskega dela

Spodaj podpisani Matjaž Hribernik,

z vpisno številko 63070193

sem avtor diplomskega dela z naslovom:

Napovedovanje izidov tekem badmintona na osnovi analize omrežij

S svojim podpisom zagotavljam, da:

- sem diplomsko delo izdelal samostojno pod mentorstvom izr. prof. dr. Marka Bajca
- so elektronska oblika diplomskega dela, naslov (slov., angl.), povzetek (slov., angl.) ter ključne besede (slov., angl.) identični s tiskano obliko diplomskega dela
- soglašam z javno objavo elektronske oblike diplomskega dela v zbirki »Dela FRI«.

V Ljubljani, dne 13.3.2014

Podpis avtorja:

Zahvalil bi se mentorju izr. prof. dr. Marku Bajcu za pomoč in usmerjanje pri izdelavi tega diplomskega dela. Posebej bi se zahvalil tudi as. dr. Lovru Šublju za njegov trud, ideje in koristne nasvete tekom izdelave.

Ne nazadnje pa gre še posebna zahvala tudi mojim staršem, bratu in starim staršem za motivacijo ter podporo ne samo pri izdelavi diplomske naloge, ampak tekom celotnega šolanja.

Kazalo

Povzetek	
Abstract	
1. Uvod	1
2. Uporabljena orodja in tehnologije	3
2.1. Razvojno okolje Eclipse	3
2.2. JUNG	3
2.3. Orange Canvas	4
2.4. Java	5
3. Badminton	7
3.1. Osnovno o badmintonu	7
3.1.1. Badminton v Sloveniji	7
3.2. Tekmovalni sistem	8
3.2.1. Jakostna lestvica A in B kategorije	9
3.3. Pridobivanje podatkov	10
4. Teorija omrežij in podatkovno rudarjenje	13
4.1. Omrežje	13
4.2. Lastnosti omrežij	14
4.2.1. Mere središčnosti	14
4.3. Podatkovno rudarjenje	17
4.3.1. Načini podatkovnega rudarjenja	17
4.3.2. Regresija	18
5. Analiza badmintonskega omrežja in napoved rezultata	23
5.1. Analiza turnirja A1-2012-2013	23
5.2. Ocena uspešnosti atributov na celotni podatkovni množici s pomočjo programa Orange ..	27
5.2.1. Napoved doseženega mesta	29
5.2.2. Napoved točk	31
5.2.3. Napoved povprečnega števila točk igralca na posamezen turnir	33
6. Primerjava rezultatov in zaključek	35
Tabele	37
Slike	39
Priloge	41

Seznam uporabljenih kratic in simbolov

Oznaka	Opis
<i>BZS</i>	Badmintonska zveza Slovenije
<i>KO</i>	Knock-out, tekmovanje na izpadanje
<i>SJLP</i>	Slovenska jakostna lestvica posameznikov
<i>G</i>	Graf
<i>N</i>	Omrežje
<i>V</i>	Množica vozlišč
<i>E</i>	Množica povezav
<i>n</i>	Število vozlišč
v_i	i-to vozlišče
<i>kNN</i>	Metoda k-najbližjih sosedov
<i>MSE</i>	Srednja kvadratna napaka
<i>RSE</i>	Relativna kvadratna napaka
<i>MAE</i>	Srednja absolutna napaka
<i>RAE</i>	Relativna absolutna napaka

Povzetek

S širjenjem svetovnega spleta in hitrejšim internetnim dostopom se spreminja tudi njegova namembnost. Spletne strani vsebujejo vse več podatkov, ki jih lahko pridobimo in obdelujemo. Takšne so tudi strani namenjene hranjenju rezultatov v različnih športih, saj je ukvarjanje s športno statistiko zelo razširjeno tako med navijači, novinarji kot tudi med športnimi analitiki.

Cilj diplomskega dela je napoved uvrstitve in doseženih točk badmintonskega igralca v Sloveniji na navidezni lestvici, ki temelji na dosežkih iz turnirjev za neko obdobje. Napoved je podana na podlagi atributov, ki se pridobijo z analizo omrežja vseh odigranih tekem v tem obdobju.

Rezultati so na voljo na naslovu www.tournamentsoftware.com, s katerega je potrebno s pomočjo programskega jezika Java in namenske knjižnice JUNG prenesti rezultate ter jih pretvoriti v obliko za analizo omrežij. S pomočjo analize omrežja se pridobijo lastnosti, ki jih kasneje uporabimo kot attribute za gradnjo napovednega modela.

Dobljeni rezultati so osnova za izvajanje statističnih metod in metod podatkovnega rudarjenja ter strojnega učenja v programu Orange.

Končni rezultat naloge je napovedni model, ki lahko služi za napovedovanje rezultatov v badmintonu na splošno, ali pa z določenimi popravki v katerem drugem športu, ki uporablja podoben način tekmovanja kot badminton. Zaradi premajhnega vzorca rednih igralcev badmintona v Sloveniji težko ocenimo, kako bi se predstavljeni napovedni model obnesel v realnih primerih.

Ključne besede: omrežje, graf, mere središčnosti, PageRank, statistika, podatkovno rudarjenje, analiza omrežij, napovedovanje rezultatov

Abstract

The expansion of World Wide Web and faster access to the Internet is changing its primary use. Websites contain a lot more information that can be acquired and processed. Such sites are also the ones that provide results for various competitions in sport. Dealing with sport statistics is well spread among fans and journalists, as well as sport analysts.

The purpose of this thesis is the prediction of rank and total points scored by badminton players in Slovenia on a virtual ranking that is based on tournament results from a certain time window. Prediction is based on attributes that are obtained with the analysis of a network representing all games played in this period.

The analysis is based on the data available through www.tournamentsoftware.com. The acquired data has first been processed and represented in the form of a network using Java programming language and JUNG network analysis library. The necessary attributes were obtained using the techniques from network analysis and then adopted to learn a prediction model.

The mentioned attributes were the basis for running various statistical, data mining and machine learning methods in Orange.

The main contribution of the thesis is a model that can be used for predicting results and standings in badminton in general or in any other sport that runs competitions or tournaments in a similar manner as badminton (with a certain adjustments). Due to a small sample size (the number of regular players in Slovenia), the performance of the model in practical applications remains unclear.

Keywords: network, graph, centrality measures, PageRank, statistics, data mining, network analysis, results prediction

Poglavje 1

1. Uvod

S širjenjem svetovnega spleta in hitrejšim internetnim dostopom se spreminja tudi njegova namembnost. Oblikovati so se začela družabna omrežja, kot sta Facebook in Twitter, druga socialna (lahko tudi družbena) omrežja ter ostale spletne strani, ki hranijo veliko količino podatkov. Takšne so tudi strani namenjene hranjenju rezultatov v različnih športih, saj je ukvarjanje s športno statistiko zelo razširjeno tako med navijači, novinarji kot tudi med športnimi analitiki. Tako so se začele širiti zbirke podatkov, ki jih lahko pridobimo s spleta. Te, poleg osnovne informacije o posameznih entitetah, hranijo tudi različne vrste relacij in odnosov med njimi. Relacije so lahko prijateljstva med ljudmi v socialnem omrežju ali pa, kot v našem primeru, odigrane tekme med tekmovalci badmintona. Prav s preučevanjem relacij med entitetami se ukvarja analiza omrežij.

Omrežja so lahko majhna z nekaj entitetami in relacijami med njimi, lahko pa so kompleksnejša z nekaj milijoni entitet in relacij. Analiza takšnih omrežij se zato razlikuje med seboj, kljub temu pa je vsem omrežjem enako, da so v osnovi sestavljena iz množice vozlišč in množice povezav med njimi. Vozlišča ponazarjajo entitete, povezave pa relacije med njimi.

Poznamo več vrst omrežij, trenutno najbolj razširjena in preiskovana so socialna omrežja (angl. *social network analysis*), raziskovalci pa se ukvarjajo tudi z biološkimi omrežji, informacijskimi omrežji, tehnološkimi omrežji, itd.

Analiza omrežij je široko področje, ki je močno povezano tudi z drugimi področji, kot so teorija grafov (angl. *graph theory*), podatkovno rudarjenje (angl. *data mining*), strojno učenje (angl. *machine learning*), statistika (angl. *statistics*) in še nekatera druga področja.

V nalogi se zgledujemo po članku *Who Is The Best Player Ever? A Complex Network Analysis of the History of Professional Tennis* [1], v katerem avtor igralce rangira glede na podatke iz turnirjev od leta 1968 do leta 2010 in jih primerja z že obstoječimi sistemi na tem področju. Namesto teniških profesionalcev so v nalogi zbrana tekmovalca slovenskih badmintonskih igralcev, ki ponazarjajo mnogo manjšo množico. V članku *Napovedovanje izidov teniških dvobojev s pomočjo omrežja* [2] avtorji v algoritmu PageRank dodajo tudi časovno komponento, na katero se sklicujemo večkrat tekom naloge.

Omrežje badmintonskih igralcev, ki so med septembrom 2010 in marcem 2014 na turnirjih v sklopu organizacije Badmintonske zveze Slovenije, odigrali vsaj eno tekmo, je nekaj manj kot 150, zato takšno omrežje spada med manjša omrežja. V nalogi preučimo, kako stopnja omrežja in lastnosti središčnosti vplivajo na končno lestvico tekmovalcev. S pomočjo podatkovnega rudarjenja in programa Orange Canvas pa preverimo, kako uspešno se da na podanih atributih napovedati mesto igralca in število točk, ki jih bo dosegel ter kako na napoved vplivajo lastnosti omrežja, ki ga preučujemo. Nato poskusimo razumeti, zakaj je prišlo do določenih rezultatov in kako se da napoved še izboljšati.

V poglavju 2 sledi pregled orodij in tehnologij uporabljenih v tej nalogi. V poglavju 3 je nekaj splošnih značilnosti badmintona v Sloveniji, način tekmovanja in točkovanje ter razlaga, na kakšen način pridemo do želenih podatkov za analizo. Poglavje 4 sestoji iz razlage teorije grafov in omrežij ter njihovih lastnosti, predvsem tistih, ki se nanašajo na to nalogo. Drugi del poglavja 4 je opis podatkovnega rudarjenja in metod za analizo uspešnosti napovednih modelov. Poglavje 5 je glavni del naloge, kjer najprej analiziramo majhno omrežje, ki je sestavljeno iz enega turnirja, nato pa ugotovitve uporabimo na večjem omrežju, ki ga sestavlja celotna podatkovna množica. Nato sledi opis urejanja podatkov, ki so primerni za poskus napovedi in na koncu še sama napoved s pomočjo ustreznih orodij. V poglavju 6 podamo ugotovitve na podlagi rezultatov in razmislimo o ustreznih popravkih ter možnostih o nadaljnji raziskavi.

Poglavje 2

2. Uporabljena orodja in tehnologije

V tem poglavju na kratko opišemo orodja in tehnologije, ki smo jih uporabili za analizo badmintonskega omrežja.

2.1. Razvojno okolje Eclipse

Eclipse je integrirano razvojno okolje (angl. *integrated development environment, IDE*) za razvoj programske opreme predvsem v programskem jeziku Java, s pomočjo programskih dodatkov pa tudi v drugih jezikih kot so Ada, C, C++, COBOL, Fortran, Haskell, JavaScript, Lasso, Perl, PHP, Python, R, Ruby (vključno z Ruby on Rails framework), Scala, Clojure, Groovy, Scheme, and Erlang.

Osnovna koda izvira iz IBM VisualAge. Eclipse software development kit (SDK) vsebuje razvojna orodja za programiranje v Javi. Uporabniki lahko prilagodijo namembnost programa z namestitvijo programskih dodatkov ali vtičnikov (angl. *plug-ins*), ki so bili napisani za Eclipse. Ti dodatki so lahko razvojna orodja za druge programske jezike, uporabniki pa lahko po potrebi napišejo tudi svoje dodatke.

Eclipse SDK je prosto dostopno odprtokodno razvojno orodje, izdano pod pogoji Eclipse Public Licence. Od leta 2006 vsako leto izide nova različica orodja Eclipse, vsaka s svojim poimenovanjem (Luna, Kepler, Juno, ...) [3].

V nalogi se Eclipse (različica Juno) uporablja za dostop do podatkov o turnirjih s spletnega mesta. Te podatke je potrebno prilagoditi za uporabo v analizi omrežij. Analiza se izvaja s pomočjo knjižnice JUNG [4].

2.2. JUNG

JUNG (Java Universal Network/Graph Framework) je programska knjižnjica za programski jezik Java. Omogoča razširitev jezika za modeliranje, analizo in prikaz podatkov, ki jih lahko predstavimo s pomočjo grafov ali omrežij.

Arhitektura omogoča, da lahko podatke predstavimo z usmerjenimi in neusmerjenimi grafi, večvrstnimi grafi, grafi s paralelnimi povezavami in hipergrafi. Podatkom lahko dodamo tudi meta podatke, s katerimi omogočimo več informacij o različnih komponentah grafa. Trenutna različica vsebuje vrsto algoritmov za uporabo v teoriji grafov, podatkovnem rudarjenju in analizi omrežij, kot so razvrščanje v skupine, dekompozicija, optimizacija, generiranje naključnih grafov, statistika in računanje razdalj, pretokov ter mer za pomembnost vozlišča (središčnost, PageRank, kazala & viri) [5].

Z orodji se lahko podatke predstavi tudi vizualno, razvijalcem pa omogoča različne mehanizme s katerimi lahko omrežja interaktivno spreminjajo in filtrirajo podatke.

JUNG je prosto dostopna odprtokodna knjižnica.

2.3. Orange Canvas

Orange je več-komponentna programska oprema za podatkovno rudarjenje in strojno učenje, ki ponuja grafični vmesnik za analizo podatkov in vizualizacijo ter knjižnice za naprednejšo analizo v programskem jeziku Python. Vsebuje komponente za predobdelavo podatkov, napovedovanje rezultatov, filtriranje in modeliranje podatkov ter različne raziskovalne tehnike. Razvit je bil v jezikih C++ in Python [6].

Orange je prosto dostopen pod licenco GPL. Razvija ga Laboratorij za bioinformatiko na Fakulteti za računalništvo in informatiko v Ljubljani. Uporaba programa Orange je možna na različnih platformah, kot so Linux, MacOS X in Microsoft Windows.

V nalogi je Orange uporabljen za analizo primernosti izbranih atributov glede na določen razred. Na podlagi dobljenih rezultatov ocenimo, kateri atributi so primerni, kakšne so njihove prednosti in pomanjkljivosti ter za katero vrsto podatkov je določen atribut najbolj primeren.

Prav tako uporabimo program za izdelavo .tab datotek, v katerih določimo attribute, njihovo vrsto in atributni razred. Te datoteke kasneje uporabimo v programu Orange Canvas kot vir podatkov za potrebe podatkovnega rudarjenja.

2.4. Java

Java je objektno usmerjeni programski jezik, ki ga je razvil James Gosling v podjetju Sun Microsystems, ki je trenutno del Oracle Corporation. Izdana je bila leta 1995 kot del Sun Microsystems' Java platforme.

Način pisanja programske kode je večinoma povzet iz programskih jezikov C in C++. Namen tega programskega jezika je »napiši enkrat, poženi kjerkoli« (angl. *write once, run anywhere* ali *WORA*), kar pomeni, da programske kode, ki je bila pognana na eni platformi, ni potrebno ponovno prevajati, da bi jo lahko pognali na neki drugi platformi. Java aplikacije so praviloma prevedene v bitno kodo (datoteka .class), ki lahko teče le na Java virtual machine (JVM) ne glede na vrsto računalnika in operacijskega sistema, na katerem poganjamo kodo.

Trenutno je Java eden najpriljubljenejših programskih jezikov, predvsem na področju spletnih aplikacij z okrog 9 milijonov razvijalcev. Leta 2007 je Sun ponovno licenciral večino Java tehnologije pod licenco GNU General Public License.

Poglavje 3

3. Badminton

V tem poglavju predstavimo badminton v Sloveniji, opišemo način igranja badmintonskih turnirjev in pravilnega beleženja rezultatov, kar nam bo pomagalo pri razumevanju badmintonskih posebnosti v nadaljevanju naloge.

3.1. Osnovno o badmintonu

Badminton je šport, pri katerem se uporabljata lopar in operjena žogica iz plutovine. Igralec poskuša udariti žogico preko mreže v nasprotnikovo polje. Točko doseže igralec, tako da spravi žogico v nasprotnikovo polje ali tako, da nasprotni igralec stori napako. Igra se lahko tudi v parih.

Definicije:

- igralec/igralka: vsaka oseba, ki igra badminton,
- igra: osnovna oblika tekmovanja v badmintonu med dvema nasprotnima stranema, ki vključujeta enega ali dva igralca,
- igra posameznikov: igra, pri kateri je na vsaki od obeh nasprotnih strani po en igralec,
- igra dvojic: igra, pri kateri sta na vsaki od obeh nasprotnih strani po dva igralca.

Sistem štetja:

- Igra sestoji iz dveh dobljenih setov, v kolikor ni dogovorjeno drugače
- Set zmagata igralec, ki prvi doseže 21 točk, razen:
 - pri rezultatu 20:20 zmagata igralec, ki prvi doseže prednost dveh točk,
 - pri rezultatu 29:29 zmagata igralec, ki prvi doseže 30 točk [7].

3.1.1. Badminton v Sloveniji

Badminton se je začel v Sloveniji igrati že leta 1955, ko so v Elanu začeli izdelovati prve loparje za badminton po nemškem vzorcu. Prvi badmintonski klub je bil ustanovljen leta 1956 v sklopu ŠD Papirničar v Vevčah v Ljubljani. Ustanovitelj je bil dr. Leander Bleiweis. Za začetnike slovenskega badmintona veljajo Janez in Danče Pohar, Tomaž Pavčič, Vida Leskobar in drugi.

Osamosvojitve v slovenskem badmintonu ni prinesla velikih sprememb, saj so že reprezentanco v nekdanji Jugoslaviji sestavljali večinoma igralci iz Slovenije [8].

Po podatkih Badminton Europe v Sloveniji igra badminton 1696 igralcev in igralk v 29 različnih klubih [9].

3.2. Tekmovalni sistem

V nalogi so uporabljeni rezultati in lestvica iz A in B turnirjev, ki jih organizira Badmintonska zveza Slovenije (BZS) [10].

A in B turnirji se igrajo po sistemu modificiranega KO sistema z razigravanjem za vsako mesto (t.i. Monrad sistem v Tournament Plannerju):

- Tekmovalci se srečujejo po parih v vsakem kolu. Zmagovalec se uvrsti v naslednje kolo, poraženec pa igra razigravanje za končna mesta.
- Glavna tabela obsega 32 moških posamezno pri A turnirju, pri B turnirju pa je velikost tabele poljubna. Kvalifikacijski kriterij za uvrstitev v tabelo je SJLP (Slovenska jakostna lestvica posameznikov). V primeru prevelikega števila prijavljenih se oblikuje čakalna lista, s katere se tekmovalci, v primeru odpovedi višje postavljenih tekmovalcev, na tekmovanje uvrstijo po vrsti, kakor so razvrščeni na čakalni listi.
- Nosilce se razporedi v tabelo glede na pravila o enojnem KO sistemu. Pri moških posamezno je 16 nosilcev.

Tekmovanja za slovensko jakostno lestvico posameznikov (v nadaljevanju SJLP) v eni sezoni so:

- najmanj 4 in največ 6 turnirjev B kategorije,
- najmanj 3 in največ 6 turnirjev A kategorije,
- državno člansko prvenstvo.

Pravico udeležbe na turnirjih B kategorije imajo:

- Vsi registrirani tekmovalci BZS

Pravico udeležbe na turnirjih A kategorije imajo:

- tekmovalci, uvrščeni na prvih 32 mest, oziroma tekmovalke, uvrščene na prvih 16 mest na slovenski jakostni lestvici (SJLP) oz. prvi naslednji po SJLP, do zapolnitve tabel.

- Pravico do udeležbe imata tudi 2 tekmovalca in 2 tekmovalki, ki ju lahko prosto določi organizator, ter 2 tekmovalca in 2 tekmovalki, ki ju lahko prosto določi BZS (wild card). Tekmovalci, katerim je wild card podeljen, nastopijo na tekmovanju kot zadnje postavljeni tekmovalci.

Pravico udeležbe na državnem članskem prvenstvu imajo:

- vsi tekmovalci, ki so slovenski državljani.

3.2.1. Jakostna lestvica A in B kategorije

Vrstni red tekmovalcev in tekmovalk na SJLP je določen s številom točk, ki jih tekmovalci zberejo na podlagi svojih uvrstitev na turnirjih.

Točke za posamezne uvrstitve se dodelijo v skladu s tabelo Tabela 3.1.

mesto	točke- A turnir	točke- B turnir	točke- DP
1	100	90	120
2	98	86	110
3	96	82	100
4	94	78	-
5	92	76	96
6	90	74	-
7	88	72	-
8	86	70	-
9	84	68	92
10	82	66	-
11	80	64	-
12	78	62	-
13	76	60	-
14	74	58	-
15	72	56	-
16	70	54	-
17	69	52	84
18	68	50	-
19	67	48	-
20	66	46	-
21	65	44	-
22	64	42	-
23	63	40	-
24	62	38	-
25	61	36	-
26	60	34	-

27	59	32	-
28	58	30	-
29	57	28	-
30	56	26	-
31	55	24	-
32	54	22	-
33-64	-	20	78
65-			70

TABELA 3.1: TOČKOVANJE ZA RAZLIČNE VRSTE TURNIRJEV

3.3. Pridobivanje podatkov

Podatki o turnirjih so pridobljeni na spletni strani www.tournamentsoftware.com, kjer so na voljo rezultati turnirjev in posameznih tekem v različnih športih, med njimi so najštevilčnejši podatki o športih z loparji, kot so badminton, tenis, namizni tenis itd. Spletna stran deluje v okviru plačljivih programov (Tournament Planner, League Planner,...), s pomočjo katerih lahko vodje tekmovanj vodijo turnir. Rezultate se lahko prikaže na spletni strani, njihov ogled pa je dovoljen komurkoli [11].

Za zbirko rezultatov smo izbrali obdobje od 1. septembra 2010 do 31. januarja 2013. Podatki so zbrani za turnirje A in B kategorije, ne pa tudi za državno prvenstvo.

Seznam članskih turnirjev, ki so upoštevani v nalogi, je prikazan v tabelah Tabela 3.2, Tabela 3.3, Tabela 3.4 in Tabela 3.5. Zraven sta še podatka o času in lokaciji turnirja. Turnirji A kategorije po navadi potekajo v dveh dneh, zato je naveden samo datum igranja moških posamezno.

Sezona 2010/2011

Razpored za A turnirje

24. oktober 2010	Zgornja Kungota	1. turnir A kategorije v sezoni 2010/2011
28. november 2010	Mirna	2. turnir A kategorije v sezoni 2010/2011
9. januar 2011	Medvode	3. turnir A kategorije v sezoni 2010/2011
13. marec 2011	Mirna	4. turnir A kategorije v sezoni 2010/2011

Razpored za B turnirje

18. september 2010	Medvode	1. turnir B kategorije v sezoni 2010/2011
14. november 2010	Lendava	2. turnir B kategorije v sezoni 2010/2011
19. februar 2011	Mirna	3. turnir B kategorije v sezoni 2010/2011
9. april 2011	Ljubljana	4. turnir B kategorije v sezoni 2010/2011
23. april 2011	Ljubljana	5. turnir B kategorije v sezoni 2010/2011

TABELA 3.2: RAZPORED UPORABLJENIH TURNIRJEV ZA SEZONO 2010/2011

Sezona 2011/2012**Razpored za A turnirje**

18. september 2011	Zgornja Kungota	1. turnir A kategorije v sezoni 2011/2012
11. december 2011	Medvode	2. turnir A kategorije v sezoni 2011/2012
15. januar 2012	Mirna	3. turnir A kategorije v sezoni 2011/2012
25. marec 2012	Ljubljana	4. turnir A kategorije v sezoni 2011/2012

Razpored za B turnirje

9. september 2011	Mirna	1. turnir B kategorije v sezoni 2011/2012
6. november 2011	Medvode	2. turnir B kategorije v sezoni 2011/2012
27. november 2011	Lendava	3. turnir B kategorije v sezoni 2011/2012
26. maj 2012	Ljubljana	6. turnir B kategorije v sezoni 2011/2012

TABELA 3.3: RAZPORED UPORABLJENIH TURNIRJEV ZA SEZONO 2011/2012

Sezona 2012/2013**Razpored za A turnirje**

23. september 2012	Medvode	1. turnir A kategorije v sezoni 2012/2013
25. november 2012	Mirna	2. turnir A kategorije v sezoni 2012/2013
20. januar 2013	Ljubljana	3. turnir A kategorije v sezoni 2012/2013
7. april 2013	Zgornja Kungota	4. turnir A kategorije v sezoni 2012/2013
19. maj 2013	Kočevje	5. turnir A kategorije v sezoni 2012/2013

Razpored za B turnirje

27. oktober 2012	Mirna	1. turnir B kategorije v sezoni 2012/2013
15. december 2012	Ljubljana	2. turnir B kategorije v sezoni 2012/2013
23. februar 2013	Medvode	3. turnir B kategorije v sezoni 2012/2013
23. marec 2013	Ljubljana	4. turnir B kategorije v sezoni 2012/2013

TABELA 3.4: RAZPORED UPORABLJENIH TURNIRJEV ZA SEZONO 2012/2013

Sezona 2013/2014**Razpored za A turnirje**

15. september 2013	Ljubljana	1. turnir A kategorije v sezoni 2013/2014
24. november 2013	Zgornja Kungota	2. turnir A kategorije v sezoni 2013/2014
15. december 2013	Medvode	3. turnir A kategorije v sezoni 2013/2014
12. januar 2014	Mirna	4. turnir A kategorije v sezoni 2013/2014

Razpored za B turnirje

26. oktober 2013	Medvode	1. turnir B kategorije v sezoni 2013/2014
-------------------------	---------	---

TABELA 3.5: RAZPORED UPORABLJENIH TURNIRJEV ZA SEZONO 2013/2014

Poglavje 4

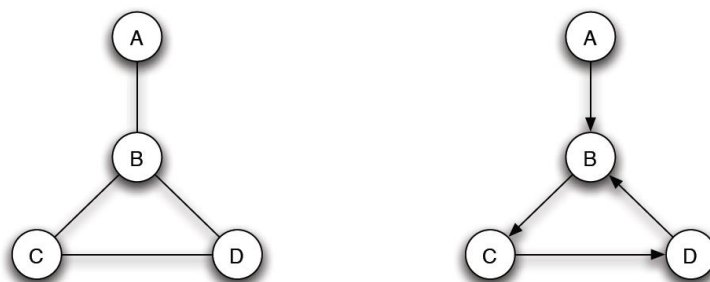
4. Teorija omrežij in podatkovno rudarjenje

V tem poglavju si ogledamo nekaj splošnih značilnosti omrežij in bistvene lastnosti podatkovnega rudarjenja.

4.1. Omrežje

Omrežje si lahko predstavljamo kot skupek vozlišč (angl. *node*) in povezav (angl. *edge*) med njimi. Vozlišča so lahko konkretni objekti, kot npr. ljudje, mesta itd. Povezava določa neko relacijo med vozliščema, ki ju povezuje. V socialnih omrežjih to lahko pomeni, da je neko vozlišče A prijatelj z vozliščem B, v našem primeru pa to pomeni, da je igralec A igral tekmo z igralcem B.

Usmerjeno omrežje ali usmerjen graf (angl. *directed graph, digraph*) je omrežje pri katerem vsaka povezava kaže na določeno vozlišče. Te povezave se imenujejo usmerjene povezave (angl. *directed edges*), prikažemo pa jih s črtami in puščico glede na smer, v katero kažejo.



SLIKA 1: PRIMER NEUSMERJENEGA IN USMERJENEGA OMREŽJA

V omrežju badmintonskih tekem se uporabljajo usmerjene povezave, saj smer določa, kateri igralec je medsebojno tekmo zmagal. Če je povezava usmerjena proti vozlišču A, to pomeni, da je ta igralec, ki ga predstavlja vozlišče A, bil zmagovalec.

Omrežje navadno predstavimo z matematičnim objektom, ki mu pravimo graf. Graf $G(V, E)$ je sestavljen iz množice vozlišč $V = \{v_1, v_2, \dots, v_n\}$, in množice povezav med njimi $E = \{e_i, e_j \mid e_i, e_j \in V\}$ [12]. Predstavimo ga lahko tudi z matriko sosednosti (angl. *adjacency matrix*). Matrika sosednosti vozlišč omrežja N z n vozlišči je $n \times n$ kvadratna matrika $A(N)$, kjer element matrike a_{ij} predstavlja število povezav iz vozlišča v_i v v_j .

Povezave med vozlišči so lahko določene binarno, torej obstajajo/ne obstajajo, lahko pa imajo tudi svoje vrednosti glede na frekvenco, pomembnost itd. V članku o napovedovanju rezultatov v tenisu, na katerega se opiramo, utež vozlišča določa čas igranja tekme. To pomeni, da imajo tekme, ki so bile odigrane nekaj let nazaj, manjšo pomembnost od tistih, ki so bile odigrane nedolgo nazaj med istima igralcema. V našem primeru časovna komponenta zaradi kratkega časovnega obsega obdelanih podatkov ni potrebna, je pa toliko pomembnejši pomen tekme. Kot že omenjeno, se v badmintonu za razliko od tenisa igra t.i. modificiran KO sistem in zmaga v finalu je pomembnejša od zmage v borbi za 9.-16. mesto. Zato je treba pri merjenju središčnosti upoštevati tudi to komponento.

Omrežje je povezano, če za poljubni dve vozlišči obstaja povezava med njima. Takim omrežjem pravimo, da so sestavljena iz ene same komponente, v splošnem pa lahko imamo omrežje sestavljeno iz največ n komponent. Omrežje v nalogi je sestavljeno iz ene komponente, saj je vsak igralec, ki je del omrežja odigral tekmo vsaj z enim drugim igralcem v omrežju.

4.2. Lastnosti omrežij

V tem delu predstavimo lastnosti omrežij in poudarimo tiste, ki so bistvene za delo z izbranim omrežjem [13] [14].

4.2.1. Mere središčnosti

Pomembna lastnost omrežij so mere središčnosti. Središčnost določa, kako pomembno je neko vozlišče (ali povezava) v določenem omrežju. Obstaja več matematičnih načinov določanja središčnosti, ki se razlikujejo po tem, po katerem konceptu se sploh določa, ali je neko vozlišče v središču omrežja.

Mere središčnosti glede na stopnjo omrežja

Enostavna in dobro razumljiva mera je *stopnja vozlišča*. V neusmerjenem omrežju je stopnja vozlišča kar število povezav, ki so povezana z določenim vozliščem, v usmerjenem pa stopnjo delimo na vhodno stopnjo (angl. *in-degree*) in izhodno stopnjo (angl. *out-degree*). Vhodna stopnja je število povezav, ki gredo proti vozlišču, izhodna pa število povezav, ki gredo stran

od vozlišča. V našem primeru vhodne povezave kažejo zmago igralca, izhodne pa poraz proti igralcu na katerega kaže povezava.

Tako središčnost imenujemo kar stopnja ali stopenjska središčnost (angl. *degree, degree centrality*). V primeru nekega tekmovanja lahko enostavno vidimo, da je igralec, ki ima veliko vhodnih povezav (zmag) boljši, kot tisti, ki ima več odhodnih povezav (porazov). Problem pri tej meri je, da favorizira količino zmag ali porazov, pri tem pa ne upošteva, koliko tekem je nek igralec odigral in proti kako močnim nasprotnikom. Ali na primeru socialnih omrežjih, nespametno bi bilo sklepati, da ima nekdo, ki ima veliko povezav, večjo veljavo kot nekdo, ki ima malo povezav. Zato je potrebno vedeti tudi, kako pomembna so vozlišča, s katerimi je trenutno vozlišče povezano.

Središčnost lastnega vektorja (angl. *eigenvector centrality*) meri pomembnost določenega vozlišča v omrežju. Nanjo lahko gledamo kot na naravni podaljšek stopnje vozlišča. K pomembnosti določenega vozlišča v omrežju prispeva dejstvo, da ima vozlišče neposredne povezave z drugimi vozlišči, ki so prav tako pomembna in vplivna v omrežju. V našem primeru to pomeni, da je zmaga nad tekmovalcem, ki je zelo uspešen, pomembnejša kot zmaga nad tekmovalcem, ki je manj uspešen. V nalogi središčnosti lastnega vektorja kot take ne uporabimo, ampak uporabimo PageRank algoritem, ki izhaja iz te središčnosti.

PageRank je algoritem, ki ga je razvil Google za rangiranje spletnih strani v njihovih iskalnih algoritmih. Razvila sta ga Larry Page in Sergey Brin leta 1996. V tistih časih so spletni iskalniki dajali prednost stranem z največjo gostoto ključnih besed. To slabost iskalnikov je bilo lahko izigrati z večkratnim ponavljanjem ključnih besed in si tako zagotoviti visok položaj med iskalnimi rezultati. Page in Brin sta prišla na idejo, da bi bile spletne strani na svetovnem spletu urejene hierarhično, po pomembnosti. Stran naj bi bila pomembna glede na število povezav, ki kažejo nanjo.

Algoritem lahko uporabimo tudi v drugih situacijah, npr. pri analizi športnih rezultatov, kot je to prikazano v nadaljevanju. V tem primeru vozlišča grafa niso spletne strani, ampak igralci. Povezave med njimi so odigrane medsebojne tekme. V primeru, da puščica kaže proti izbranemu vozlišču, to pomeni, da je igralec premagal nasprotnika iz katerega povezava prihaja. V primeru, da je izbrani igralec izgubil pa smer povezave kaže proti nasprotnemu igralcu. Ker PageRank upošteva proti kako uspešnemu igralcu je trenutni igralec igral, to pomeni, da je zmaga nad bolje rangiranim igralcem vrednejša od zmage nad slabše rangiranim.

PageRank, za razliko od nekaterih drugih mer središčnosti, deluje zato, ker v primeru, da nek visoko postavljeni igralec igra z veliko drugimi igralci, to še ne pomeni, da bodo ostali igralci po njem podedovali njegov visok položaj, ampak se bo le-ta porazdelil med vse odhodne povezave tega igralca. Tako nek igralec ni dober samo zato, ker je igral z najbolje postavljenim igralcem.

Slabost algoritma PageRank v našem primeru je v tem, da se za razliko od tenisa na turnirjih BZS igrajo tudi tekme za razvrstitev. To posledično lahko pomeni, da bi igralec, ki bi izgubil v prvem krogu in nato zmagal vse preostale tekme do konca, imel višji PageRank kot igralec, ki je osvojil 4. mesto, za katerega pa velja, da je izgubil samo zadnji dve tekmi. V primeru, da je baza rezultatov dovolj velika, se ta nepravilnost težje zazna, vendar je še vedno lahko prepoznavna. Možna rešitev je v tem, da se tekem za razvrstitev ne upošteva pri računanju PageRank-a. V primeru omrežja badmintonskih tekem v Sloveniji, ki ga uporabljamo kot vir, to ni mogoče, saj bi že tako majhno množico podatkov še dodatno zmanjšalo.

Vse dosedanje mere središčnosti najbolj nagradijo vozlišče, na katerega kaže veliko število pomembnih povezav. Obstajajo pa tudi situacije, v katerih bi vozlišče, ki kaže na čim več drugih vozlišč, moralo imeti večjo mero središčnosti. Takšen primer je pisanje člankov, v katerih avtorji na kratko opišejo neko tematiko, za bolj poglobljeno analizo pa predstavijo vire. Tej meri središčnosti pravimo *kazala in viri* (angl. *Hubs & Authorities*). V takšnih omrežjih obstajata dve vrsti vozlišč. Prva so viri, ki vsebujejo ključne informacije o iskani tematiki, druga pa so kazala, ki nam povejo, kateri so najboljši viri za neko tematiko. Algoritem za računanje te mere je HITS (angl. *hyperlink-induced topic search*).

Mere središčnosti glede na dostopnost omrežja

Bližinska središčnost (angl. *closeness centrality*) meri povprečno razdaljo določenega vozlišča do vseh drugih vozlišč v omrežju. V splošnem to pomeni, koliko korakov moramo v povprečju narediti, da pridemo od izbranega vozlišča do kateregakoli drugega vozlišča v omrežju. Do tistih vozlišč, ki imajo majhno vrednost bližinske središčnosti, informacija pride hitreje ali pa imajo bolj neposreden vpliv na ostala vozlišča. Bližinska središčnost je primerna predvsem za uporabo v družabnih omrežjih.

Ekscentrična središčnost (angl. *eccentricity centrality*) označuje najdaljšo razdaljo izmed vseh najkrajših razdalj, ki jih lahko določimo za neko vozlišče v omrežju. Če je ekscentrična središčnost visoka za določeno vozlišče, to pomeni, da so ostala vozlišča v njegovi bližini.

Mere središčnosti glede na vmesnost omrežja

Središčnost vmesnosti (angl. *betweenness centrality*) nam pove pogostost nekega vozlišča na poti med drugimi vozlišči. Tista vozlišča, ki so pogosto na poti med drugimi vozlišči, naj bi imela več informacij in s tem večji vpliv v celotnem omrežju. Odstranitev takega vozlišča iz omrežja bi pomenila drastično spremembo pretoka informacij skozi omrežje.

Obstaja tako več vrst središčnosti, vendar se mere središčnosti glede na vmesnost omrežja in glede na dostopnost omrežja niso izkazale kot potrebne pri razumevanju in analizi danega omrežja.

4.3. Podatkovno rudarjenje

Podatkovno rudarjenje je veja računalništva, pri kateri lahko odkrivamo vzorce v velikih podatkovnih množicah s pomočjo umetne inteligence, strojnega učenja in statistike. Cilj podatkovnega rudarjenja je izločiti čim več in čim bolj kvalitetne informacije iz množice podatkov, jih poskušati razumeti, da jih lahko kasneje uporabimo na novih podatkih ali tehnikah.

Pred začetkom je potrebno pridobiti množico podatkov, nad katero se bo rudarjenje izvajalo. Podatke je treba analizirati in ugotoviti, kateri so dovolj kvalitetni za nadaljnjo obdelavo in katere se lahko zavrže. Po navadi podatkovno množico razdelimo na učno množico in testno množico. Na učni množici se s pomočjo napovednega modela učimo, kako se ta obnaša, nato pa se na testni množici prepričamo ali je naučeno znanje pravilno.

4.3.1. Načini podatkovnega rudarjenja

Odkrivanje anomalij (angl. *anomaly detection*) – identifikacija nenavadnih podatkov in napak v podatkih, ki bi lahko bile zanimive za nadaljnje preiskovanje.

Učenje asociacijskih pravil (angl. *association rule learning*) – iskanje razmerij med spremenljivkami. Na primer, v trgovini zbirajo podatke o nakupovalnih navadah njihovih

kupcev. Z uporabo posebnih metod lahko v trgovini določijo, kateri izdelki so kupljeni skupaj in te ugotovitve uporabijo za oglaševanje izdelkov. Tej metodi pravimo tudi analiza nakupovalne košarice.

Gručenje, razvrščanje v skupine (angl. *clustering*) – odkrivanje skupin in struktur v podatkih, ki so si na nek način podobne, ne da bi poznali dejansko strukturo podatkov.

Klasifikacija (angl. *classification*) – razdelitev podatkov iz določene podatkovne množice v več razredov. Na primer, program za prejemanje elektronske pošte lahko pošto označi kot legitimno ali kot »spam« glede na lastnosti atributov. Razred je diskretna spremenljivka.

Regresija (angl. *regression*) – statistični proces, ki se uporablja pri ocenjevanju razmerja med spremenljivkami. Razred je zvezna spremenljivka.

Povzemanje (angl. *summarization*) – prinaša bolj zgoščeno predstavitev podatkovne množice, vključno z vizualizacijo podatkov in ustvarjanjem poročil.

Pri podatkovnem rudarjenju lahko ne namenoma pridobljene informacije napačno razumemo ali predstavimo. Lahko se zgodi, da naučene informacije, ki delujejo na testnih podatkih, ne delujejo na novih podatkih. Glavni razlog za to je, da se pri iskanju pravih atributov preveč osredotočamo na testno množico in tako svojo izbiro prilagodimo samo testni množici (angl. *overfitting*). Zato je potrebno pridobljene informacije analizirati in se prepričati, da le-te delujejo tudi na širši podatkovni množici, kot je testna množica [15].

Pri naši nalogi poskušamo ugotoviti, kateri atributi so ključni pri napovedovanju rezultata in kakšna so razmerja med njimi, zato uporabimo regresijske tehnike.

4.3.2. Regresija

Regresija je statistični proces, ki se uporablja pri ocenjevanju razmerja med spremenljivkami. Bolj natančno, regresijska analiza pomaga razumeti, kako na vrednost odvisne spremenljivke vpliva spreminjanje ene neodvisne spremenljivke, medtem ko ostale neodvisne spremenljivke ostajajo enake. Za razliko od klasifikacije, kjer so atributi diskretni, morajo biti atributi, ki se uporabljajo pri regresiji zvezni [16].

4.3.2.1. Regresijske metode

Za uporabo v podatkovnem rudarjenju lahko uporabimo različne regresijske metode:

Metoda k-najbližjih sosedov (angl. *k-nearest neighbours, kNN*) napoveduje vrednost vozlišča tako, da upošteva le *k* najbližjih testnih primerov. kNN algoritem pri regresiji je prilagojen za napovedovanje zveznih spremenljivk, pri klasifikaciji pa za diskretne spremenljivke [17] [18].

Regresijsko drevo (angl. *regression tree*) se uporablja, kadar so nam vrednosti učne množice znane vnaprej. Razred mora biti zvezen, medtem ko so atributi lahko zvezni ali diskretni. Razdelitev v množico se dela z metodo najmanjših kvadrantov. Drevo je sestavljeno iz vozlišč, ki predstavljajo attribute, vej, ki predstavljajo podmnožice vrednosti atributov in listov, kateri predstavljajo funkcije, ki preslikajo vektor vrednosti atributov v zvezni razred.

Naključni gozd (angl. *random forest*) je algoritem, kjer se hkrati uporablja večje število regresijskih dreves. Drevo je odvisno od vrednosti naključno vzorčenega vektorja. V različnih drevesih se uporabljajo različni vektorji, ki pa morajo biti enako porazdeljeni [19].

SVM algoritem (angl. *support vector machine*) – pri tem algoritmu model, ki se dobi kot rezultat, ni zgrajen na celotno učno množico, ampak le na del nje, z razlogom, da se ignorirajo tisti učni podatki, ki so po vrednosti približno enaki kot tisti, ki jih je model napovedal. Namenjen je za učenje na velikih množicah z velikim številom manj pomembnih atributov.

Linearna regresija (angl. *linear regression*) je metoda, ki z izgradnjo linearnega modela, določi odnos med odvisnimi in neodvisnimi spremenljivkami. Če linearno regresijo uporabljamo za napovedovanje, izračunamo linearni model na učni množici, katerega nato uporabimo za napovedovanje razreda.

4.3.2.2. Analiza uspešnosti regresijskih metod

Za analizo uspešnosti regresijskih metod se uporabljajo:

Srednja kvadratna napaka (angl. *mean squared error, MSE*) je definirana kot povprečni kvadrat razlike med napovedano in pravo vrednostjo.

Predstavlja jo formula:

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2,$$

kjer je n število vseh primerov danega problema, \hat{Y}_i napovedana vrednost primera i in Y_i prava vrednost primera i .

Relativna kvadratna napaka (angl. *relative squared error, RSE*). MSE je odvisna od razpona možnih vrednosti funkcije, zato je bolj smiselno uporabiti relativno kvadratno napako.

Predstavlja jo formula:

$$RSE = \frac{n * MSE}{\sum_{i=1}^n (Y_i - \bar{Y})^2},$$

kjer je n število vseh primerov danega problema, MSE je srednja kvadratna napaka, \bar{Y} je povprečna vrednost pravih vrednosti in Y_i prava vrednost primera i .

Srednja absolutna napaka (angl. *mean absolute error, MAE*) meri povprečno absolutno razliko med napovedano vrednostjo in pravo vrednostjo.

Predstavlja jo formula:

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - Y_i|,$$

kjer je n število vseh primerov danega problema, f_i napovedana vrednost primera i in Y_i prava vrednost primera i .

Relativna absolutna napaka (angl. *relative absolute error, RAE*): MAE je odvisna od razpona možnih vrednosti funkcije, zato je bolj smiselno uporabiti relativno absolutno napako.

Predstavlja jo formula:

$$RAE = \frac{n * MAE}{\sum_{i=1}^n |f_i - Y_i|},$$

kjer je n število vseh primerov danega problema, MAE srednja absolutna napaka, f_i napovedana vrednost primera i in Y_i prava vrednost primera i .

Poglavje 5

5. Analiza badmintonskega omrežja in napoved rezultata

V poglavju najprej analiziramo majhno omrežje, ki je sestavljeno iz enega turnirja, nato pa ugotovitve uporabimo na večjem omrežju, ki ga sestavlja celotna podatkovna množica. Sledi opis urejanja podatkov, ki so primerni za poskus napovedi in na koncu še sama napoved s pomočjo ustreznih orodij.

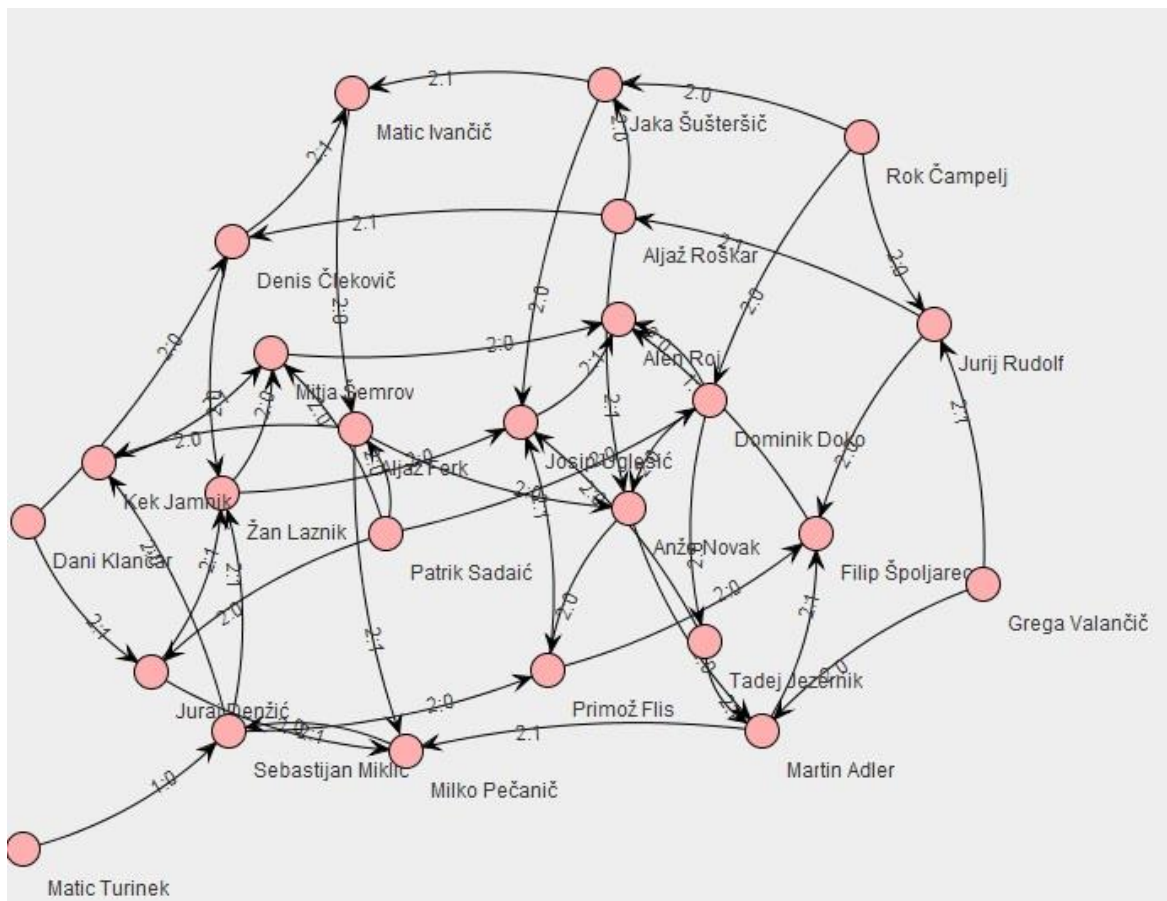
5.1. Analiza turnirja A1-2012-2013

Najprej pokažemo delovanje mer za analizo omrežja in njihove prednosti ter pomanjkljivosti za posamezen turnir. S tem lažje razumemo obnašanje kompleksnejšega omrežja.

Kot primer vzamemo 1. A turnir v sezoni 2012/2013. Na turnirju je nastopilo 25 igralcev po modificiranem KO sistemu. Ker je igralcev manj kot 33, so igralci razporejeni v drevo z 32 listi po ključu nosilcev kot je nakazano v pravilniku BZS. Igralec, ki zmagaja napreduje po drevesu, tisti, ki pa izgubi pa se seli v poddrevo, kjer se bori za posamezna mesta. Če igralec npr. izgubi v 1. krogu, se seli v poddrevo za mesta od 17.-24. mesta, če v drugem, od 9.-16., če v tretjem, od 5.-8. in v četrtem od 3.-4. mesta. Če igralec, ki igra za mesta 5.-8. zmagaja, se nato bori za mesta 5.-6., če pa izgubi pa za mesta 7.-8.. Podobno je tudi pri ostalih poddrevesih. Če bi bilo na turnirju 33 ali več igralcev, bi se ti razdelili v drevo s 64 listi.

Vsak igralec bi torej moral odigrati 5 tekem, razlogov zakaj je nek igralec odigral manj tekem pa je več:

- ker je nastopilo manj kot 32 igralcev, to pomeni, da so lahko nekateri igralci v prvih dveh krogih prosti. V prvem krogu so to nosilci, ki so prosti po ključu razporeditve igralcev v tabelo, v drugem krogu pa je v primerih, kjer se igra za mesta, lahko prost igralec, ki je izgubil z nosilcem
- tekma je bila dodeljena v korist enega od igralcev, ker je drugi igralec dvoboj predal zaradi poškodbe ali pa se ni pojavil na dvoboju. Ti dvoboji se ne štejejo v računanje mer, ker je iz pridobljenih podatkov nemogoče napovedati, kateremu od igralcev je bila dodeljena zmaga, brez da bi morali preveriti s kom igra v naslednjem krogu.



SLIKA 2: PRIMER GRAFA ODIGRANIH TEKEM ZA POSAMEZEN TURNIR

Mesto	Tekmovalac	PageRank	mPR	dPR	Uteženi Page Rank	mUPR	dUPR	InDeg	OutDeg
1	Alen Roj	0.188262	1	0	0.282099	1	0	4	0
2	Mitja Šemrov	0.072609	2	0	0.124613	3	-1	3	1
3	Kek Jamnik	0.042891	9	-6	0.081474	5	-2	2	1
4	Filip Špoljarec	0.058142	5	-1	0.115654	4	0	3	1
5	Josip Uglešič	0.069960	3	+2	0.137614	2	+3	4	1
6	Žan Laznik	0.049683	7	-1	0.078862	7	-1	3	2
7	Primož Flis	0.046115	8	-1	0.081226	6	+1	2	2
8	Sebastijan Miklič	0.065522	4	+4	0.067884	9	-1	2	3
9	Milko Pečanič	0.050095	6	+3	0.078640	8	+1	3	1
10	Martin Adler	0.040325	11	-1	0.056958	10	0	3	2
11	Juraj Denzič	0.020306	18	-7	0.018173	18	-7	2	2
12	Tadej Jezernik	0.017656	20	-8	0.018513	16	-4	1	2
13	Anže Novak	0.035647	12	+1	0.051382	11	+2	3	2
14	Aljaž Ferko	0.042091	10	+4	0.038164	12	+2	2	3
15	Dominik Doko	0.018549	19	-4	0.017085	20	-5	2	3
16	Patrik Sadaić	0.012400	21	-5	0.007835	21	-6	0	4
17	Matic Ivančič	0.031829	13	+4	0.037407	13	+4	2	1
18	Jaka Šušteršič	0.021979	15	+3	0.024302	14	+4	2	2
19	Denis Člekovič	0.023735	14	+5	0.023016	15	+4	2	2
20	Aljaž Roškar	0.021404	16	+4	0.018513	17	+3	1	3
21	Jurij Rudolf	0.021184	17	+4	0.017085	19	+2	2	2
22	Matic Turinek	0.012400	21	+1	0.007835	21	+1	0	1
23	Rok Čampelj	0.012400	21	+2	0.007835	21	+2	0	3
24	Dani Klančar	0.012400	21	+3	0.007835	21	+3	0	2
25	Grega Valančič	0.012400	21	+4	0.007835	21	+4	0	2

TABELA 5.1: ANALIZA POSAMEZNEGA TURNIRJA

Legenda:

mPR ... mesto igralca glede na izračunano PageRank vrednost

dPR ... razlika med mPR in dejanskim mestom

mUPR ... mesto igralca glede na izračunano uteženo PageRank vrednost

dUPR ... razlika med mUPR in dejanskim mestom

InDeg ... število zmag tekmovalca (vhodna stopnja vozlišča)

OutDeg ... število porazov tekmovalca (izhodna stopnja vozlišča)

Prav zaradi načina tekmovanja, torej da se po prvem porazu igra še naprej za mesta, pokažeta stopnja vozlišča in PageRank nekoliko drugačne rezultate od zelenih. Če so odigrane vse tekme, ima peto-uvrščeni igralec več zmag (4 proti 3) in manj porazov (1 proti 2) kot četrto-uvrščeni. Podobno velja tudi za nekatera ostala mesta naprej. Stopnja tako ne

pokaže dejanskega stanja, nekoliko bolje pa je z algoritmom PageRank, saj ta upošteva proti katerim igralcem (boljšim ali slabšim) je izbrani igralec igral.

Linearno povezanost PageRank algoritma z doseženimi mesti izmerimo s Pearsonovim koeficientom, ki je za ta turnir enak -0.768. Podobno velja tudi za ostale turnirje, kjer se Pearsonov koeficient giblje med -0,70 in -0,85, kot je razvidno v tabeli 5.2.

Vendar pa se da PageRank algoritem tudi izboljšati, tako da se upošteva krog tekmovanja in dosedanja uspešnost tekmovalca. Idealno bi bilo, da bi bilo v podatkih za vsako tekmo določeno, za kateri del tekmovanja gre. Tega pa iz pridobljenih podatkov ni mogoče razbrati, saj se lahko zgodi, da igralec do trenutnega kroga ni zbral maksimalnega števila nastopov zaradi razlogov, ki so navedeni zgoraj. Tako upoštevamo samo, kolikokrat je igralec izgubil do trenutnega kroga. Če igralec še ni izgubil, bo trenutna tekma imela utež 1, če pa je že izgubil pa bo utež med 0 in 1, odvisno od tega kolikokrat je igralec do tedaj že izgubil.

Rezultati Pearsonovega koeficienta se spremenijo za približno 3% v povprečju na turnir. Menimo, da se da PageRank še nekoliko izboljšati, če bi imeli na voljo popolnejše podatke. Prav tako je možno v PageRank algoritem dodati tudi časovno komponento, ki določa, kdaj sta igralca med seboj igrala dvoboj. Dvoboj pred petimi leti bi imel manjšo težo kot dvoboj pred enim letom. V tem primeru bi bila časovna komponenta nesmiselna, saj gre za rezultate enega turnirja, kjer se igralca ne moreta ponovno srečati.

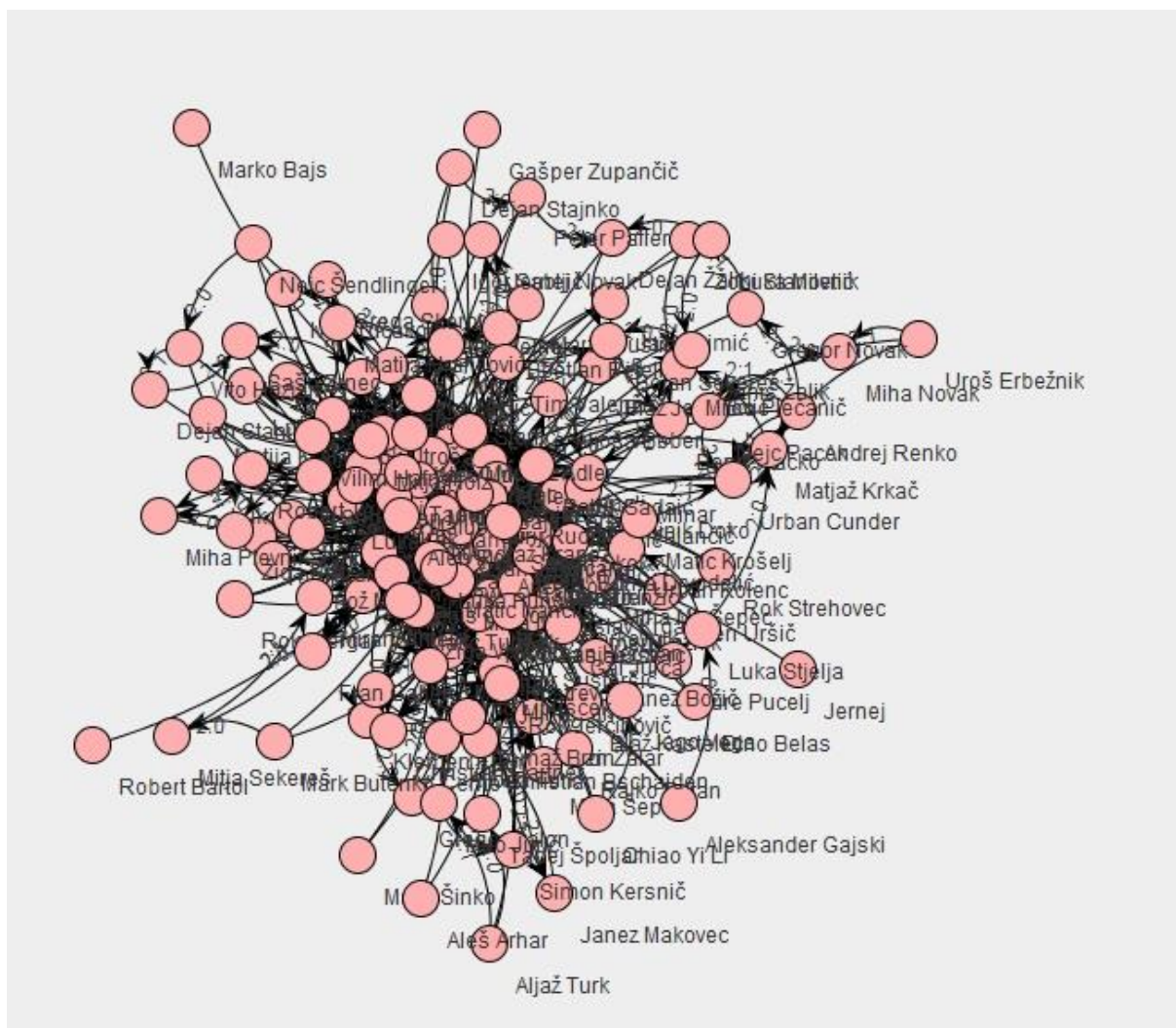
Tabela izračunanih Pearsonovih koeficientov s turnirjev za sezono 2012/2013:

Turnir	Pearsonov koeficient za PageRank	Pearsonov koeficient za uteženi PageRank	Sprememba
A1	0.7683917806491373	0.8268428285194427	0,041331132
A2	0.7469738294604790	0.7847637449435966	0,026721505
A3	0.8162063658666854	0.8393798836253626	0,016386152
A4	0.7128673675346898	0.7836667113381948	0,050062696
A5	0.7530898229579733	0.7957442432422246	0,03016123
B1	0.7943665282105483	0.8213308501228607	0,019066655
B2	0.7748275142541976	0.8093877088632673	0,024437748
B3	0.8892031989770359	0.8744522147785205	0,010430521
B4	0.7863211058888085	0.7927350474399844	0,004535342

TABELA 5.2: PREGLED IZRAČUNA PERSONOVEGA KOEFICIENTA ZA PAGERANK ALGORITEM V SKLOPU TURNIRJEV ENE SEZONE

5.2. Ocena uspešnosti atributov na celotni podatkovni množici s pomočjo programa Orange

Za množico podatkov vzamemo rezultate A in B turnirjev od septembra 2010 do januarja 2014, ki štejejo za AB lestvico BZS. Iz tega časovnega obdobja je v podatkovni množici upoštevanih 32 turnirjev. Iz pridobljenih podatkov smo izvzeli rezultate državnih prvenstev, saj ima na teh turnirjih pravico nastopati vsak državljan RS, posledično pa to pomeni, da nekaj igralcev igra samo na tem turnirju, ki sicer prav tako šteje v člansko AB lestvico, na ostalih pa ne. Še posebej to velja za najboljše slovenske tekmovalce, ki drugače igrajo v tujih klubih na različnih mednarodnih tekmovanjih, v sklopu te lestvice pa odigrajo samo DP. Za njih je namenjena SJLP, ki upošteva tudi mednarodne turnirje. Množica igralcev ima 139 elementov, kateri so igrali vsaj eno tekmo v tem obdobju, skupaj pa je bilo do konca odigranih 1318 tekem.

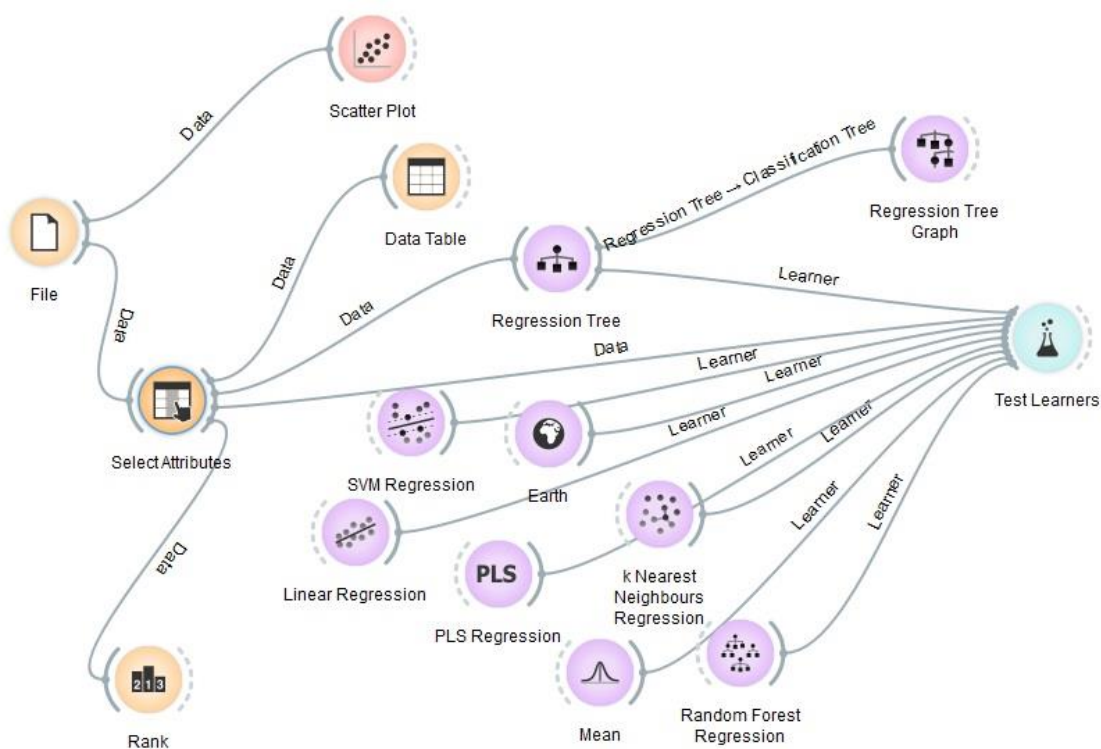


SLIKA 3: CELOTNO OMREŽJE BADMINTONSKIH TEKEM

Ker lestvica za uporabljeno obdobje ne obstaja, ampak se določa le za posamezno sezono, je potrebno ročno izračunati navidezno AB lestvico, ki bi veljala za celotno obdobje iz katerega se črpajo podatki. Na spletu je na voljo le trenutna AB lestvica, zato je potrebno izračunati lestvico še za ostala leta s pomočjo točkovnika.

Izračunana lestvica se nahaja v prilogi.

Iz zbranih podatkov ustvarimo datoteko z atributi, ki smo jo uvozili v Orange. Atributi so Število turnirjev, InDeg (število zmag), OutDeg (število porazov), InDeg-OutDeg (razlika med številom zmag in porazov), Število turnirjev, InDegree/turnir (število zmag na turnir), InDeg-OutDeg/turnir (razlika med številom zmag in porazov na turnir), PageRank Score (vrednost navadnega PageRank algoritma) in Uteženi PageRank. Poskušamo ugotoviti, kateri atributi najbolj napovejo mesto igralca na skupni lestvici in število njegovih točk. Kot validacijska metoda je uporabljeno 10-kratno prečno preverjanje.



SLIKA 4: SHEMA V PROGRAMU ORANGE

5.2.1. Napoved doseženega mesta

Najprej preverimo kateri od atributov InDeg, OutDeg, Število turnirjev, PageRank Score in Uteženi PageRank posamezno najboljše poskušajo napovedati mesto igralca. Rezultati so prikazani v tabeli Tabela 5.3:

Atribut	MAE	RAE
InDeg		
Random Forest	14.1893	0.4083
Regression Tree	17.2499	0.4964
SVM Regression	30.8662	0.8883
kNN	14.7781	0.4253
Linear Regression	21.3097	0.6133
Mean	34.9311	1.0053
OutDeg		
Random Forest	15.7279	0.4526
Regression Tree	20.5868	0.5925
SVM Regression	32.3193	0.9301
kNN	16.0030	0.4605
Linear Regression	21.9812	0.6326
Mean	34.9311	1.0053
Število turnirjev		
Random Forest	7.3204	0.2107
Regression Tree	8.4011	0.2418
SVM Regression	30.8286	0.8872
kNN	7.5571	0.2175
Linear Regression	19.5134	0.5616
Mean	34.9311	1.0053
PageRank Score		
Random Forest	21.0145	0.6048
Regression Tree	17.6412	0.5077
SVM Regression	34.9944	1.0071
kNN	17.3919	0.5005
Linear Regression	34.6540	0.9973
Mean	34.9311	1.0053
Uteženi PageRank		
Random Forest	18.8019	0.5411
Regression Tree	22.6544	0.6520

SVM Regression	32.8350	0.9449
kNN	18.7291	0.5390
Linear Regression	25.8976	0.7453
Mean	34.9311	1.0053

TABELA 5.3: NAPOVED DOSEŽENEGA MESTA S POMOČJO POSAMEZNIH ATRIBUTOV

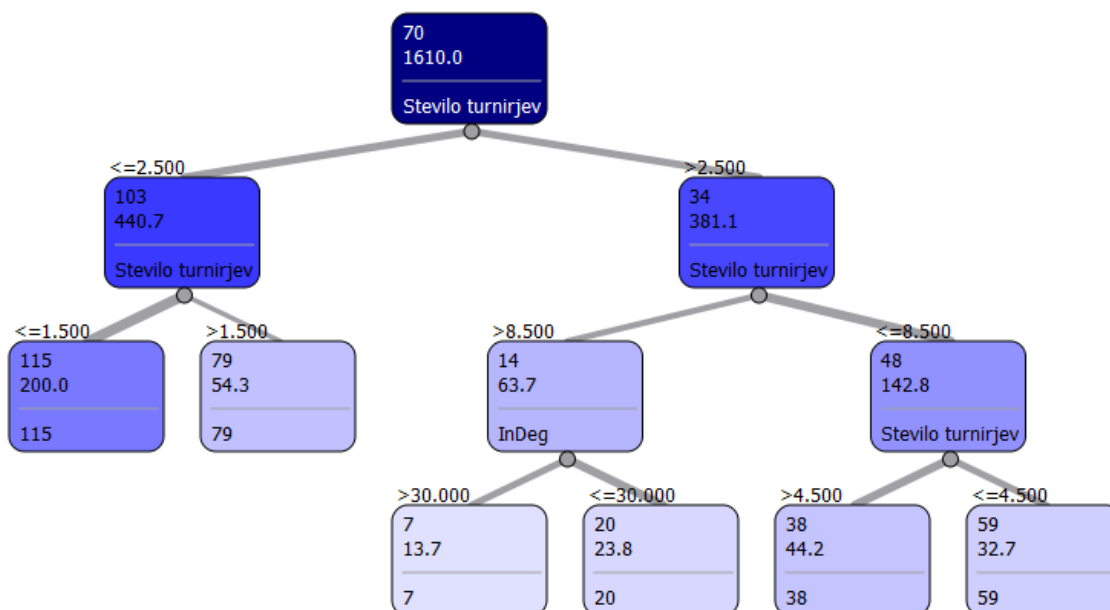
Razvidno je, da najbolj uspešno napove mesto tekmovalca atribut Število turnirjev, ki mesto napove na nekaj več kot 7 mest natančno, kar pomeni, da se napovedni model zmoti za 21,07%. Rezultat ni presenetljiv, saj je razumljivo, da bodo igralci, ki imajo odigranih več turnirjev, višje uvrščeni.

Zato nas zanima ali lahko z ostalimi atributi, ki smo jih pridobili s pomočjo analize omrežij kakorkoli izboljšamo rezultat. Izkaže se, da rezultat najboljše napove kombinacija atributov Število turnirjev, InDeg, OutDeg in Uteženi PageRank.

Metoda	MAE	RAE
Random Forest	7.5160	0.2163
Regression Tree	8.3673	0.2408
SVM Regression	31.5757	0.9087
kNN	6.3888	0.1839
Linear Regression	19.5630	0.5630
Mean	34.9311	1.0053

TABELA 5.4: NAPOVED DOSEŽENEGA MESTA ZA KOMBINACIJO ATRIBUTOV ŠTEVILO TURNIRJEV, INDEG, OUDEG IN UTEŽENI PAGERANK

S temi atributi lahko mesto napovemo na 6,38 mesta natančno, napovedni model pa se zmoti za 18,39%.



SLIKA 5: NAPOVEDNI MODEL ZA DOSEŽENA MESTA

5.2.2. Napoved točk

Na poskusu napovedovanja točk si ogledamo razliko med učinkovitostjo navadnega PageRank algoritma in uteženega PageRank algoritma.

Tudi v tem primeru najbolje deluje napovedni model z atributi Število turnirjev, InDeg, OutDeg in Uteženi PageRank, zanima pa nas, koliko slabši je napovedni model z navadnim PageRank algoritmom.

Navadni PageRank:

Metoda	MAE	RAE
Random Forest	58.1575	0.1830
Regression Tree	77.1902	0.2430
SVM Regression	269.2776	0.8475
kNN	44.5108	0.1401
Linear Regression	23.0892	0.0727
Mean	319.6028	1.0059

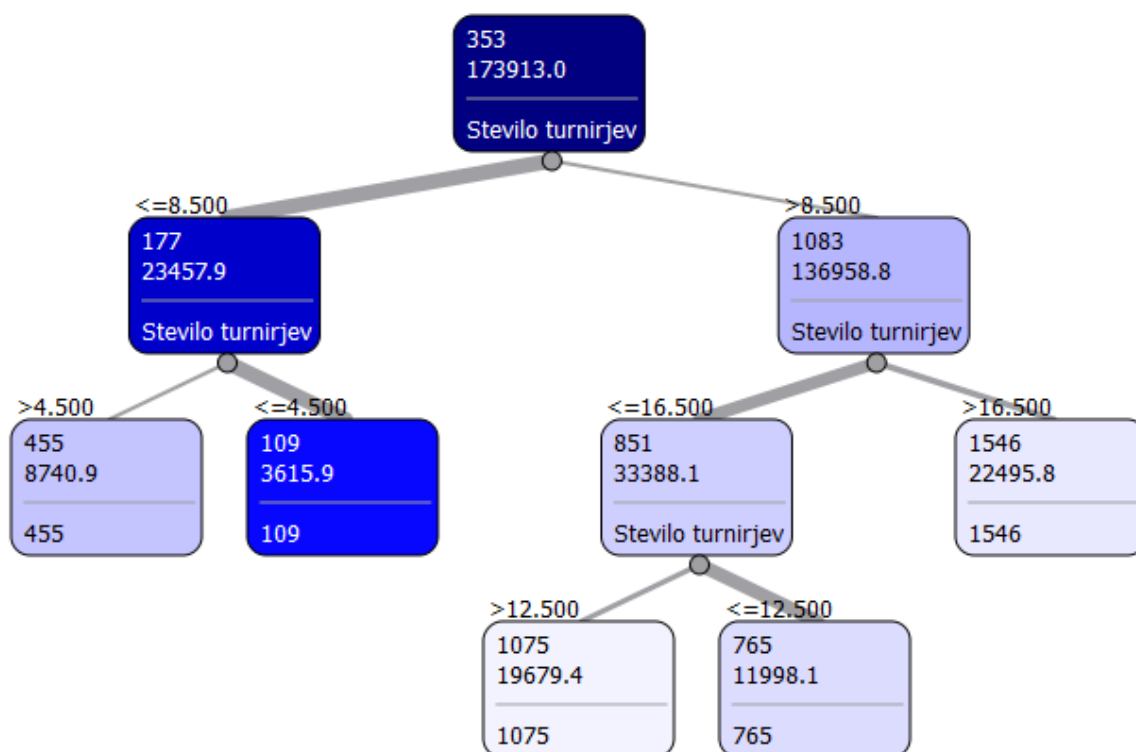
TABELA 5.5: NAPOVED TOČK ZA KOMBINACIJO ATRIBUTOV ŠTEVILO TURNIRJEV, INDEG, OUTDEG IN PAGERANK

Uteženi PageRank:

Metoda	MAE	RAE
Random Forest	54.9678	0.1730
Regression Tree	74.1638	0.2334
SVM Regression	268.8273	0.8461
kNN	40.4351	0.1273
Linear Regression	19.7204	0.0621
Mean	319.6028	1.0059

TABELA 5.6: NAPOVED ŠTEVILA TOČK ZA KOMBINACIJO ATRIBUTOV ŠTEVILO TURNIRJEV, INDEG, OUTDEG IN UTEŽENI PAGERANK

S kombinacijo atributov v kateri je navadni PageRank algoritem lahko najboljše napovemo število doseženih točk na 23 točk natančno, kjer je napaka napovednega modela 7,27%. Če pa namesto navadnega PageRank algoritma uporabimo Uteženi PageRank, se napoved izboljša na nekaj manj kot 20 točk natančno in 6,21% zmoto napovednega modela.



SLIKA 6: NAPOVEDNI MODEL ZA TOČKE

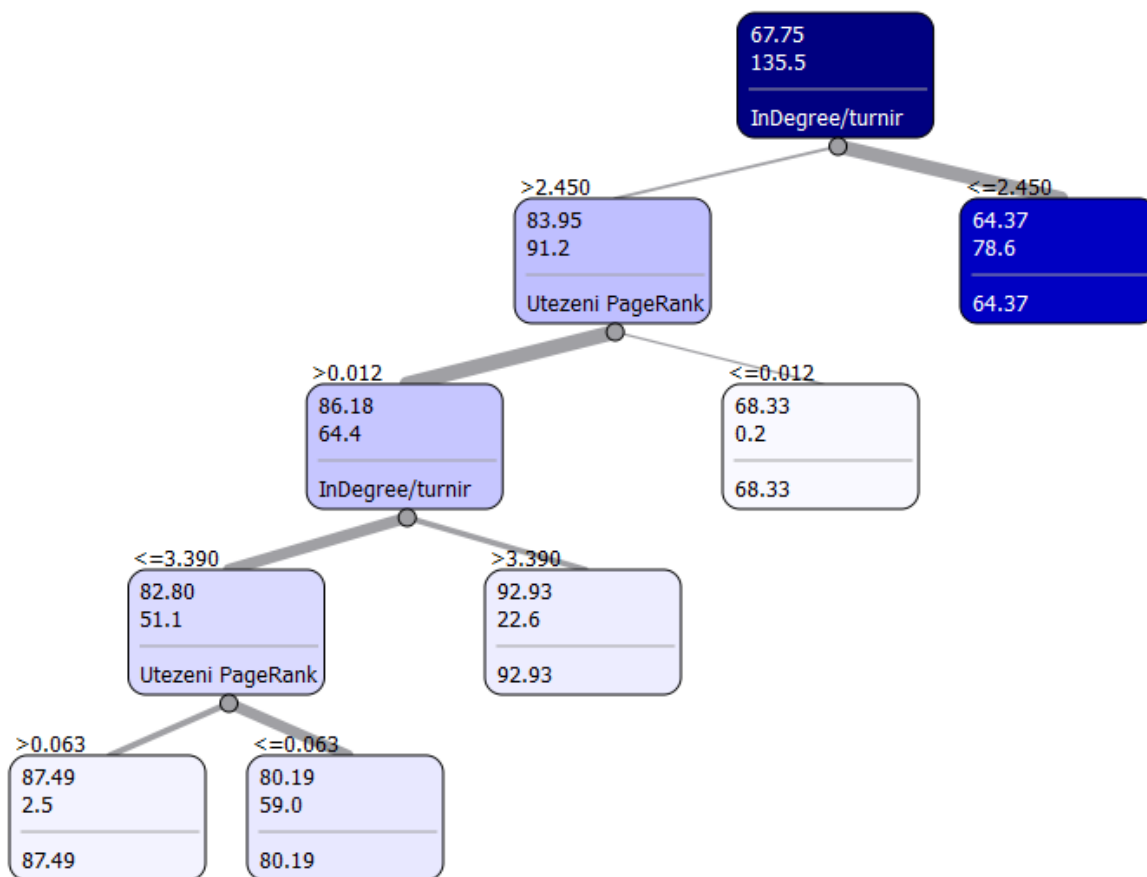
5.2.3. Napoved povprečnega števila točk igralca na posamezen turnir

Naslednja možna napoved je povprečno število točk, ki jih bo igralec dosegel na posameznem turnirju. V tem primeru so atributi, ki določajo število zmag, porazov itd. v povprečju na turnir bolj primerni od absolutnih atributov. Glede na lastnosti bi morala dobro delovati tudi oba PageRank algoritma.

Pri napovedi se najbolje obnese kombinacija atributov Število turnirjev, InDegree/turnir in Uteženi PageRank. Vendar pa rezultati, ki so prikazani v tabeli Tabela 5.7 niso najboljši. Rezultat lahko napovemo na 5,42 točke natančno, kar pomeni, da se napovedni model zmoti za 59,85%. Napoved na 5 točk natančno se zdi na prvi pogled v redu, saj je možnih točk za zmago na turnirju 100, vendar je napaka skoraj 60%. Razlog za slabši rezultat gre iskati v tem, da je razpon točk, ki jih tekmovalec doseže na turnirju majhen. Na turnirjih A kategorije je razlika med posameznim mestom dve točki in ker na turnirju igra v povprečju nekaj več kot 20 igralcev je razlika v točkah med prvim in zadnjim le okrog 35% vseh možnih točk. Ker pa je podatkovna množica sorazmerno majhna in včasih med rezultate ne štejejo vse odigrane tekme igralca, lahko pride do večje napake.

Metoda	MAE	RAE
Random Forest	6.2764	0.6931
Regression Tree	7.8542	0.8673
SVM Regression	7.3034	0.8065
kNN	6.3047	0.6962
Linear Regression	5.4203	0.5985
Mean	9.1072	1.0057

TABELA 5.7: NAPOVED POVPREČNEGA ŠTEVILA TOČK NA TURNIR ZA KOMBINACIJO ATRIBUTOV ŠTEVILO TURNIRJEV, INDEGREE/TURNIR IN UTEŽENI PAGERANK



SLIKA 7: NAPOVEDNI MODEL ZA ŠTEVILO TOČK NA POSAMEZEN TURNIR

Poglavje 6

6. Primerjava rezultatov in zaključek

Iz rezultatov gre sklepati, da atributi, ki jih pridobimo z analizo omrežja ne moremo dovolj natančno napovedati mesta, nekoliko bolje pa lahko napovemo točke posameznega tekmovalca. Razlog, zakaj PageRank algoritem v našem primeru ne prikaže realnega stanja na lestvici, je v tem, da imamo za razredni atribut absolutno število točk, ki jih je nek igralec osvojil v tem obdobju. Tako na vrhu lestvice niso nujno najboljši igralci, ampak v večini tisti, ki imajo odigranih največ turnirjev. Naslednja težava je, da je število turnirjev, ki so jih igralci odigrali v tem obdobju, zelo različno. Nekateri so odigrali skoraj vse turnirje, spet drugi pa samo enega ali dva. Poleg tega najboljši slovenski igralci igrajo v tujini in praviloma domače turnirje izpuščajo.

Rešitev bi bila podatkovna množica, ki bi temeljila na daljšem časovnem obdobju, kar v tem primeru ni bilo mogoče, saj se v Sloveniji program Tournament Planner uporablja šele v zadnjem času. Tako so javni podatki o rezultatih turnirjev na voljo le za zadnjih nekaj let in še to ne za povsem vse turnirje. Idealno bi bilo, da bi vsi slovenski igralci, torej tudi tisti, ki igrajo v tujini, igrali vse turnirje pod okriljem BZS. To realno ni možno, saj tako kot v tenisu vsi igralci ne igrajo na vseh turnirjih, ampak samo na izbranih. Vendar za razliko od svetovnega tenisa, kjer je na voljo več deset turnirjev na sezono, je v slovenskem badmintonu turnirjev le okrog deset, od tega približno polovica v najkakovostnejši kategoriji, čeprav igralci iz A kategorije redno igrajo tudi na turnirjih B kategorije, še posebej ko se iščejo točke za čim višjo uvrstitev na končni lestvici.

Ko bodo na voljo podatki za daljše časovno obdobje, bo uporabna tudi časovna komponenta za uravnoteženje povezav glede na čas tekme. Izboljšanje prinese, posebej za način tekmovalca, ki se igra pri badmintonu, tudi dodatna obtežitev povezav, ki je uporabljena v tej nalogi. Kot je bilo nakazano te povezave niso vedno pravilno utežene zaradi različnega števila veljavno odigranih tekem posameznega tekmovalca. Če bi bilo možno tudi iz predane ali neodigrane tekme iz podatkov določiti, kdo je zmagovalec dvoboja, bi se uteženi PageRank obnesel bolje.

Menimo, da stopnja vozlišča in algoritem PageRank dobro pokažeta, kako dober je v resnici igralec. Vendar pa s temi metrikami težko napovemo rezultate na tekmah slovenskega badmintona. Za to je predvsem pomembna dovolj velika množica podatkov, v kateri so igralci, ki so igrali na podobnem številu turnirjev. S časovno komponento in dodatnim

obteževanjem povezav glede na pomembnost tekme lahko dosežemo še nekoliko boljšo napoved, vendar pa je glavni predpogoj, da je množica dovolj velika in primerno porazdeljena.

Tabele

Tabela 3.1: Točkovanje za različne vrste turnirjev	10
Tabela 3.2: Razpored uporabljenih turnirjev za sezono 2010/2011	10
Tabela 3.3: Razpored uporabljenih turnirjev za sezono 2011/2012	11
Tabela 3.4: Razpored uporabljenih turnirjev za sezono 2012/2013	11
Tabela 3.5: Razpored uporabljenih turnirjev za sezono 2013/2014	11
Tabela 5.1: Analiza posameznega turnirja	25
Tabela 5.2: Pregled izračuna Personovega koeficienta za PageRank algoritem v sklopu turnirjev ene sezone	26
Tabela 5.3: Napoved doseženega mesta s pomočjo posameznih atributov.....	30
Tabela 5.4: Napoved doseženega mesta za kombinacijo atributov Število turnirjev, InDeg, OuDeg in Uteženi PageRank	30
Tabela 5.5: Napoved točk za kombinacijo atributov Število turnirjev, InDeg, OutDeg in PageRank	31
Tabela 5.6: Napoved števila točk za kombinacijo atributov Število turnirjev, InDeg, OutDeg in Uteženi PageRank.....	32
Tabela 5.7: Napoved povprečnega števila točk na turnir za kombinacijo atributov Število turnirjev, InDegree/Turnir in Uteženi PageRank.....	33

Slike

Slika 1: Primer neusmerjenega in usmerjenega omrežja	13
Slika 2: Primer grafa odigranih tekem za posamezen turnir	24
Slika 3: Celotno omrežje badmintonskih tekem.....	27
Slika 4: Shema v programu Orange	28
Slika 5: Napovedni model za dosežena mesta	31
Slika 6: Napovedni model za točke.....	32
Slika 7: Napovedni model za število točk na posamezen turnir.....	34

Priloge

Izračunana lestvica AB turnirjev v obdobju od septembra 2010 do januarja 2014:

Mesto	Tekmovalec	Število turnirjev	Tocke	InDeg	OutDeg	Tocke/Turnir	Mesto T/T
1	Jurij Rudolf	26	1806	40	53	69,46	51
2	Dani Klančar	21	1658	42	40	78,95	30
3	Sebastijan Miklič	22	1629	46	42	74,05	10
4	Milko Pečanič	23	1589	40	39	69,09	60
5	Primož Flis	18	1560	52	22	86,67	26
6	Matic Ivančič	22	1518	38	49	69,00	39
7	Mitja Šemrov	17	1510	47	27	88,82	19
8	Martin Adler	18	1396	43	30	77,56	7
9	Anže Novak	18	1248	38	30	69,33	32
10	Alen Roj	13	1246	47	7	95,85	13
11	Žan Laznik	16	1215	15	6	75,94	91
12	Simon Skok	16	1079	31	28	67,44	92
13	Matevž Bajuk	11	972	38	7	88,36	28
14	Tadej Jezernik	12	948	25	19	79,00	5
15	Aljaž Roškar	14	929	23	31	66,36	93
16	Rok Čampelj	13	907	21	28	69,77	20
17	Tomaz Uršič	10	844	19	17	84,40	52
18	Luka Pungerčar	12	826	23	26	68,83	17
19	Kek Jamnik	9	808	29	10	89,78	94
20	Matjaž Adler	9	768	26	9	85,33	14
21	Miha Ivanič	10	758	27	14	75,80	2
22	Andraž Krapež	10	756	25	18	75,60	95
23	Nejc Zaviršek	12	705	14	24	58,75	96
24	Aljaž Ferk	9	667	19	19	74,11	97
25	Martin Cerkovnik	9	667	23	14	74,11	8
26	Josip Uglesič	7	650	27	6	92,86	49
27	Vanja Osterc	9	625	20	13	69,44	98
28	Urban Turk	7	608	22	6	86,86	99
29	Jaka Šusteršič	9	606	11	21	67,33	11
30	Luka Petrič	6	588	24	1	98,00	21
31	Nejc Juršič	8	533	14	18	66,63	22
32	Nejc Čeplak	6	532	18	8	88,67	69
33	Tomislav Krga	8	531	12	19	66,38	70
34	Valentin Dupouey	7	510	19	13	72,86	24
35	Grega Valančič	7	483	7	14	69,00	25
36	Miha Ivančič	8	461	9	21	57,63	3
37	Urban Herman	7	459	9	19	65,57	55
38	Rok Jerčinovič	7	458	8	14	65,43	100
39	Ales Murn	5	454	18	5	90,80	101
40	Robert Špoljarec	6	442	20	10	73,67	40

41	Filip Malus	7	428	5	20	61,14	34
42	Žiga Vračar	8	421	1	8	52,63	71
43	Niko Alič	6	420	14	8	70,00	102
44	Denis Čleković	6	401	8	14	66,83	103
45	Juš Mlinar	6	352	2	15	58,67	43
46	Damjan Repar	5	324	12	9	64,80	104
47	Mitja Holz	5	323	8	10	64,60	105
48	Igor Šušteršič	5	320	5	14	64,00	106
49	Juraj Denzić	4	310	7	7	77,50	107
50	Matic Turinek	5	307	8	10	61,40	16
51	Iztok Utroša	3	298	11	1	99,33	72
52	Samo Lipušček	3	256	12	1	85,33	1
53	Domen Laznik	4	247	6	11	61,75	27
54	Gal Jurca	4	232	1	12	58,00	9
55	Jure Stanič	3	222	6	6	74,00	4
56	Patrik Sadaic	3	207	2	10	69,00	6
57	Vasja Osterc	4	197	2	9	49,25	35
58	Mišo Žalik	3	196	6	6	65,33	56
59	Tilen Zalar	3	191	4	6	63,67	73
60	Miha Šepec ml.	2	190	5	3	95,00	108
61	Vilm Hafner	3	184	2	9	61,33	18
62	Borna Drvodelič	3	181	2	8	60,33	74
63	Aleks Miklič	3	176	2	7	58,67	109
64	Primož Makuc	3	174	1	8	58,00	110
65	Robert Božič	3	173	1	8	57,67	111
66	Tim Valenčič	3	168	2	8	56,00	12
67	Janez Božič	3	155	0	8	51,67	29
68	Matic Krošelj	3	154	1	5	51,33	44
69	Matevž Šrekl	2	150	3	4	75,00	31
70	Blaž Kastelec	2	149	3	4	74,50	33
71	Christian Bscheiden	2	144	3	5	72,00	15
72	Franc Horvat	2	139	2	4	69,50	75
73	Tomaž Brzin	2	138	3	4	69,00	76
74	Nejc Pacek	2	136	3	4	68,00	77
75	Tilen Uršič	2	132	0	4	66,00	37
76	Dominik Doko	2	132	2	6	66,00	38
77	Urban Cunder	2	132	4	3	66,00	58
78	Matjaž Krkač	2	130	4	3	65,00	78
79	Janez Makovec	2	126	4	3	63,00	46
80	Urban Kolenc	2	126	3	5	63,00	47
81	Luka Miletič	2	126	1	3	63,00	48
82	Robert Tomažič	2	122	2	3	61,00	112
83	Jon Schwarzmann	2	120	1	6	60,00	113
84	Mitja Kostrevc	2	117	0	6	58,50	114
85	Dejan Stanjko	2	114	1	2	57,00	115
86	Andrej Renko	2	112	1	5	56,00	116

87	Luka Stelja	2	112	2	4	56,00	117
88	Peter Paller	2	108	1	3	54,00	59
89	Miran Vihtelič	2	90	2	6	45,00	79
90	Dejan Glazer	2	90	2	9	45,00	80
91	Fran Galjer	1	88	3	2	88,00	81
92	Matija Hranilović	1	88	2	2	88,00	118
93	Luka Vlahek	1	86	2	2	86,00	119
94	Simon Kersnič	1	80	1	2	80,00	120
95	Miro Jurič	1	78	1	3	78,00	121
96	Jure Kelher	1	78	3	1	78,00	53
97	Grega Skerbiš	1	78	1	2	78,00	50
98	Tadej Špoljar	1	76	1	2	76,00	61
99	Miha Novak	1	76	2	1	76,00	41
100	Jure Pucelj	1	74	0	2	74,00	82
101	Jernej Novak	1	74	2	2	74,00	62
102	Aljaž Turk	1	72	0	3	72,00	83
103	Žoni Stanovnik	1	72	1	2	72,00	122
104	Klemen Kramar	1	70	2	2	70,00	123
105	Boštjan Petelinc	1	70	1	3	70,00	124
106	Aljoša Beber	1	70	1	3	70,00	125
107	Gregor Novak	1	70	1	3	70,00	23
108	Christian Hartner	1	69	4	1	69,00	45
109	Igor Sabljic	1	68	0	2	68,00	63
110	Bojan Sekereš	1	68	4	1	68,00	84
111	Sašo Zrnec	1	68	3	1	68,00	54
112	Matic Šinko	1	64	0	2	64,00	64
113	Borut Vračko	1	64	1	2	64,00	126
114	Gregor Balon	1	64	1	3	64,00	127
115	Ivan Tucaković	1	64	1	2	64,00	65
116	Denis Žalik	1	64	2	2	64,00	36
117	Žiga Bec	1	64	2	2	64,00	85
118	Jaša Dimic	1	63	0	3	63,00	128
119	Rok Strehovec	1	62	1	2	62,00	66
120	Jago Vega	1	62	1	2	62,00	86
121	Uroš Erbežnik	1	62	0	3	62,00	87
122	Miha Plevnik	1	60	2	2	60,00	129
123	Dejan Žalik	1	60	2	2	60,00	130
124	Rajko Petan	1	60	0	2	60,00	88
125	Chiao Yi Li	1	59	0	3	59,00	131
126	Aleksander Gajevski	1	58	0	3	58,00	132
127	Vito Hozjan	1	58	1	3	58,00	133
128	Dino Belas	1	57	0	3	57,00	42
129	Nejc Šendlinger	1	56	0	3	56,00	134
130	Blaž Jerčinovič	1	56	0	3	56,00	135
131	Aljoša Turk	1	54	1	2	54,00	67
132	Marko Bajs	1	54	0	1	54,00	68

133	Martin Justin	1	54	0	3	54,00	136
134	Robert Bartolj	1	52	0	1	52,00	137
135	Rok Stergar	1	52	0	3	52,00	57
136	Mitja Sekereš	1	50	1	2	50,00	89
137	Matija Kraljić	1	50	0	3	50,00	90
138	Aleš Arhar	1	44	0	2	44,00	138
139	Mark Butenko Černe	1	44	0	3	44,00	139

Literatura

- [1] F. Radicchi, Who Is the Best Player Ever? A Complex Network Analysis of the History of Professional Tennis. Dostopno na:
<http://filrad.homelinux.org/My Papers/journal.pone.0017249.pdf>
(dostop: 7. marec 2014)
- [2] A. Krašček, B. Milosheška in Ž. Zupanec, Napovedovanje izidov teniških dvobojev s pomočjo analize omrežja, 2013
- [3] Eclipse. Wikipedia. Dostopno na:
http://en.wikipedia.org/wiki/Eclipse_%28software%29
(dostop: 8. marec 2014)
- [4] J. O'Madadhain, D. Fisher, S. White, P. Smyth in Y.-B. Boey, Analysis and visualization of network data using JUNG, Journal of Statistical Software, 2005
- [5] JUNG. Dostopno na:
<http://jung.sourceforge.net/>
(dostop: 8. marec 2014)
- [6] Orange. Dostopno na:
<http://orange.biolab.si/>
(zadnji dostop: 8. marec 2014)
- [7] Badmintonzveza Slovenije, Badmintonzveza pravila, 2006. Dostopno na:
http://www.badminton-zveza.si/wp-content/uploads/2012/11/badmintonzveza_pravila2006.pdf
(dostop: 8. marec 2014)
- [8] M. Kersnik, Zgodovina badmintona v Sloveniji, diplomsko delo, Fakulteta za šport, 2006. Dostopno na:
<http://www.badminton-zveza.si/wp-content/uploads/2012/11/Diploma22050090KersnikMaja.pdf>
- [9] Badminton Europe. Dostopno na:
<http://badmintoneurope.com/CMS/?cmsid=239&pageid=5730>
(dostop: 8. marec 2014)
- [10] Badmintonzveza Slovenije, Tekmovalni pravilnik. Dostopno na:
<http://www.badminton-zveza.si/wp-content/uploads/2012/11/Tekmovalni-pravilnik2013.pdf>
(dostop: 8. marec 2014)
- [11] Tournament software. Dostopno na:
<http://www.tournamentsoftware.com/>
(zadnji dostop: 8. marec 2014)
- [12] L. Šubelj, N. Blagus, Š. Furlan, B. Klemenc, A. Kumer, D. Lavbič, A. Zrnc, S. Žitnik in M. Bajec, Analiza kompleksnih omrežij: Osnovni pojmi in primeri uporabe v praksi, 2010
- [13] M. E. J. Newman, Networks: An Introduction, Oxford University Press, 2010
- [14] D. Easley, J. Kleinberg, Networks, Crowds, and Markets: Reasoning about a Highly Connected World, Cambridge University Press, 2010
- [15] Data Mining. Wikipedia. Dostopno na:

http://en.wikipedia.org/wiki/Data_mining
(dostop: 9. marec 2014)

- [16] A. O. Sykes, An Introduction To Regregration Analysis. Dostopno na:
http://www.law.uchicago.edu/files/files/20.Sykes_.Regression.pdf
(dostop: 9. marec 2014)
- [17] K. M. Leung, k-Nearest Neighbor Algorithm for Classification, 2007. Dostopno na:
<http://cis.poly.edu/~mleung/FRE7851/f07/k-NearestNeighbor.pdf>
(dostop: 9. marec 2014)
- [18] Z. Yao, A Regression-based K nearest neighbor algorithm for gene function prediction from heterogeneous data, 2006. Dostopno na:
<http://www.biomedcentral.com/1471-2105/7/S1/S11>
(dostop: 9. marec 2014)
- [19] L. Breiman, Random Forests, 2001. Dostopno na:
<http://oz.berkeley.edu/~breiman/randomforest2001.pdf>
(dostop: 9. marec 2014)