

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Matic Tribušon

**Lastnosti športnih tekmovanj in
povezanost s tržno zanimivostjo**

MAGISTRSKO DELO

MAGISTRSKI PROGRAM DRUGE STOPNJE
RAČUNALNIŠTVO IN INFORMATIKA

MENTOR: doc. dr. Lovro Šubelj
SOMENTOR: doc. dr. Erik Štrumbelj

Ljubljana, 2017

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

©2017 MATIC TRIBUŠON

ZAHVALA

Zahvaljujem se mentorju doc. dr. Lovru Šublju in somentorju doc. dr. Eriku Štrumblju za vso strokovno pomoč, nasvete, popravke in mentorstvo pri izdelavi in pisanju magistrskega dela.

Zahvaljujem se tudi staršem za vso podporo tekom študija.

Matic Tribušon, 2017

Kazalo

Povzetek

Abstract

1	Uvod	1
1.1	Problemska domena	3
1.2	Začetne hipoteze	4
2	Pregled področja	7
3	Metodologija	17
3.1	Pridobivanje in obdelava podatkov	17
3.2	Gradnja omrežij tekmovanj	22
3.3	Analiza omrežij	23
3.4	Masseyjeva metoda	30
4	Analiza in rezultati	35
4.1	Analiza izenačenosti športnih tekmovanj	36
4.2	Analiza korelacije med zanimivostjo in izenačenostjo	51
4.3	Primerjava z Masseyjevo metodo	55
5	Sklepne ugotovitve	59
A	Relativna entropija PageRank	63
B	Korelacija med izenačenostjo in zanimivostjo	73

Seznam uporabljenih kratic

kratica	angleško	slovensko
MLB	major league baseball	severnoameriška liga bejzbola
NFL	national football league	severnoameriška liga ameriškega nogometa
NBA	national basketball association	severnoameriška košarkarska liga
NHL	national hockey league	severnoameriška hokejska liga
PDF	probability density function	funkcija gostote verjetnosti
CDF	cumulative distribution function	zbirna funkcija verjetnosti

Povzetek

Naslov: Lastnosti športnih tekmovanj in povezanost s tržno zanimivostjo

Šport je povsod po svetu izjemno priljubljena vrsta zabave. Posledično je zelo pomemben tudi s poslovnega vidika. V delu raziskujemo vpliv izenačenosti športnega tekmovanja na njegovo tržno zanimivost. Cilj je ugotoviti izenačenost različnih športnih tekmovanj in raziskati korelacijo med izenačenostjo ter različnimi metrikami njihove tržne zanimivosti. V delu predlagamo metodo za ocenjevanje izenačenosti poljubnega ekipnega športnega tekmovanja. Metoda temelji na omrežni analizi in prilagojenem algoritmu PageRank. Kot vhod potrebuje le rezultate odigranih tekem v opazovanem obdobju, njen izhod pa je indeks izenačenosti športnega tekmovanja v tem obdobju. Predlagano metodo primerjamo z Masseyjevo metodo in predstavimo rezultate. Prav tako predstavimo različne metrike tržne zanimivosti športnih tekmovanj in izračunamo korelacijo med mero izenačenosti ter metrikami tržne zanimivosti izbranih športnih tekmovanj. Rezultati analize izenačenosti potrjujejo hipotezo, da so tekmovanja v Severni Ameriki bolj izenačena. Iz rezultatov je razvidno, da je bilo izbrano časovno obdobje prekratko za uspešen izračun korelacije na ravni sezone, saj so intervali zaupanja preveliki. Če izračune ponovimo na ravni posamezne tekme, je podatkov dovolj, korelacija pa potrjuje začetno hipotezo, da so bolj izenačena tekmovanja tržno zanimivejša.

Ključne besede

*omrežna analiza, PageRank, športna tekmovanja, izenačenost, tržna zanimi-
vost*

Abstract

Title: Properties of sports competition and correlation to market value

Sports are a very popular type of entertainment around the world. Consequently, they have become very important from a business perspective. In this thesis, we research the influence of sports competition competitive balance on its market value. The aim is to determine the competitive balance of various sports competitions and explore the correlation between competitive balance and market value of sports competitions. We propose a method for assessing competitive balance of an arbitrary team sport competition. The method is based on network analysis and a version of the PageRank algorithm. As an input, it only requires results of played matches in the observed sports competition and the output is an index of competitive balance. We compare the proposed method with Massey's method and evaluate the results. We also propose different metrics that are related to market value of sports competitions and research the correlation between the competitive balance, obtained with the proposed method, and the mentioned metrics. The results of competitive balance analysis confirm the hypothesis that North American sports competitions are more balanced. From the results of correlation analysis it is clear that the chosen time period was too short for per-season correlation analysis. Thus, the confidence intervals of correlation results are too wide. Correlation analysis on per-game basis confirms the hypothesis that competitive balance of sports competitions has a positive effect on their market value.

Keywords

network analysis, PageRank, sports competitions, competitive balance, market value

Poglavje 1

Uvod

Zgodovina športa sega v zgodnjo dobo človeštva. Z leti je popularnost nekaterih športov skokovito rasla in tekmovanja so dobivala vedno večjo težo tudi z vidika poslovanja ter zaslužkov. Vrednost športnega trga naj bi samo v Severni Ameriki do leta 2019 zrasla na 73,5 milijarde ameriških dolarjev [1]. Zaradi razmaha poslovanja, ki je povezano s športom, se je le-ta lahko zelo hitro razvijal in napredoval. Že nekaj desetletij ni nenavadno, da so najboljši športniki med najboljše plačanimi ljudmi na svetu [2]. Kljub izjemno velikim odhodkom, ki jih predstavljajo plače igralcev, so najuspešnejše ekipe in klubi priljubljenih športov, kot sta nogomet in košarka, med najbolj dobičkonosnimi organizacijami. Zaslužke si zagotavljajo ne le z uporabo klasičnih kanalov, kot so prodaja vstopnic in televizijskih pravic, temveč tudi z razvojem in trženjem blagovne znamke [3]. Tako kot v vseh vejah poslovanja, je glavni cilj športnih klubov povečanje prihodkov oziroma dobička. Osnova za povečanje dobička športnega kluba je njegova prepoznavnost, to pa je najlažje doseči s konkurenčnimi rezultati in zvenečimi imeni v igralskem kadru. Vse to zahteva visoke denarne vložke, ki v najbolj popularnih tekmovanjih različnih ekipnih športov iz leta v leto skokovito rastejo. Vse športne ekipe torej stremijo k čim večji kakovosti, kar jim na dolgi rok prinese več prihodkov. Poleg ekip so tu še krovne zveze ali lige tekmovanj, v katerih ekipe nastopajo. Tudi te so v zadnjih desetletjih uspele znatno povečati prihodke

in so izjemno dobičkonosne [4]. Prihodke si zagotavljajo s trženjem tekmovanja, torej s trženjem ekip, ki nastopajo v tem tekmovanju, kjer je glavni vir prihodkov vezan na gledanost tekmovanja. V zadnjih nekaj desetletjih se je pojavilo več del in raziskav, ki se ukvarjajo z ugotavljanjem korelacije med izenačenostjo tekmovanja in njegovo gledanostjo [5]. Zaradi pozitivnih rezultatov tovrstnih raziskav, si vodstvo tekmovanja prizadeva ustvariti čim bolj izenačeno in napeto tekmovanje.

Vedno večja dostopnost športnih podatkov [6] v zadnjih letih je privedla do povečanega obsega števila del, ki se ukvarjajo z različnimi pristopi analize športnih podatkov. Večina se osredotoča na analizo posameznih igralcev znotraj določene ekipe ali na analizo kakovosti posamezne ekipe znotraj določenega tekmovanja. V delu smo se osredotočili na širšo analizo tekmovanj različnih ekipnih športov z vidika izenačenosti skozi časovno obdobje večih sezon. Analiza temelji na izsledkih analize omrežij. Omrežja, ki predstavljajo posamezno tekmovanje v opazovani sezoni, so zgrajena iz vozlišč, ki predstavljajo posamezne ekipe tekmovanja in povezav, ki predstavljajo odigrane tekme med povezanima ekipama. Omrežja so torej zgrajena na podlagi podatkov o posameznih tekmah določenega tekmovanja v obdobju več sezon. Prvi del naloge predstavi pridobivanje podatkov in gradnjo omrežij tekmovanj različnih ekipnih športov v več opazovanih sezonah. V drugem delu se osredotočimo na analizo zgrajenih omrežij, katere cilj je ugotoviti izenačenost posameznega tekmovanja v posamezni sezoni. Po pridobivanju podatkov, ki kažejo na tržno zanimivost tekmovanja, v zadnjem delu naloge z različnimi metodami ugotavljamo korelacijo med izenačenostjo izračunano v drugem delu in pridobljenimi podatki tržne zanimivosti tekmovanja.

Razvili smo več metod za analizo ekip znotraj tekmovanja poljubnega ekipnega športa. Na podlagi omenjenih metod smo razvili še metodo za analizo izenačenosti tekmovanja kot celote, z upoštevanjem analize posameznih ekip znotraj tekmovanja. Metoda temelji na različici priljubljenega algoritma za omrežno analizo PageRank [7], ki smo ga prilagodili specifikam analiziranih omrežij. Metoda je dovolj splošna, da jo je mogoče uporabiti za analizo

omrežij tekmovanj različnih ekipnih športov, kar v delu tudi pokažemo. Razvita metoda služi kot osnova za ugotavljanje korelacije med izenačenostjo tekmovanja in tržno zanimivostjo in s tem predstavlja osnovo za potrditev ali zavrnitev osnovne hipoteze dela.

1.1 Problemska domena

Kot smo že omenili, so določene raziskave izenačenosti in gledanosti športnih tekmovanj potrdile korelacijo med omenjenima metrikama. Vodstva športnih tekmovanj si zato prizadevajo ohraniti čim višjo stopnjo izenačenosti med ekipami, kar naj bi privedlo do večje zanimivosti tekmovanja in posledično večjih prihodkov. Pristopi za doseganje omenjenih ciljev se od tekmovanja do tekmovanja razlikujejo. Najbolj radikalni so v Severni Ameriki. Tam so v košarkarski ligi NBA že leta 1947 uvedli nabor igralcev (angl. *draft*), kjer najslabše ekipe končane sezone prve izbirajo na naboru novincev [8]. Leta 1984 so uvedli še pravilo, ki ekipam omejuje količino denarja, ki ga lahko namenijo za plače (angl. *salary cap*) [9]. Ekipe, ki za plače namenijo največ denarja, pa od leta 2002 [10] plačujejo tudi t.i. davek na luksuz (angl. *luxury tax*), ki se razdeli med ekipe, ki za plače igralcev namenijo najmanj denarja. Vsa ta pravila pripomorejo k manjšanju razlik med najboljšim in najslabšim ekipo v tekmovanju, vsaj kar se tiče njihovega izhodišča. Takšna pravila imajo lahko tudi negativne posledice, saj lahko privedejo do visoke stopnje kalkuliranja ekip, saj jim je lahko poraz predvsem v izdihljajih sezone vreden več kot zmaga. Z leti je vodstvo prilagajalo določene podrobnosti takih pravil, da bi se znebilo negativnega vpliva [11]. V Evropi je takih pravil precej manj oziroma jih sploh ni. V zadnjih letih se pojavljajo približki zgoraj omenjenih ukrepov, ki jih v Ameriki izvajajo že več desetletij. V evropskem nogometu se je nedavno pojavil t.i. finančni *fair play*, ki ekipam preprečuje prekomerno zapravljanje denarja v namene nakupa novih igralcev. V nalogi preučujemo, ali omenjeni ukrepi vplivajo na izenačenost tekmovanja in v kolikšni meri. Prav tako nas zanima, kako se različna tekmovanja primerjajo med seboj po

izenačenosti in kaj se je z izenačenostjo tekmovanj dogajalo skozi čas. Glavni del naloge zajema razvoj in uporabo metodologij za ugotavljanje in analizo izenačenosti športnih tekmovanj. Omejili smo se na ekipne športe, cilj dela pa je razvoj splošne metode, s katero lahko analiziramo ekipna športna tekmovanja različnih športov na nivoju ene sezone. Vhodni podatki so zgolj podatki rezultatov odigranih tekem, rezultat metode pa je indeks izenačenosti določenega športnega tekmovanja v opazovani sezoni. Že iz naslova dela je razvidno, da je analiza izenačenosti le del naloge, saj nas zanima ugotavljanje korelacije med izračunanim indeksom izenačenosti tekmovanja in njegovo tržno zanimivostjo. Tržna zanimivost tekmovanja je zelo širok pojem in le stežka bi izbrali eno metriko, ki odraža tržno zanimivost. Pridobili smo podatke več metrik, ki bi lahko odražale tržno zanimivost, jih primerno obdelali in z različnimi metodami ugotavljali korelacijo med obdelanimi podatki tržne zanimivosti in izračunano izenačenostjo športnega tekmovanja.

1.2 Začetne hipoteze

V poglavju 1.1 so opisane velike razlike v pravilih, ki spodbujajo izenačenost ekip med tekmovanji v evropskem in ameriškem prostoru. Te razlike se najverjetneje odražajo v izenačenosti tekmovanj in v tem delu bomo z razvito metodo za ugotavljanje izenačenosti med drugim poskusili to hipotezo potrditi. Omenjena pravila imajo lahko tudi negativne vplive na izenačenost.

Vsa opazovana ameriška tekmovanja so razdeljena na redni del in končnico. V končnico se uvrsti le določen delež najboljših ekip rednega dela, za ostale je sezona po rednem delu končana. Ker nimajo koncepta izpada v nižje lige, kar je praviloma prisotno v evropskih športnih ligah, je ekipam, ki ne tekmujejo v končnici vseeno, na katerem mestu končajo sezono. Zaradi pravil, kot je nabor igralcev, pa je zanje celo boljše, da sezono končajo nižje, saj si s tem zagotovijo boljši izbor novincev v naslednji sezoni. Posledica tega je nemalokrat načrtno izgubljanje tekem konca sezone, ko ekipe nimajo več možnosti uvrstitve v končnico tekmovanja. To lahko ob zaostrenem boju za mesta pri

vrhu in mesta, ki še vodijo v končnico, negativno vpliva na izenačenost. S tem se namreč povečajo točkovne razlike med ekipami, ki se borijo za uvrstitev v končnico in najslabšimi ekipami na dnu točkovne lestvice.

Namen drugega dela naloge je ugotavljanje korelacije med različnimi metriki tržne zanimivosti in izenačenostjo tekmovanja. Naša hipoteza je, da obstaja korelacija med tržno zanimivostjo in izenačenostjo tekmovanja, dočim pa je zaradi številnih dejavnikov, ki vplivajo na izenačenost, hipoteza težko dokazljiva. Pri ugotavljanju korelacije je zelo pomembna izbira metrik, ki naj bi predstavljale tržno zanimivost, saj je nabor možnosti precej širok. Nekatere metrike so lahko zaradi različnih dejavnikov neprimerne ali zelo pristranske. Kot metrike tržne zanimivosti smo izbrali obiskanost tekem, povprečje vsote tržne vrednosti igralcev posamezne ekipe v tekmovanju in volumen stav na tekme.

Poglavje 2

Pregled področja

V zadnjih letih se je omrežna analiza uveljavila kot zelo primerna in priljubljena metoda reševanja problemov na različnih področjih. V raziskavah se pogosto pojavi potreba identifikacije entitet z določenimi lastnostmi v izjemno veliki množici entitet ali rangiranje večje množice entitet. V takih primerih lahko za uspešno analizo omrežja uporabimo metode naključnih sprehodov po omrežju. Naključni sprehodi so znani in raziskovani že vrsto let, dočim obstaja več različic in implementacij [12]. Če se omejimo na problem rangiranja entitet v omrežju, je osnova tudi tu navadno modeliranje naključnega sprehoda. Vse metode za rangiranje entitet v omrežju v osnovi uporabljajo neko vrsto naključnega sprehoda, razlikujejo pa se v načinu izračuna ocene posamezne entitete, ki je na koncu metrika rangiranja entitet.

Tudi na področju raziskav in analiz športnih tekmovanj je pogosto prisotna omrežna analiza v povezavi z naključnimi sprehodi po omrežju in eno izmed metod rangiranja entitet. Na tem področju je bila v preteklih desetletjih največja težava pridobivanje potrebnih podatkov za samo gradnjo omrežij. V zadnjih letih se je to spremenilo, saj se je dostopnost podatkov znatno povečala. Podatki večjih športnih turnirjev so prosto dostopni na spletu in hitro so se pojavila zanimiva dela, ki za osnovo uporabljajo tovrstne podatke. Ker je količina podatkov precej velika, tudi če se omejimo na

nivo enega turnirja, je omrežna analiza zelo priljubljena izbira za predstavitev in analizo tovrstnih podatkov.

Kljub nekoliko različnim ciljem so si marsikatera dela na področju omrežne analize športnih podatkov, kar se tiče uporabljenih metod in algoritmov, precej podobna. Razlog je predvsem v zelo podobni strukturi zgrajenih omrežij ne glede na to, na katerega izmed ekipnih športov se omejimo. Prav tako ni zelo pomembno ali smo osredotočeni na posamezno tekmovanje ali posamezno sezono.

Zato večina del, ki se ukvarja z rangiranjem entitet v specifičnem omrežju, uporablja nekakšno različico algoritma PageRank [7]. Ker je bil PageRank v osnovi zasnovan za rangiranje spletnih strani, v mnogih primerih ni najprimernejši za analizo športnih omrežij, vseeno pa predstavlja dobro osnovo. Po izgradnji omrežij navadno najprej sledi uvodna analiza, s katero pridobimo vpogled v osnovne lastnosti omrežij. S pridobljenim znanjem o lastnostih analiziranih omrežij lahko prilagodimo algoritem PageRank tako, da bolje ustreza potrebam analize.

Z modeliranjem naključnega sprehoda skozi graf športnega tekmovanja in rangiranjem z uporabo algoritma PageRank se ukvarjajo v [13]. Za namene pridobitve boljšega rangiranja športnih ekip predlagajo prilagoditve algoritma, ki vplivajo na verjetnosti posameznih prehodov v naključnem sprehodu. Zaradi specifik športnih tekmovanj je pogosto potrebno za pravilne rezultate analize algoritem PageRank prilagoditi. Le-ta je bil namreč zasnovan z namenom rangiranja spletnih strani, omrežje spleta pa je po strukturalnih lastnostih zelo drugačno od omrežja, ki ga navadno zgradimo za namene analize športnih tekmovanj. Prvo ima predvsem precej večje število vozlišč in povezav. Zaradi tega je verjetnost, da ima določeno vozlišče le vhodne povezave, izjemno majhna. V primeru analize določenega ekipnega športnega tekmovanja po metodologiji, ki je podobna naši, pa je verjetnost takega pojava precej večja. Ob pogledu na sam postopek izgradnje omrežja ugotovimo, da se to lahko zgodi v primeru, ko določena ekipa v posameznem športnem tekmovanju ni bila poražena. Tak pojav ni izjemno pogost, je pa povsem

možen, saj se tudi v najboljših in najbolj konkurenčnih športnih tekmovanjih dogaja, da je določena ekipa tako dobra, da skozi celoten potek tekmovanja ne izkusi poraza. To je posebej problematično v primeru analiz, ki temeljijo na operacijah nad matrikami, ki predstavljajo strukturo povezav v omrežju. S preprosto modifikacijo matrike je mogoče ta problem odpraviti [14], vendar s tem vnesemo določeno mero pristranskosti do neporažene ekipe, saj jo za potrebe analize predstavimo kot nekoliko slabšo.

Zaradi narave algoritma PageRank, ki v model naključnega sprehoda preko posebnega parametra vnese določeno mero naključnosti, ki jo predstavljajo občasni naključni skoki med vozlišči omrežja, ta pojav ni usoden in ga ni potrebno posebej obravnavati.

Pri analizi športnih tekmovanj skozi čas v analizo vnesemo novo dimenzijo. Algoritem PageRank je definiran kot statično stanje naključnega sprehoda, kar pomeni, da je predpogoj za uporabo tega algoritma statično omrežje, na katerem izvajamo analize. V primerih analiz več časovnih obdobj temu ni tako. Najbolj osnovna rešitev omenjenega problema je gradnja omrežij, ki predstavljajo določeno športno tekmovanje v različnih časovnih obdobjih. Tako namesto enega omrežja na analizirano športno tekmovanje za vsako analizirano športno tekmovanje dobimo eno omrežje za vsako sezono tekmovanja. S takim pristopom število omrežij, ki jih je potrebno analizirati, skokovito narašča s številom opazovanih sezon. V namene analize dinamičnih omrežij, kjer se povezave spreminjajo v časovni dimenziji, bi bilo potrebno povezave predstaviti kot nekoliko drugačne entitete z atributom časovne značke. Tako bi lahko vedeli, v katerem časovnem obdobju obstaja določena povezava. To informacijo je nato mogoče uporabiti v algoritmu PageRank, ki ga lahko prilagodimo tako, da se zaveda časovne dimenzije in minljivosti posameznih povezav v omrežju [15]. Tako prilagojen algoritem PageRank je primeren za analizo dinamike posameznega omrežja, kjer se pomembnost oziroma moč posameznih vozlišč v omrežju s časom spreminja. V našem primeru takšna analiza ni ključnega pomena, saj nas zanima analiza na nivoju posameznih sezon. Tako lahko na koncu ugotovljamo korelacijo med pridobljenimi po-

datki o izenačenosti in tržni zanimivosti tekmovanj. Vsekakor pa bi bil zanimiv vidik analize različnih športnih tekmovanj tudi analiza moči posameznih vozlišč, ki predstavljajo posamezne klube, skozi daljša časovna obdobja.

Med športi po priljubljenosti prevladuje nogomet, ki se na vrh lestvice uvršča tako po gledanosti kot tudi po številu aktivnih igralcev [16]. Prav zaradi največje priljubljenosti tega športa veliko število del na področju analize športnih tekmovanj temelji na analizi nogometnih turnirjev in tekmovanj. Zaradi zanimivosti taktičnega aspekta športnih tekmovanj in konec koncev tudi njegove pomembnosti, se veliko del ukvarja s tovrstnimi analizami.

Cilj vsake športne ekipe so čim boljši rezultati. Zato ekipe naredijo vse, kar je v njihovi moči, da bi bile čim bolj konkurenčne. Trenerji so zadolženi za to, da na igrišče pošljejo najboljše igralce, ki jih ekipa premore. Izbira igralcev je torej zelo pomembna, veliko se govori o sodelovanju in t.i. kemiji ekipe. S porastom uporabe omrežne analize se je le-ta pojavila tudi v tovrstnih raziskavah. Na podlagi različnih statističnih podatkov posameznih tekem, je mogoče zgraditi omrežje igralcev posamezne ekipe in prikazati različne aspekte njihovega sodelovanja tekom tekme. Na podlagi izsledkov analiz je mogoče pokazati, kateri igralci dobro igrajo skupaj in prinašajo boljše rezultate ekipi kot celoti [17]. Prav tako je mogoče metode omrežne analize uporabiti v namene raziskovanja taktične postavitve nasprotne ekipe in ugotavljanja dinamike njihove igre [18].

Začeni s svetovnim nogometnim prvenstvom leta 2006 sta dve največji krovni nogometni organizaciji na svetu, evropska UEFA in svetovna FIFA, začeli prosto dostopno objavljati večjo količino statističnih podatkov o posameznih tekmah. To je omogočilo bolj podrobne analize ne samo na nivoju posameznega turnirja temveč tudi na nivoju posamezne tekme. Ogromna količina dostopnih podatkov o posameznih tekmah večjih turnirjev omogoča podrobno analizo posameznih tekem in taktik posameznih ekip.

Omrežja zgrajena v namene analize posamezne tekme so precej drugačna od tistih, ki se uporabljajo za analize tekmovanj. Tukaj vozlišča predstavljajo posamezne igralce, povezave pa določeno razmerje med njimi. Za analizo tak-

tik in ugotavljanje načina igre posamezne ekipe je zelo pomembna strategija prenašanja žoge. Z analizo omrežja, ki temelji na igralcih kot vozliščih in podajah med njimi kot povezavah, je mogoče pridobiti pomembne informacije o načinu igre posamezne ekipe. Takšne informacije so lahko zelo koristne trenerjem ekip, ki morajo svojo ekipo kar najbolje pripraviti na soočenja z drugimi ekipami [19]. Za analizo omrežja podaj se lahko poslužujemo različnih metod, kot so t.i. dostopnost (angl. *closeness centrality*) [20], in vmesnost (angl. *betweenness centrality*) [21]. Prva nam pove, kako preprosto je priti do posameznega vozlišča v omrežju, kar se na omrežju podaj med igralci prevede v pomembnost igralca pri podajanju. Na podlagi takih podatkov se lahko igralci nasprotne ekipe med branjenjem osredotočajo na nasprotne igralce z višjo oceno *closeness centrality* in tako ohromijo pretok žoge med igralci nasprotne ekipe. Druga metoda se lahko uporabi za iskanje ozkega grla v podajalnem omrežju. Tako lahko identificiramo igralce, ki so na primer ključni za prenašanje žoge iz obrambe v napad.

Kljub drugačni zgradbi omrežja in ciljem avtorji v [6] za ugotavljanje najpomembnejših igralcev v podajalnem omrežju posamezne ekipe uporabljajo tudi algoritem PageRank, ki je podobno kot v našem delu prilagojen za analizo usmerjenih uteženih omrežij. Zanimiv je tudi pogled na analizo položaja na igrišču, kjer igrajo posamezni najpomembnejši igralci podajalnega omrežja v različnih klubih. To razkriva taktično postavitve in taktične zamisli ekipe [22].

Yamamoto in Yokoyama sta šla v analizi omrežij podaj še nekoliko dlje in sta z analizo uspela potrditi korelacijo med določenim vzorcem podaj in uspešnimi napadi nogometne ekipe [23]. Ker gre pri tovrstni analizi za podatke s časovno dimenzijo, sta avtorja iskala trikotnike povezanih vozlišč, ki predstavljajo igralce. Ugotovila sta, da obstaja korelacija med številom trikotnikov in številom napadov nogometne ekipe.

Do podobnih zaključkov so prišli tudi v delu [24], kjer so analizo izvajali na podatkih svetovnega nogometnega prvenstva leta 2010 v Južni Afriki. V delu so analizirali tekme zmagovalca prvenstva, t.j. reprezentanca Španije.

Omrežja so bila sestavljena na podoben način kot v ostalih sorodnih delih, ki se ukvarjajo z analizo posamezne tekme. Temelj analize je opazovanje koeficienta gručenja (angl. *clustering*) skozi različna časovna obdobja tekme. Prav španska reprezentanca je v načinu igre in načinu kroženja žoge nekoliko nenavadna, saj igra t.i. tiki taka slog igre. Za ta slog je značilno večje število podaj na manjšem prostoru, kar se odraža v višjem koeficientu gručenja. Z analizo posameznih tekem španske reprezentance so ugotovili, da povečan koeficient gručenja navadno privede do doseženega zadetka, kar potrjuje pogosto nogometno prepričanje, da večja posest žoge in dobro sodelovanje igralcev zvezne vrste prineseta zmago. Ugotavljanje korelacije med doseženimi goli in različnimi metrikami igre je zanimivo predvsem z vidika izboljšanja igre posamezne ekipe. Če lahko pokažemo korelacijo med povečano metriko, ki meri določen element igre, in doseženimi goli, se lahko pri razvoju ekipe osredotočimo na to metriko in posledično upamo na boljše rezultate. Zato se s tem vprašanjem ukvarja več del. Manjšo korelacijo je bilo mogoče pokazati med več metrikami omrežja, izstopala pa je metrika gostote omrežja, ki je imela poleg koeficienta gručenja najvišjo korelacijo z doseženimi goli na posamezni tekmi [25].

Kljub velikemu številu del, ki se osredotočajo na analizo nogometa, nogomet ni edini šport, kjer je tovrstna metodologija primerna. Podoben pristop se lahko uporabi za analizo večine ekipnih športov. Veliko del se ukvarja z analizo ameriških lig, kjer so podatki praviloma lažje dostopni. PageRank kot osnova za rangiranje se je izkazal za uspešnega tako pri rangiranju igralcev v ameriški bejzbol ligi MLB [26] kot tudi pri rangiranju ekip univerzitetne lige ameriškega nogometa [27, 28].

Z vidika športnih trenerjev, vodstva ekip in tudi igralcev samih, je zelo zanimiva analiza posameznih športnih klubov, igralcev in njihove perspektivnosti. Analize perspektivnosti posameznih igralcev se lahko lotimo na različne načine. Zelo pomemben vidik pri razvoju mladih igralcev so njihovi soigralci, saj z njimi na dnevni ravni opravljajo treninge. Gradnja omrežja v take namene zato poteka nekoliko drugače, saj je potrebno zajeti podatke

o sodelovanju med igralci. Tudi struktura omrežja je nekoliko drugačna, še vedno pa lahko za analizo uporabimo podobno metodologijo, ki temelji na algoritmu PageRank [29]. Izjemno pomembni pri razvoju igralcev so seveda tudi klubi, kjer igralci igrajo. Pogosto se namreč zgodi, da perspektivni mladinci po prestopu v največje klube padejo v formi, saj tam ne dobijo rednih priložnosti za igranje na tekmah. Marsikdaj je z vidika razvoja mladega igralca bolj smiselna izbira nekoliko slabšega kluba, kjer mu je omogočeno redno igranje. Za uspešno identificiranje klubov, kjer se mladinci hitro razvijejo, potrebujemo podatke o prestopih med klubi. Tako lahko zgradimo omrežje prestopov, pri tem pa ob določanju uteži upoštevamo pomembnost kluba, kamor je bil igralec prodan. Nato lahko z uporabo t.i. metode *betweenness centrality* ugotovimo, kateri klubi so nekakšne odskočne deske za igralce mlajših generacij [29]. Omrežje prestopov lahko uporabimo tudi za analize klubov z drugačnimi cilji. S pomočjo metode PageRank in *betweenness centrality* lahko identificiramo klube, ki so najpomembnejši pri trgovanju z igralci oziroma rangiramo klube glede na sposobnost posredovanja pri prestopih med klubi [30]. Tako lahko identificiramo lige, kjer se igralci razvijajo, in lige, ki so denarno močnejše in kupujejo najboljše igralce. V ligah prvega tipa klubi veliko zaslužijo od prestopov perspektivnih oziroma že razvitih igralcev, v ligah drugega tipa pa klubi veliko denarja namenijo za okrepitve in kupujejo najboljše in najbolj vredne igralce.

Zelo zanimiv problem, ki ga je mogoče reševati z uporabo omrežne analize, je napovedovanje rezultata ali uspešnosti določene ekipe v posamezni sezoni. V obeh primerih gre za neko vrsto rangiranja športne ekipe znotraj športnega tekmovanja. Ta podatek je zelo pomemben stavnicam, saj jim to omogoča pravilno definiranje kvot za dogodke na posameznih tekmah. Zato že vrsto let razvijajo napredne tehnologije za avtomatsko določanje kvot, ki temeljijo na rangiranju nasprotnikov. Tudi krovne športne organizacije, ki periodično objavljajo jakostne lestvice, se poslužujejo podobnih metod. Lestvice in kvote so tako dober približek moči določene ekipe.

Ti podatki niso na voljo za vse tekme, predvsem pa jih ni enostavno pri-

dobiti za daljša zgodovinska obdobja. Zato se pojavlja potreba po razvoju metode, ki je sposobna s podobno natančnostjo predvideti zmagovalca tekme, ali rangirati ekipe posameznega tekmovanja v opazovani sezoni glede na njihovo moč. Pri tem je pomembno, da metoda kot vhod potrebuje le preproste podatke, kot so na primer podatki o rezultatih odigranih tekem. S pomočjo takih podatkov lahko zgradimo dokaj preprosto omrežje ekip, kjer povezave predstavljajo posamezne medsebojne tekme. Z uporabo različice algoritma PageRank je mogoče pridobiti tako napovedi o zmagovalcu posamezne tekme kot tudi napovedi o razpletu celotne sezone določenega športnega tekmovanja [31]. Mi smo šli še nivo višje, saj smo razvili podobno metodo za ugotavljanje moči posamezne ekipe na podlagi osnovnih podatkov o odigranih tekmah. Podatke o moči posameznih ekip, pridobljene z omenjeno metodo, pa smo uporabili v namene ocenjevanja izenačenosti oziroma uravnoteženosti posameznega športnega tekmovanja.

Opazimo lahko, da večina del, ki se ukvarja z rangiranjem, za osnovo uporablja algoritem PageRank ali nekakšno prilagojeno različico tega priljubljenega algoritma. Seveda za rangiranje obstajajo tudi druge metode. Večina del, ki uporablja algoritem PageRank, analize izvaja na omrežjih, ki predstavljajo končno stanje sezone ali drugega opazovanega časovnega obdobja. Pri tovrstnih analizah je PageRank zelo primeren. Drugače pa je, če želimo analizo izvesti na neki poljubni točki v času. Ob neki poljubni točki v času so opazovane ekipe odigrale le delež vseh tekem v sezoni. To z vidika analize ni problematično, lahko pa je analiza zelo pristranska, saj je pomemben sam razpored tekem. Neka ekipa ima lahko odličen rezultatski izkupiček, vendar je v začetku sezone večino tekem odigrala proti slabšim ekipam, na koncu pa jo čakajo težje tekme in obratno. Metoda, ki tega ne vzame v zakup, bi tako bila naklonjena ekipam z lažjim razporedom v opazovanem časovnem obdobju. Zato je za tovrstne analize bolj primerna Masseyjeva metoda [32], ki rang ekipe c v poljubnem obdobju t definira glede na razliko med točkami, ki so jih nasprotnne ekipe osvojile proti ekipi c , in točkami, ki jih je ekipa c osvojila proti nasprotnim ekipam, ter rang ekip, s katerimi se je ekipa c pomerila

v časovnem obdobju t . V kombinaciji z omrežno analizo lahko z uporabo Masseyjeve metode, ki ji kot vhod podamo podatke pridobljene iz lastnosti omrežja, učinkovito analiziramo in rangiramo vozlišča v omrežju [33]. Tako lahko v poljubni točki v sezoni zgradimo omrežje ekip, kjer povezave predstavljajo določene tekme, in s pomočjo Masseyjeve metode izračunamo moč oziroma rang posamezne ekipe. Vemo, da se forma in posledično moč ekip s časom spreminja. Nemalokrat se zgodi, da določena ekipa v nekem obdobju sezone igra veliko bolje ali veliko slabše kot v drugem delu sezone. V osnovni Masseyjevi metodi se ta razlika v formi tekom opazovanega obdobja t pri izračunih ne upošteva. Tako se lahko zgodi, da sta ekipi a in b obe premagali ekipo c v različnih obdobjih. Ekipa a se je z ekipo c pomerila v obdobju, ko je bila ekipa c superiorna, ekipa b pa se je z ekipo c pomerila v obdobju, ko je bila ekipa c v slabši formi. Masseyjeva metoda tega zmagama daje enako težo, kar pogosto ni najbolj primerno. Z določenimi spremembami v postopku lahko za vsako tekmo v času t_1 izračunamo moči obeh nasproti stoječih ekip in rezultatu priredimo bolj primerno težo [34].

Ker v pričujočem delu med drugim analiziramo izenačenost športnih tekmovanj, ki temelji na analizi moči posameznih ekip v tem športnem tekmovanju, je zanimiva analiza ranga posameznih športnih ekip skozi čas in njihove raznolikosti. Avtorji so v [35] analizirali več tekmovanj različnih športov. Ugotavljajo, da je raznolikost ranga v vseh analiziranih športnih zelo podobna. Poleg tega so prišli do zaključka, da se na mestih pri vrhu lestvice v istem časovnem obdobju navadno zamenja manjše število ekip kot na dnu lestvice. Tak pojav lahko negativno vpliva na izenačenost posameznega športnega tekmovanja, ki jo analiziramo v pričujočem delu. Tovrstne analize bi bilo zanimivo opraviti še na podatkih katere izmed najpopularnejših severnoameriških športnih lig (t.j. NBA, NHL, NFL), kjer so prisotna v poglavju 1 omenjena pravila, ki pomagajo slabšim ekipam tekmovanja z namenom povečanja izenačenosti lige.

Omrežna analiza kot pristop za raziskave na področju športa ni omejena le na dosedaj omenjene športe. Zelo podoben pristop grajenja omrežja ekip in

analiza z uporabo algoritma PageRank sta se izkazala za uspešno tudi pri analiziranju večje količine podatkov tekem kriketa in iskanju najboljše reprezentance zadnjih 40 let [36].

Učinkovitost omrežne analize in algoritma PageRank prav tako ni omejena le na ekipne športe. Pri analizi posameznih športov so že podatki, ki predstavljajo osnovo analize, povsem drugačni. Pri marsikaterem posameznem športu na tekmi ne tekmujeta le dva tekmovalca. V teh primerih je izgradnja omrežja drugačna, drugačne pa so tudi njegove lastnosti. Obstajajo tudi taki športi, kjer si nasproti stojita dva tekmovalca. Tukaj lahko izgradnja omrežja poteka po podobnih načelih kot pri ekipnih športih. Prav tako je analiza lahko podobna in temelji na priljubljenem algoritmu PageRank. Šport, pri katerem lahko uporabimo podobno metodologijo analize je tenis. S pomočjo algoritma PageRank so avtorji v [37] razvili metodo za rangiranje igralcev tenisa. V [38] avtorji predlagajo metodo ocenjevanja uspešnosti posameznih igralcev tenisa, ki prav tako temelji na omrežni analizi. Aplicirali so jo na podatke tekem od leta 1968 do leta 2010 in razkrili najboljšega igralca glede na predlagano metodo.

Obstojećih del na področju analize omrežij, ki temeljijo na športnih podatkih, je torej veliko. Nekatera se osredotočajo na bolj ozko usmerjene analize tekem z namenom preučevanja taktik športnih ekip in ugotavljanja ključnih igralcev. Druga se ukvarjajo s problemom rangiranja ekip. Večini je skupna osnova, saj za analizo uporabljajo nekakšno različico algoritma PageRank. Naše delo se od večine razlikuje že v sami raziskovalni domeni, saj smo analize izvajali na večji količini podatkov, ki ne izvirajo samo iz tekmovalcev iz različnih držav in kontinentov temveč tudi različnih športov. Skupni razpon opazovanih sezon se razteza čez daljše časovno obdobje dolžine skoraj 40 let. Poleg razlik v vhodnih podatkih ima pričujoče delo tudi širše zastavljene cilje, saj je rangiranje ekip znotraj posameznega tekmovanja le predpogoj za uspešno ugotavljanje izenačenosti posameznega športnega tekmovanja. Ta podatek pa predstavlja osnovo za končno ugotavljanje korelacije med izenačenostjo športnega tekmovanja in njegovo tržno zanimivostjo.

Poglavje 3

Metodologija

3.1 Pridobivanje in obdelava podatkov

3.1.1 Zajem podatkov

Poleg podatkov o odigranih tekmah v različnih športnih tekmovanjih smo za potrebe ocenjevanja tržne zanimivosti in kasnejšega ugotavljanja korelacije potrebovali podatke o različnih pokazateljih tržne zanimivosti. Te podatke je bilo potrebno pridobiti za čim več evropskih nogometnih lig in ameriško košarkarsko ligo NBA za čim širši razpon sezon. Vse pridobljene podatke smo po zajemu obdelali, kot je opisano v nadaljevanju poglavja, in kasneje uporabili za ugotavljanje korelacije med izenačenostjo športnega tekmovanja in tržno zanimivostjo.

3.1.1.1 Zajem podatkov obiskanosti tekem

Eden najbolj očitnih pokazateljev tržne zanimivosti tekmovanja je gledanost. Bolj kot je športno tekmovanje zanimivo, več ljudi si ga želi ogledati. Zato se tržna zanimivost lige odraža tudi v številu prodanih vstopnic za posamezne tekme. S spleta smo pridobili podatke o obisku tekem posameznih ekip v različnih tekmovanjih v različnih sezonah. S spletne strani <http://www.european-football-statistics.co.uk/> smo pridobili

podatke o obisku tekem v osmih najmočnejših evropskih nogometnih ligah. Podatke o obisku tekem posameznih ekip v ligi NBA smo pridobili s spletnega mesta <http://www.apbr.org/>. Podatke gledanosti smo uspeli pridobiti za vse opazovane sezone, torej od leta 2000 do leta 2015.

3.1.1.2 Zajem podatkov o vrednosti nogometnih igralcev

Znano je, da so najbolj poznane športne lige na svetu zelo gledane, kar jim prinese veliko denarja. Del tega denarja se razdeli med klube, ki si posledično lahko privoščijo boljše igralce. Zato lahko podatek o ocenjeni tržni vrednosti igralcev, ki igrajo v posamezni ligi, odraža tržno zanimivost te lige. Podatke o ocenjeni tržni vrednosti posameznih igralcev v najmočnejših evropskih nogometnih ligah v vsaki posamezni sezoni smo pridobili s spletnega mesta <https://www.transfermarkt.de/>. Ker se je tržno vrednost igralcev bolj podrobno, javno dostopno in pogosto ocenjevalo šele od leta 2005 dalje, so podatki dostopni od tega leta dalje.

3.1.1.3 Zajem stavnih podatkov

Stave so z leti postale pomemben del vsakega športa. Zelo veliko ljudi, ki jih zanimajo športna tekmovanja, tudi bolj ali manj redno stavi. Ti podatki so za našo analizo in ugotavljanje korelacije med zanimivostjo in izenačenostjo tekmovanj zelo primerni. Predvsem nas zanimata podatka volumna stav na določeno tekmo (količina denarja vplačanega na stave izida te tekme), ki lahko odraža tržno zanimivost, in porazdelitve kvot za določen izid tekme, ki lahko odraža izenačenost.

Pridobili smo podatke o stavah največje borze stav na svetu *Betfair*. V podatkih je za vsak dogodek, na katerega je mogoče staviti, in vsako enolično kvoto za ta dogodek, ki se skozi čas spreminja, poleg drugih meta-podatkov zapisan tudi nam pomemben volumen stav.

Ker je te podatke težje pridobiti, saj jih stavnice ne ponujajo kot prosto dostopne, smo uspeli pridobiti podatke le od leta 2005 do leta 2013. V zajetih podatkih so, poleg drugih, vključene tudi lige, ki smo jih podrobneje analizi-

rali. To so najmočnejše evropske nogometne lige in najmočnejša košarkarska liga NBA.

3.1.2 Obdelava podatkov

3.1.2.1 Obiskanost tekem

Tržna zanimivost lige se odraža tudi v številu prodanih vstopnic za posamezne tekme, saj si bolj zanimive lige želi ogledati več športnih navdušencev. Pridobljene podatke o obisku stadionov posameznih ekip v različnih sezonah v različnih ligah smo obdelali tako, da smo izračunali povprečni obisk tekem lige v posamezni sezoni. Indeks gledanosti $A_{l,s}$ v ligi l v sezoni s smo izračunali po formuli

$$A_{l,s} = \frac{\sum_{i=1}^n a_{i,s}}{n} , \quad (3.1)$$

kjer $a_{i,s}$ predstavlja povprečno obiskanost domačih tekem ekipe i v sezoni s , n pa število ekip v ligi.

3.1.2.2 Vrednost nogometnih igralcev

Glede na hipotezo, da si tržno zanimivejši klubi lahko privoščijo boljše in posledično dražje igralce, je bilo potrebno pridobljene podatke o vrednosti posameznih igralcev primerno združiti. Da lahko pridobimo indeks vrednosti igralcev za posamezno ligo, je potrebno najprej izračunati vrednosti posameznih ekip v ligi, nato pa izračunati povprečno vrednost ekipe. Vrednost ekipe v posamezni sezoni smo izračunali po formuli

$$V_{i,s} = \sum_{j=1}^n v_{j,s} , \quad (3.2)$$

kjer $v_{j,s}$ predstavlja vrednost j -tega igralca v opazovani ekipi v sezoni s , n število igralcev v opazovani ekipi, i posamezno ekipo, s pa sezono.

Indeks vrednosti igralcev $\bar{V}_{l,s}$ smo izračunali po formuli

$$\bar{V}_{l,s} = \frac{\sum_{i=1}^m V_{i,s}}{m}, \quad (3.3)$$

kjer m predstavlja število ekip v ligi l v sezoni s .

3.1.2.3 Negotovost izida tekem glede na stavniške kvote

Stavnice določajo kvote glede na lastne podatke o jakosti posameznih ekip. Zato je kvota zelo dober približek negotovosti določene tekme. Če sta si kvoti za zmago ene in druge ekipe zelo blizu, pomeni, da je izid te tekme zelo negotov in obratno. Tako lahko z obdelavo podatkov o stavah pridobimo približke negotovosti na nivoju posameznih tekem v ligah. V pridobljenih podatkih o stavah vsaka vrstica predstavlja določen stavni dogodek, ki je vezan na dogodek na posamezni tekmi. V našem primeru smo se omejili na stavne dogodke, ki predstavljajo končni izid tekme, saj se tu najbolj odraža izenačenost tekme. V teh primerih je množica stavnih dogodkov E običajno omejena na dogodke E_d, E_g, E_n , kjer E_d predstavlja zmago domače ekipe, E_g zmago gostujoče ekipe, E_n pa neodločen izid. Stavnemu dogodku pripadata tudi kvota in volumen, ki predstavlja količino denarja vplačanega na ta dogodek s točno določeno kvoto. Ker se kvota s časom spreminja, je za ta stavni dogodek prisotnih več zapisov z različnimi kvotami. Za določanje negotovosti posamezne tekme moramo enake dogodke združiti. Združimo jih tako, da kvoto uteženo povprečimo glede na volumen. Združena kvota o_e za posamezni stavni dogodek e je torej

$$o_e = \frac{\sum_{i=1}^n o_{e,i} \cdot v_{e,i}}{\sum_{j=1}^n v_{e,j}}, \quad (3.4)$$

kjer $o_{e,i}$ predstavlja kvoto stavnega dogodka e v času i , $v_{e,i}$ pa volumen stav na stavni dogodek e v času i pri kvoti $o_{e,i}$.

Tako dobimo utežene kvote za posamezni dogodek. Te kvote pa se zaradi spreminjanja skozi čas, uteženega povprečja in politike stav, ki zaradi lastnega zaslužka določene dogodke opredeljujejo kot bolj verjetne, seštejejo v

vsoto, ki je večja od 1. Zato kvote najprej pretvorimo v psevdo verjetnosti po formuli

$$l_e = \frac{1}{o_e} \quad , \quad (3.5)$$

nato pa jih normaliziramo. Verjetnost izida posamezne tekme p_e za dogodek e smo izračunali po formuli

$$p_e = \frac{l_e}{\sum_{i \in E} l_i} \quad , \quad (3.6)$$

kjer E predstavlja množico opazovanih stavnih dogodkov.

3.1.2.4 Volumen stav na tekme v ligi

Ker so stave izjemno povezane s športnimi tekmovanji, je lahko količina denarja vplačanega na stave določene tekme ali na stave tekem določenega tekmovanja dober pokazatelj tržne zanimivosti tega tekmovanja. Ker ljudje običajno stavijo na tekme, ki si jih tudi ogledajo, večji volumen stav pomeni tudi večje zanimanje za to tekmo oziroma tekmovanje. V podatkih je za vsak dogodek, na katerega je mogoče staviti, in vsako unikatno kvoto za ta dogodek, ki se skozi čas spreminja, poleg drugih podatkov zapisan tudi volumen stav. Za potrebe našega dela smo se osredotočili na stave o izidu tekme oziroma zmagovalcu. Za vsako tekmo smo lahko izračunali skupni volumen stav tako, da smo sešteli volumne različnih kvot skozi čas za enak izid tega dogodka in tudi volumne različnih izidov dogodka. Skupni volumen smo izračunali po formuli

$$v_e = \sum_{r \in E} \sum_{o \in O} v_{e,r,o} \quad , \quad (3.7)$$

kjer e predstavlja opazovani stavni dogodek, r predstavlja enega izmed možnih izidov, o pa posamezno vrednost kvote v množici vseh kvot O , ki so se pojavile za določen izid tega dogodka.

3.2 Gradnja omrežij tekmovanj

Za analizo izenačenosti, ki je predpogoj za ugotavljanje korelacije med izenačenostjo in tržno zanimivostjo tekmovanja, smo uporabili omrežno analizo. Zato je bilo potrebno pred analizo zgraditi omrežja, ki opisujejo določeno tekmovanje v posamezni opazovani sezoni. Omrežje torej predstavlja posamezno športno tekmovanje v določenem časovnem obdobju. Lahko si ga predstavljamo kot prikaz poteka sezone v določenem športnem tekmovanju oziroma ligi.

Omrežje je sestavljeno iz vozlišč, ki predstavljajo ekipe in povezav, ki predstavljajo tekme med povezanima ekipama v opazovani sezoni. Vsaka ekipa je povezana z vsemi ekipami, s katerimi je odigrala vsaj eno tekmo. Število povezav med ekipama predstavlja število odigranih tekem v opazovani sezoni. Vse povezave so usmerjene in kažejo od poražene ekipe proti zmagovalni. V primeru izenačenega rezultata se ustvari povezava z utežjo enako nič.

3.2.1 Določanje uteži povezav

Da bolje orišemo premoč ene ekipe nad drugo, imajo povezave med ekipami različne uteži. Uteži predstavljajo rezultatsko razliko med ekipama na igrani tekmi. Tekom analize smo preizkusili več metod za določanje uteži. Različne metode določanja uteži so lahko bolj primerne za različne skupinske športe, saj so rezultati lahko pri različnih športih precej drugačnih velikostnih redov (npr. nogomet ali košarka). Prav zaradi velikih rezultatskih razlik med opazovanimi športi se je izkazalo, da je potrebno pri določanju uteži uporabiti logaritemsko ali sigmoidno funkcijo. Med vsemi preizkušenimi metodami določanja uteži nismo opazili velikih razlik v omrežjih in kasnejši analizi omrežij. Tekom razvoja metode za ugotavljanje izenačenost smo preizkusili uporabo logaritemske funkcije, ki temelji na točkovni razliki v rezultatu. Utež

posamezne povezave smo izračunali po formuli

$$w = \ln \left(1 + \frac{s_i - s_j}{\max(1, s_i)} \right) , \quad (3.8)$$

kjer s_i predstavlja rezultat zmagovalne ekipe, s_j pa rezultat poražene ekipe. V primeru enakega rezultata bo utež po zgornji formuli enaka 0. Za končne analize smo za določanje uteži uporabili logistično funkcijo in uteži izračunali po formulah

$$d = \alpha \cdot \frac{s_j - s_i}{s_i + s_j} , \quad (3.9)$$

$$w = \frac{1}{1 + e^{-d}} , \quad (3.10)$$

kjer s_i predstavlja rezultat zmagovalne ekipe, s_j rezultat poražene ekipe, α pa je parameter, ki se giblje med $\frac{1}{2}$ in 2 in je odvisen od športa. Pri športih z večjimi vrednostmi končnega rezultata je α manjši in obratno.

3.3 Analiza omrežij

Po uspešni gradnji omrežja smo na pridobljenih omrežjih izvedli vrsto analiz, s katerimi smo ugotavljali njihove lastnosti. Začeli smo z osnovnimi analizami omrežja, kasneje pa smo se osredotočili na bolj usmerjene analize. Glavni cilj analize omrežij je pridobiti mero izenačenosti posameznega tekmovanja v posamezni sezoni, ki bo v nadaljevanju služila kot osnova za izračun korelacije med izenačenostjo in tržno zanimivostjo posameznega tekmovanja.

3.3.1 Osnovna analiza omrežja

Namen osnovne analize omrežja je bila validacija pravilne izgradnje in pridobivanje osnovnih informacij, na podlagi katerih smo lahko izbrali primerne metode za nadaljnjo analizo. V preliminarni analizi omrežja smo s pomočjo

orodja za izgradnjo in analizo omrežij **NetworkX**¹ izračunali naslednje lastnosti omrežij:

- število vozlišč n ,
- število povezav m ,
- gostota omrežja $\rho = \frac{m}{n(n-1)}$,
- povprečna stopnja vozlišč $\bar{k} = \frac{2m}{n}$,
- povprečna moč vozlišč $\bar{s} = \frac{\sum_{i=1}^n \sum_{j=1}^m w_{i,j}}{n}$,
- premer (največja razdalja med vozlišči) omrežja D .

Da bi bolje razumeli porazdeljenost povezav v omrežju smo po formulah

$$\bar{k}^{in} = \frac{\sum_{i=1}^n k_i^{in}}{n} \quad , \quad (3.11)$$

$$\bar{k}^{out} = \frac{\sum_{i=1}^n k_i^{out}}{n} \quad , \quad (3.12)$$

kjer k_i^{in} predstavlja vhodno stopnjo, k_i^{out} izhodno stopnjo i -tega vozlišča v grafu, n pa število vozlišč v grafu, izračunali povprečno vhodno in izhodno stopnjo vozlišča v grafu. Po formulah

$$\sigma_{k^{in}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (k_i^{in} - \bar{k}^{in})^2} \quad , \quad (3.13)$$

$$\sigma_{k^{out}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (k_i^{out} - \bar{k}^{out})^2} \quad (3.14)$$

pa smo izračunali odklon vhodnih in izhodnih stopenj vozlišč v grafu.

¹<https://networkx.github.io/>

3.3.2 Analiza stopenj v omrežju

Stopnje v omrežju so zaradi načina gradnje omrežja zelo pomembne in imajo veliko težo pri ugotavljanju izenačenosti. Število vhodnih povezav nekega vozlišča, ki predstavlja določeno ekipo, namreč predstavlja število zmag, število izhodnih povezav pa število porazov te ekipe v opazovani sezoni. Analizirali smo torej tako vhodne kot tudi izhodne stopnje vozlišč v omrežju. S prikazom teh podatkov lahko dobimo boljšo predstavo o izenačenosti opazovane lige. Za lažje razumevanje izenačenosti smo analizirali tudi porazdelitve vhodnih in izhodnih stopenj vozlišč. V prvi fazi smo prikazali, koliko vozlišč v omrežju ima določene vhodne in izhodne stopnje. Za vsako možno vhodno ali izhodno stopnjo j lahko po formulah

$$n_j^{in} = \sum_{i=1}^n [k_{v_i}^{in} = j] \quad , \quad (3.15)$$

$$n_j^{out} = \sum_{i=1}^n [k_{v_i}^{out} = j] \quad (3.16)$$

dobimo število vozlišč n_j^{in} z vhodno stopnjo j in število vozlišč n_j^{out} z izhodno stopnjo j .

Za lažje ugotavljanje izenačenosti tekmovanja na podlagi vhodnih in izhodnih stopenj smo analizirali in prikazali še porazdelitve glede na funkcijo gostote verjetnosti (PDF) in zbirno funkcijo verjetnosti (CDF). Funkcija gostote verjetnosti je opredeljena kot

$$f(t) = \sum_{i=1}^n p_i \delta(t - x_i) \quad , \quad (3.17)$$

kjer so x_1, \dots, x_n diskretne vrednosti, ki jih lahko zaseda spremenljivka x (v našem primeru možne vrednosti vhodne ali izhodne stopnje vozlišča), p_1, \dots, p_n pa njim pripadajoče verjetnosti. Izrisali smo histogram funkcije gostote verjetnosti. Za lažjo analizo rezultatov smo določili intervale za histogram, ki zajemajo več vrednosti. Smiselna velikost intervala je odvisna od lastnosti omrežja, predvsem od števila vozlišč, saj je od tega neposredno od-

visna največja možna vrednost vhodnih ali izhodnih stopenj, če upoštevamo, da je število povezav med dvema vozliščema konstantno. Primerno velikost intervala s lahko predstavimo s formulo

$$s = \frac{3 \cdot k_{max}}{n} \quad , \quad (3.18)$$

kjer k_{max} predstavlja najvišjo možno vhodno ali izhodno stopnjo v omrežju, n pa število vozlišč v omrežju. Po izbiri velikosti intervala s lahko definiramo vse intervale in za vsakega izračunamo gostoto verjetnosti.

Zbirno funkcijo verjetnosti za posamezno vrednost spremenljivke x lahko predstavimo s formulo

$$F_x(x) = P(X \leq x), \quad (3.19)$$

kjer desna stran predstavlja verjetnost, da slučajna spremenljivka X (v našem primeru vrednost vhodnih ali izhodnih stopenj vozlišča) zasede vrednost, ki je manjša ali enaka x .

Ker delamo z uteženimi grafi in uteži predstavljajo rezultat oziroma premoč na določeni tekmi, velja iste analize opraviti še z vsoto uteži. Metodologija ostane enaka kot pri analizi samih uteži, le da namesto vhodnih in izhodnih stopenj analiziramo vsoto vhodnih in izhodnih uteži vozlišča j , ki sta definirani po formulah

$$W_j^{in} = \sum_{i \in \Gamma_j} w_{i,j} \quad , \quad (3.20)$$

$$W_j^{out} = \sum_{i \in \Gamma_j} w_{j,i} \quad , \quad (3.21)$$

kjer $w_{i,j}$ predstavlja utež povezave od vozlišča i do vozlišča j .

3.3.3 Analiza PageRank

Za potrebe določanja jakosti posameznih ekip znotraj lige smo uporabili različico priljubljenega algoritma za analizo omrežja PageRank. Podobno

kot v [29] smo osnovno različico tega algoritma priredili tako, da deluje na uteženih usmerjenih grafih. Algoritem deluje po psevdokodi, ki je zapisana spodaj.

Algorithm 1 Prirejeni algoritem PageRank

```

1:  $P \leftarrow$  vektor  $n$  skalarjev vrednosti  $\frac{1}{n}$ 
2: while  $\delta > \epsilon$  do
3:    $U \leftarrow$  vektor  $n$  skalarjev vrednosti 0
4:   for vozlišča  $i \in N$  do
5:     for predhodnik  $j \in$  sosedi  $i$  do
6:        $U_i \leftarrow U_i + P_j \cdot \frac{w_{j,i} \cdot \alpha}{k_j^{out}}$ 
7:     end for
8:   end for
9:    $u \leftarrow \|U\|$ 
10:  for vozlišča  $i \in N$  do
11:     $U_i \leftarrow U_i + \frac{1-u}{n}$ 
12:  end for
13:   $\delta \leftarrow \|P - U\|$ 
14:   $P \leftarrow U$ 
15: end while
16: return  $P$ 

```

3.3.3.1 Ocena PageRank

V prvi fazi PageRank analize smo za vsako vozlišče, ki predstavlja posamezno ekipo v opazovanem tekmovanju v opazovani sezoni, izračunali oceno PageRank, ki predstavlja jakost te ekipe v sezoni. Z analizo razporejenosti ocen PageRank lahko ocenimo izenačenost tekmovanja oziroma vidimo, če katera izmed ekip izstopa v jakosti v pozitivno ali negativno smer. Izračunali smo tudi porazdelitev ocen PageRank, ki nam pove kako je porazdeljena jakost ekip v tekmovanju. Ker je PageRank zvezna spremenljivka, je v tem primeru določitev intervalov za histogram funkcije gostote verjetnosti nujna. Za prikaze histogramov funkcije gostote verjetnosti ocen PageRank smo velikost

intervala določili po formulah

$$k = \frac{n}{3} \quad , \quad (3.22)$$

$$s = \frac{p_{max}}{k} \quad , \quad (3.23)$$

kjer k predstavlja število intervalov, n število vozlišč v omrežju, p_{max} pa najvišjo oceno PageRank, ki jo zavzame eno izmed vozlišč v omrežju.

3.3.3.2 Relativna entropija ocene PageRank

Ker same ocene PageRank posameznih ekip v določeni ligi ne povedo veliko o izenačenosti te lige kot celote, smo za ta namen razvili metodo določanja izenačenosti, ki temelji na relativni entropiji ocen PageRank posameznih ekip v ligi. Z algoritmom PageRank simuliramo naključnega gledalca. Gledalec se preko različnih tekem spreha od ekipe do ekipe. Z oceno PageRank dobimo predstavo koliko časa gledalec preživi ob gledanju tekem posamezne ekipe tekmovanja, relativna entropija ocen PageRank pa pokaže ali gledalec večino časa preživi ob gledanju tekem ene ekipe ali je njegov čas gledanja tekem bolj enakomerno porazdeljen med vse ekipe tekmovanja.

Entropijo PageRank izračunamo po formulah

$$p_i = \frac{pr_i}{\sum_{j=1}^n pr_j} \quad , \quad (3.24)$$

$$H = - \sum_{i=1}^n (p_i \cdot \ln p_i) \quad , \quad (3.25)$$

kjer je pr_i ocena PageRank posamezne (i -te) ekipe v ligi, n pa število ekip v ligi. Le-ta pa je izračunana po algoritmu predstavljenem v poglavju 3.3.3.

Relativna entropija PageRank, ki predstavlja izenačenost lige in je ključna za nadaljnje računanje korelacij, je izračunana po formuli

$$H_{rel} = \frac{H}{\ln(n)} \quad . \quad (3.26)$$

3.3.4 Ugotavljanje korelacije

Glavni del naloge predstavlja ugotavljanje korelacije med tržno zanimivostjo tekmovanja in njegovo izenačenostjo oziroma negotovostjo. Vsi dosedanja izračuni so služili kot osnova za ugotavljanje korelacije. Ker tržno zanimivost tekmovanja lahko predstavimo oziroma tretiramo na različne načine, smo korelacijo izračunali z več različnimi vhodnimi podatki o izenačenosti tekmovanja. Vsi rezultati so predstavljeni v poglavju 4. Vsem izračunom pa je skupna metodologija ugotavljanja korelacije, ki je opisana v nadaljevanju poglavja. Izračunali smo tako Pearsonov kot tudi Spearmanov koeficient korelacije. Zaradi same narave podatkov med katerimi računamo korelacijo, je bolj primeren Spearmanov koeficient korelacije, ki deluje na rangih in ne na samih vrednostih.

Pearsonov koeficient korelacije med dvema vrstama smo izračunali po formuli

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad , \quad (3.27)$$

kjer sta x in y vektorja, med katerima računamo korelacijo.

Spearmanov koeficient smo izračunali po isti formuli kot Pearsonov koeficient korelacije, le da pri Spearmanovem koeficientu za x in y uporabimo vektorja rangov, pridobljena z rangiranjem osnovnih vektorjev vrednosti, med katerima računamo korelacijo.

3.3.4.1 Metoda samovzorčenja

Pri izračunih korelacije moramo biti pozorni na interval zaupanja. V našem primeru, ko posamezna športna tekmovanja opazujemo na nivoju sezon in smo precej omejeni glede razpona podatkov, se pogosto zgodi, da je ocena korelacije zelo negotova. Zato moramo poleg izračuna korelacije vedno oceniti tudi interval zaupanja. Eden izmed možnih načinov je uporaba metode samovzorčenja (angl. *bootstrap*), ki nam iz omejenega nabora vzorcev pomaga pridobiti večje število vzorcev. Na podlagi tega pa lahko ocenimo interval

zaupanja.

Metoda samovzorčenja deluje na podlagi naključnega vzorčenja s ponavljanjem. S pomočjo naključnega vzorčenja lahko naš omejeni vzorec večkrat ponovno naključno vzorčimo in tako pridobimo večje število vzorcev. Pridobljene vzorce uporabimo za ponovne izračune in si zapomnimo vse rezultate. Iz množice vseh pridobljenih rezultatov, ki je velikosti števila ponovitev naključnega vzorčenja, lahko izračunamo poljubni interval zaupanja tako, da najdemo števili δ_{low} in δ_{high} , da velja

$$P(\delta_{low} \leq x \leq \delta_{high}) = \alpha \quad , \quad (3.28)$$

kjer x predstavlja naključno izbrani rezultat iz množice vseh pridobljenih rezultatov, α pa željeni interval zaupanja izražen kot decimalno število. Po tej formuli je torej α interval zaupanja enak $[\delta_{low}, \delta_{high}]$.

3.4 Masseyjeva metoda

Po pridobljenih rezultatih smo želeli predlagano metodo primerjati z Masseyjevo metodo [32], ki je prav tako namenjena rangiranju športnih ekip. Metoda temelji na oceni premoči določene ekipe, ki jo za tekmo k ter ekipi i in j lahko izračunamo po formuli

$$y_k = r_i - r_j \quad , \quad (3.29)$$

kjer r_i in r_j predstavljata oceni ekip i in j . Če analiziramo vse tekme v opazovani sezoni, kjer je n ekip odigralo m tekem, dobimo linearni sistem

$$Xr = y \quad , \quad (3.30)$$

kjer je X matrika velikosti $m \times n$. Vsaka vrstica predstavlja odigrano tekmo in ima na vseh mestih vrednost 0, razen na mestih i in j . Na mestu i je prisotna številka 1, če je zmagala ekipa i , oziroma -1 sicer. Na mestu j

je prisotna številka 1, če je zmagala ekipa j , oziroma -1 sicer. Tako lahko zgradimo matriko M po formuli

$$M = X^T X \quad (3.31)$$

in vektor p po formuli

$$p = X^T y \quad . \quad (3.32)$$

Matrika M ima na mestu i, j negativno število odigranih tekem med ekipama i in j , če i ni enako j , in število tekem, ki jih je odigrala ekipa i , če je i enako j . Vektor p predstavlja seštevek rezultatske razlike na dosedanjih tekmah za vsako ekipo v tekmovanju in ima dolžino enako številu ekip - n . V vektorju se na mestu i torej nahaja seštevek rezultatske razlike ekipe i na vseh odigranih tekmah.

Masseyjeva metoda je definirana kot sistem

$$Mr = p \quad . \quad (3.33)$$

Matriko M lahko predstavimo kot

$$M = D - A \quad , \quad (3.34)$$

kjer je D diagonalna matrika s številom tekem, ki jih je odigralo posamezno moštvo na diagonali, A pa je matrika, kjer $A_{i,j}$ predstavlja število tekem, ki jih je ekipa i odigrala proti ekipi j . Linearni sistem tako lahko zapišemo s formulo

$$Dr - Ar = p \quad . \quad (3.35)$$

Po formulah

$$r = D^{-1}(Ar + p) = D^{-1}Ar + D^{-1}p \quad , \quad (3.36)$$

$$r_i = \frac{1}{D_{i,i}} \sum_{j=1}^n A_{i,j} r_j + \frac{p_i}{D_{i,i}} \quad , \quad (3.37)$$

lahko izračunamo Masseyjevo oceno r za poljubno ekipo i .

3.4.1 Relativna entropija Masseyjeve ocene

Da smo lahko Masseyjevo metodo primerjali s predlagano metodo, smo izračunali relativno entropijo Masseyjeve ocene, ki predstavlja mero izenačenosti opazovanega športnega tekmovanja. Ker ocene Masseyjeve metode niso omejene na interval od 0 do 1, je bilo potrebno ocene prevesti na ta interval. Relativno entropijo Masseyjeve metode smo nato lahko izračunali po formulah

$$m_i = \frac{r_i}{\sum_{j=1}^n r_j} \quad , \quad (3.38)$$

$$H^m = - \sum_{i=1}^n (m_i \cdot \ln m_i) \quad , \quad (3.39)$$

kjer je r_i Masseyjeva ocena posamezne (i -te) ekipe v ligi, n pa število ekip v ligi. Relativna entropija Masseyjeve ocene je izračunana po formuli

$$H_{rel}^m = \frac{H}{\ln(n)} \quad . \quad (3.40)$$

Iz enačbe 3.37 je razvidno, da je Masseyjeva ocena za ekipo i sestavljena iz dveh delov. Prvi predstavlja jakostno oceno ekip, s katerimi se je ekipa i pomerila, drugi pa predstavlja rezultatski izkupiček ekipe i na teh tekmah. V našem primeru analize izvajamo na ravni sezone, kar pomeni, da so se vse ekipe že pomerile med seboj. Zato je prvi del Masseyjeve ocene enak za vse ekipe. Drugi del je neposredno odvisen od rezultatske razlike. To je ena izmed razlik glede na predlagano metodo, kjer za določanje uteži povezav, ki so odvisne od rezultatske razlike, uporabljamo sigmoidno funkcijo. Pri športih kot je nogomet, kjer so razlike v rezultatih manjše, ima to majhen vpliv. Vpliv je lahko večji pri športih kot je košarka, kjer so rezultati višji. V predlagani metodi predstavimo tudi parameter α , ki vpliva na določanje

uteži povezav in ga lahko prilagodimo potrebam opazovanega športa. Tako lahko predlagano metodo samo s prilagajanjem parametra uporabljamo za analizo različnih športov, ki imajo drugačne velikostne rede rezultatov.

Poglavje 4

Analiza in rezultati

Po uspešnem pridobivanju in obdelavi podatkov, določitvi poteka in metodologije gradnje omrežja ter analize smo se lotili same analize. Najprej smo opravili analizo izenačenosti športnih tekmovanj. Da smo bolje razumeli lastnosti posameznih omrežij in njihovo spreminjanje skozi čas, smo opravili analizo posameznih športnih tekmovanj v različnih časovnih obdobjih oziroma sezonah. Nadaljevali smo s primerjavo izenačenosti športnih tekmovanj na podlagi različnih analiz. Ko smo pridobili vse potrebne podatke o izenačenosti tekmovanj, smo opravili končno analizo korelacije med pridobljenimi podatki o izenačenosti in prej obdelanimi podatki o tržni zanimivosti tekmovanj.

V analizo so vključene lige petih različnih ekipnih športov, t.j. nogomet, košarka, hokej, odbojka in rokomet. Zaradi omejitev, ki jih predstavlja otežen dostop do podatkov, smo za večino analiz opazovali tekmovanja med leti 2000 in 2015. Pri vseh tekmovanjih, ki so razdeljena na redni del in končnico, smo zaradi lažje primerljivosti pri analizi upoštevali le tekme rednega dela. Za lažje branje so v tem poglavju predstavljeni le najpomembnejši in najbolj zanimivi rezultati ter grafi, vse ostale pa si bralec lahko ogleda v prilogi A in B.

4.1 Analiza izenačenosti športnih tekmovanj

Večji del naloge predstavlja analiza izenačenosti športnih tekmovanj, saj je le-ta osnova za ugotavljanje korelacije med izenačenostjo in tržno zanimivostjo športnih tekmovanj. Cilj analize je pridobiti oceno izenačenosti za vsako opazovano tekmovanje v vsaki posamezni sezoni. Le tako lahko ugotavljamo korelacijo med podatki pridobljenimi s to analizo, ki predstavljajo izenačenost tekmovanja in obdelanimi podatki o tržni zanimivosti tekmovanja v posameznih sezonah.

4.1.1 Analiza posameznega športnega tekmovanja

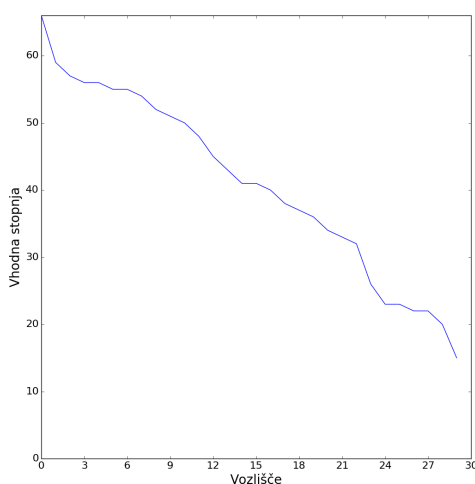
Za boljše razumevanje omrežij in pridobivanje podrobnejših informacij o spremembah omrežij posameznih športnih tekmovanj skozi čas, smo se najprej lotili analize omrežij posameznih športnih tekmovanj v posamezni sezoni. V nadaljevanju so predstavljeni rezultati analize pomembnejših sklopov opravljenih analiz. Za podrobnejšo analizo posameznega športnega tekmovanja smo izbrali ameriško košarkarsko ligo NBA, za katero smo uspeli pridobiti največ potrebnih podatkov. Z vidika analize izenačenosti je med bolj zanimivimi sezona 2007/2008, ko je po rednem delu na vrhu prevladovala ekipa Boston Celtics, na dnu pa je bila precej slabša od ostalih ekipa Miami Heat.

4.1.1.1 Stopnje vozlišč

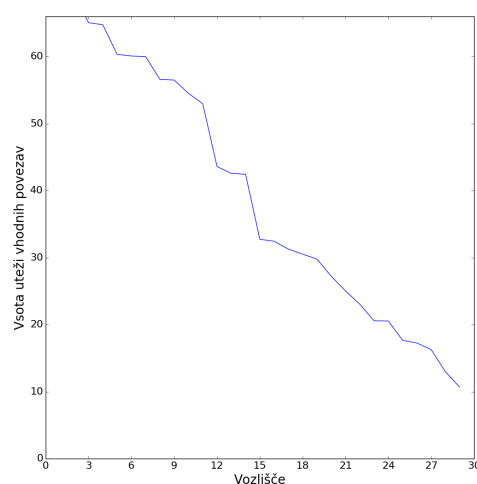
Zaradi strukture omrežja, ki temelji na izidih posameznih tekem opazovane sezone tekmovanja, so stopnje vozlišč pomemben pokazatelj moči ekipe, ki jo predstavlja vozlišče. Zmage določene ekipe se odražajo v vhodni stopnji vozlišča, ki predstavlja to ekipo, porazi pa se odražajo v izhodni stopnji tega vozlišča. Zato je smiselno analizirati stopnje, saj lahko tako dobimo občutek o izenačenosti tekmovanja, ki ga predstavlja omrežje, z uporabo dokaj enostavnih metod za analizo. Ker so povezave utežene in utež predstavlja rezultatsko razliko na tekmi, lahko analizo nadgradimo še z analizo vsote uteži vhodnih in izhodnih povezav in tako v analizo vpeljemo še koncept rezultatske premoči.

Izris vhodnih in izhodnih stopenj in vsote uteži vhodnih in izhodnih stopenj vozlišč

Najbolj osnovna analiza stopenj vozlišč je izris vhodnih in izhodnih povezav vozlišč, kjer za vsako vozlišče, od tistega z najvišjo vrednostjo do tistega z najnižjo, izrišemo vhodno in izhodno stopnjo. Strmina lomljenke nam predstavlja razliko v moči med močnejšimi in šibkejšimi ekipami.



Slika 4.1: Vhodne stopnje vozlišč v ligi NBA v sezoni 2007.



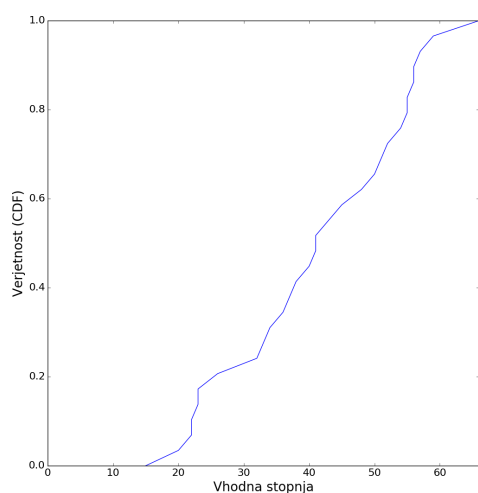
Slika 4.2: Vsota uteži vhodnih povezav vozlišč v ligi NBA v sezoni 2007.

Na sliki 4.1 lahko vidimo, da ima vozlišče 1 veliko višjo vhodno stopnjo od ostalih. To vozlišče predstavlja ekipo Boston Celtics, ki je v opazovani sezoni dominirala v rednem delu s 66 zmagami. Na drugi strani lahko vidimo, da ima ena ekipa občutno nižjo vhodno stopnjo. To je ekipa Miami Heat, ki je v tej sezoni nabrala le 15 zmag. Ob pogledu na graf kot celoto lahko opazimo tudi, da je bilo kar nekaj ekip tik pod vrhom precej izenačenih v številu zmag. Enako je z ekipami tik nad dnom. Na sliki 4.2 je predstavljena vsota uteži vhodnih povezav. Tukaj je poleg zmag pomembna tudi sama rezultatska razlika. Lomljenka na tem grafu je bolj odsekana na določenih delih, kar predstavlja večje skoke med ekipami. Na vrhu vodilni Boston Celtics niso več tako oddaljeni od zasledovalcev. Če pogledamo samo statistiko tiste

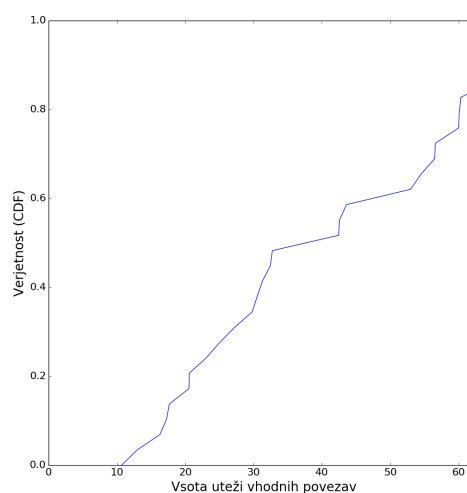
sezone lahko opazimo, da njihova razlika v koših na koncu sezone res ni bila veliko večja od drugih visoko uvrščenih ekip.

Izris porazdelitev vhodnih in izhodnih stopenj in vsote uteži vhodnih in izhodnih povezav vozlišč

Ker nam prikaz samih stopenj ali vsote uteži stopenj ne pove veliko in jo je težko interpretirati, je smiselni naslednji korak uporaba katere izmed verjetnostnih funkcij. Z uporabo formule (3.19) smo izračunali lomljenke zbirne funkcije verjetnosti vhodnih stopenj in vsote njihovih uteži.



Slika 4.3: Zbirna funkcija verjetnosti vhodnih povezav vozlišč v ligi NBA v sezoni 2007.



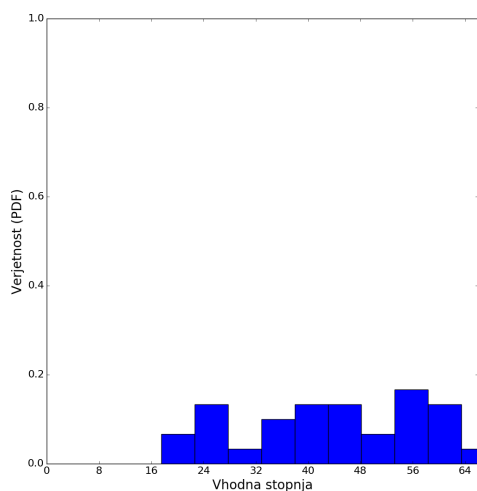
Slika 4.4: Zbirna funkcija verjetnosti vsote uteži vhodnih povezav vozlišč v ligi NBA v sezoni 2007.

Sliki 4.3 in 4.4 prikazujeta zbirno funkcijo verjetnosti vhodnih stopenj in vsote uteži vhodnih stopenj. Tukaj večji naklon ob določeni vrednosti abscisne osi pomeni več ekip s podobnim izkupičkom, manjši naklon pa ravno obratno. Na desni strani, kjer so večje vrednosti vhodnih stopenj oziroma večja vsota uteži le-teh, lahko vidimo nižji naklon lomljenke, kar je posledica premoči ene ekipe. Enako je na levi strani grafa, kjer so nižje vrednosti. Grafi zbirnih funkcij verjetnosti vhodnih stopenj oziroma vsote njihovih uteži nam lahko na ta način povedo veliko o izenačenosti tekmovanja. V tej sezoni je jasno

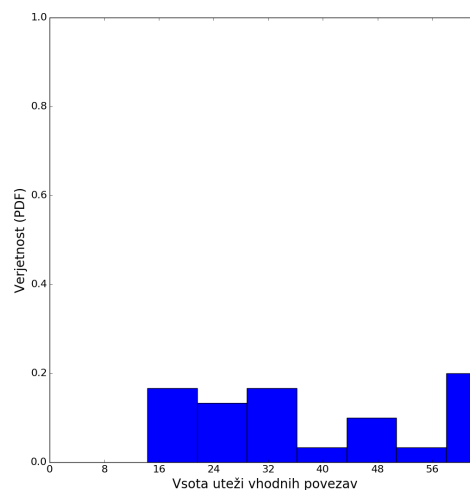
vidna premoč ekipe na vrhu in nemoč ekipe na dnu.

Izris funkcije gostote verjetnosti vhodnih in izhodnih stopenj in vsote vhodnih in izhodnih povezav vozlišč

Za uspešno analizo funkcije gostote verjetnosti vhodnih stopenj, izhodnih stopenj in vsote uteži vhodnih in izhodnih stopenj, je potrebno po formuli (3.18) izbrati primerno velikost intervala, nato pa lahko vrednost funkcije za vsak interval izračunamo z uporabo formule (3.17). Tako dobimo histogram funkcije gostote verjetnosti.



Slika 4.5: Histogram funkcije gostote verjetnosti vhodnih povezav vozlišč v ligi NBA v sezoni 2007.



Slika 4.6: Histogram funkcije gostote verjetnosti vsote uteži vhodnih povezav vozlišč v ligi NBA v sezoni 2007.

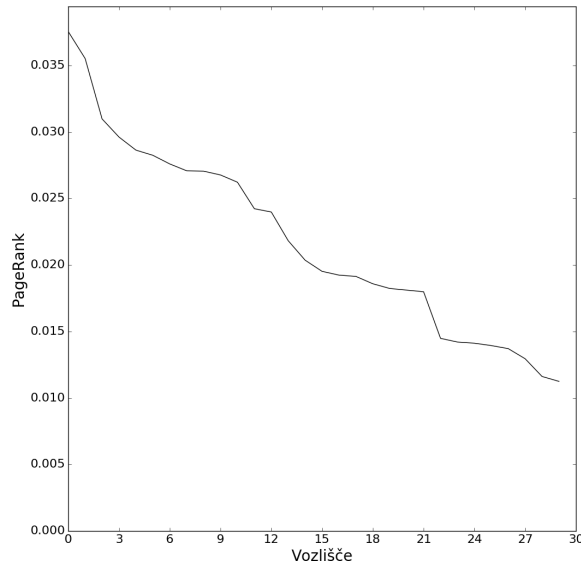
Z analizo grafov histogramov na slikah 4.5 in 4.6 lahko pridobimo dober vpogled v porazdelitev zmag med ekipami oziroma rezultatskih izkupičkov. Zaradi premoči ene ekipe je zadnji stolpec na sliki 4.5 zelo nizek, saj je v opazovani sezoni le ena izmed ekip dosegla več kot 60 zmag. Drugače je na sliki 4.6, kjer je zadnji stolpec višji zaradi manjših razlik v rezultatskem izkupičku tekem ekip pri vrhu. Več zaporednih stolpcev z višjimi vrednostmi na sliki 4.5 kaže na veliko izenačenost ekip v številu zmag tik pod vrhom in

v sredini lestvice.

4.1.1.2 Ocena PageRank

Izris ocene PageRank vozlišč

Z analizo vhodnih in izhodnih stopenj dobimo dober vpogled v zmage posameznih ekip in njihovo rezultatsko premoč. Za bolj napredno omrežno analizo, ki poleg zmag upošteva tudi moč nasprotnika, je potrebno poseči po bolj kompleksnih metodah. To lahko dosežemo z nekoliko prirejeno različico algoritma PageRank, ki ga prilagodimo tako, da deluje na uteženih usmerjenih grafih. Tako dobimo algoritem, ki je predstavljen v poglavju 3.3.3. Izris ocen PageRank posameznih ekip v opazovani sezoni nam prikaže njihovo uspešnost oziroma moč. Tako lahko podobno kot v poglavju 4.1.1.1 pridobimo vpogled v izenačenost tekmovanja s pomočjo analize razlik v ocenah PageRank, oziroma njihovih funkcij gostote verjetnosti in zbirnih funkcij verjetnosti.



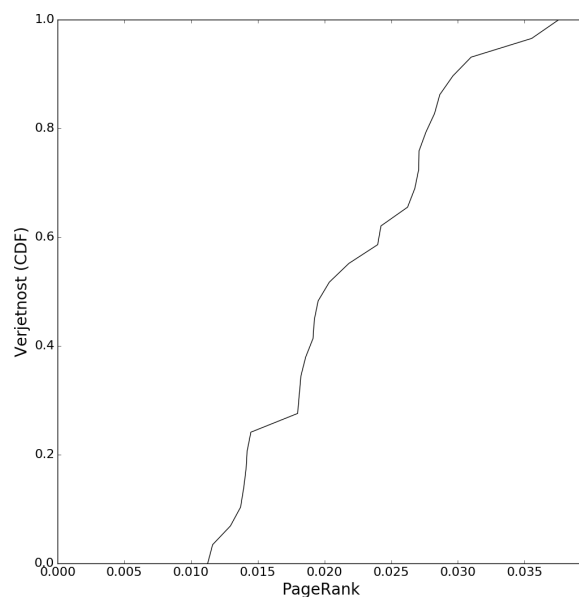
Slika 4.7: Ocena vozlišč PageRank v ligi NBA v sezoni 2007.

Slika 4.7 prikazuje ocene PageRank posameznih vozlišč, ki si na abscisni osi sledijo od najboljšega do najslabšega. Večja strmina lomljenke pomeni večje

razlike v moči ekip. Zaradi premoči na vrhu lestvice lahko vidimo večjo strmino na skrajni levi strani. Zaradi občutno slabših rezultatov nekaj ekip z dna lestvice je opaziti večji naklon lomljenke tudi na skrajni desni strani. V sredinskem delu, kjer je potekal boj za uvrstitev v končnico tekmovanja, je vidna večja izenačenost, kar razkriva bolj položna lomljenka. Ta pojav je pogosto viden pri ameriških tekmovanjih pri različnih ekipnih športih, saj se ekipe sredine lestvice krčevito borijo za uvrstitev v končnico. Ekipe z dna lestvice, ki za uvrstitev nimajo več možnosti, pa zaradi specifičnih pravil lige, ki nagrajujejo najslabše ekipe vsake sezone, nemalokrat tudi načrtno izgubljajo zadnje tekme. Zaradi tega pride do še večjih razlik med sredino lestvice in njenim repom.

Izris zbirne funkcije verjetnosti ocene PageRank vozlišč

Tako kot pri analizi vhodnih in izhodnih stopenj smo nadaljevali analizo z izrisom zbirne funkcije verjetnosti ocene PageRank. Ta nam pomaga oceniti razlike v samih ocenah PageRank med posameznimi vozlišči. Skladno s sliko 4.7 je tudi na sliki 4.8 videti premoč na vrhu, ki se kaže v položnejši lomljenki na desni strani grafa.

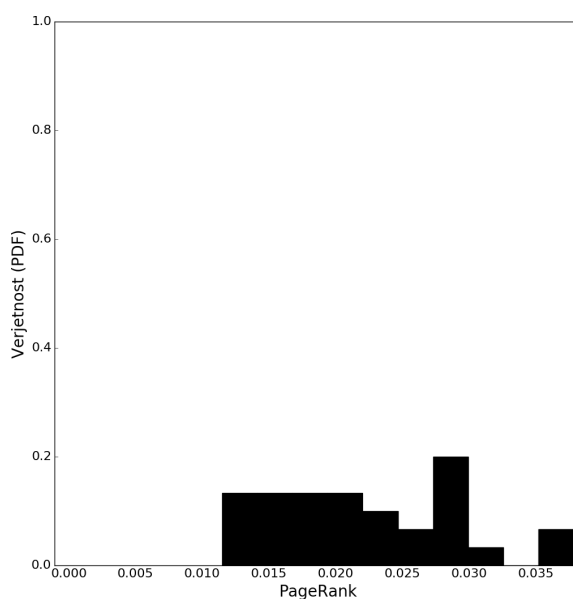


Slika 4.8: Zbirna funkcija verjetnosti ocene PageRank vozlišč v ligi NBA v sezoni 2007.

Izris funkcije gostote verjetnosti ocene PageRank vozlišč

Podobno kot pri analizi stopenj nam tudi pri analizi PageRank funkcija gostote verjetnosti pomaga lažje razpoznati porazdelitev ocen PageRank in njena nihanja med ekipami.

Na sliki 4.9, kjer so izrisane verjetnosti za posamezni interval ocene PageRank, je videti premoč na samem vrhu, saj so tam verjetnosti najnižje in velik skok okoli ocene PageRank, kjer se nahajajo ekipe, ki so se borile za uvrstitev v končnico.



Slika 4.9: Histogram funkcije gostote verjetnosti ocene PageRank vozlišč v ligi NBA v sezoni 2007.

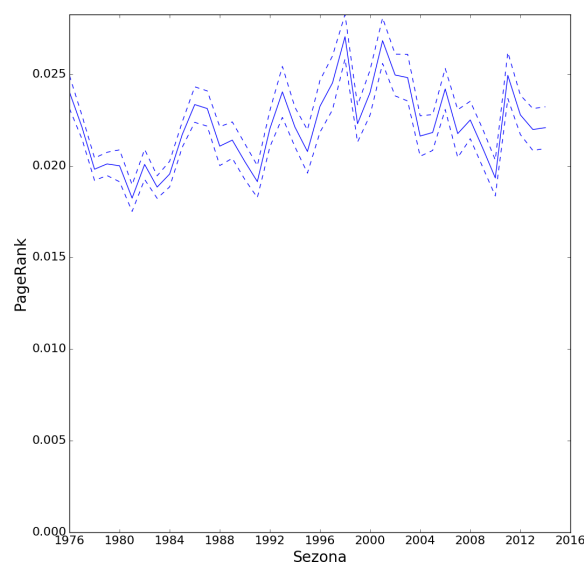
Izris povprečne ocene PageRank vozlišč v različnih sezonah

Ker ocene PageRank vozlišč dobro orišejo moč posameznih ekip v tekmovanju, je to lahko osnova za ugotavljanje izenačenosti tekmovanja. Če se premaknemo na nivo sezone, bi lahko rekli, da standardni odklon ocen vozlišč PageRank do neke mere predstavlja njihovo neuravnoteženost oziroma neizenačenost.

Iz slike 4.10 je razvidna povprečna ocena PageRank ekip v ameriški košarkarski ligi NBA v posamezni sezoni in njen standardni odklon. Vidimo lahko, da so nihanja skozi sezone prisotna, kar je pričakovano. Tudi standardni odklon se z leti spreminja. Težko pa bi na podlagi tega grafa ocenili trend izenačenosti.

4.1.2 Primerjava izenačenosti športnih tekmovanj

Po analizi posameznih športnih tekmovanj smo za lažjo primerjavo različnih športov in tudi primerjavo športov v različnih državah ali kontinentih sveta, analizo izvajali na več tekmovanjih hkrati. Omejili smo se na analize, ki



Slika 4.10: Povprečna ocena vozlišč PageRank v ligi NBA v različnih sezonah in ocena standardne napake.

so se izkazale kot najbolj obetavne med analiziranjem posameznih športnih tekmovanj. Zato smo se v zaključnih fazah dela osredotočili na PageRank analizo, ki smo jo nekoliko priredili tako, da lahko na podlagi ocen PageRank posameznih ekip znotraj tekmovanja pridobimo indeks izenačenosti opazovanega tekmovanja v posamezni sezoni. Videli smo, da sama povprečna ocena PageRank in njen standardni odklon nista dovolj. Zato smo za potrebe ugotavljanja izenačenosti posameznega tekmovanja v posamezni sezoni z uporabo formule (3.26) pridobili relativno entropijo PageRank tekmovanja v opazovani sezoni. Vse nadaljnje analize izenačenosti smo izpeljali z uporabo relativne entropije PageRank.

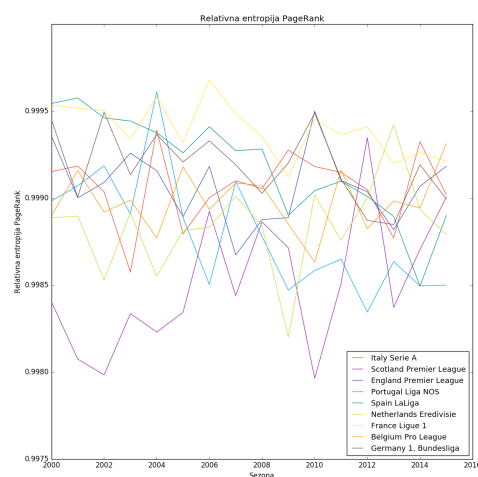
4.1.2.1 Relativna entropija PageRank

Po uspešni gradnji omrežja za vsako opazovano športno tekmovanje v vsaki opazovani sezoni smo s pomočjo algoritma 1, ki je predstavljen v poglavju 3.3.3, izračunali ocene PageRank posameznih vozlišč. Z uporabo formule (3.26) smo za vsako tekmovanje na podlagi ocen PageRank posameznih

ekip tega tekmovanja izračunali relativno entropijo PageRank tega tekmovanja, ki predstavlja izenačenost tekmovanja. Večja relativna entropija PageRank pomeni večjo izenačenost. Za lažjo predstavo o razlikah v izenačenosti med različnimi športnimi tekmovanji, smo relativno entropijo PageRank več športnih tekmovanjih v različnih sezonah izrisali na en graf.



Slika 4.11: Relativna entropija PageRank najmočnejših 5 evropskih nogometnih lig.



Slika 4.12: Relativna entropija PageRank najmočnejših 9 evropskih nogometnih lig.

Relativna entropija PageRank nogometnih lig na sliki 4.11 pokaže precejšnjo izenačenost znotraj petih najmočnejših evropskih lig. Pri vseh ligah je zaznati nekaj odstopanj v relativni entropiji PageRank od sezone do sezone, kar je pričakovano. Velik del opazovanega obdobja ima najvišjo relativno entropijo PageRank francoska prva nogometna liga, kar nakazuje na večjo izenačenost te lige. Pogled na končno razvrstitev te lige v letih med 2000 in 2015 pokaže, da je bilo v tem obdobju precej različnih zmagovalcev lige. Prav tako so bile ekipe zelo izenačene v srednjem delu lestvice, kar je kljub dominaciji ekipe Lyon v sredini tega časovnega obdobja pripomoglo k večji entropiji PageRank. V zadnji sezoni je bila ta nekoliko nižja, kar je v veliki meri verjetno posledica dominance ekipe Paris Saint Germain.

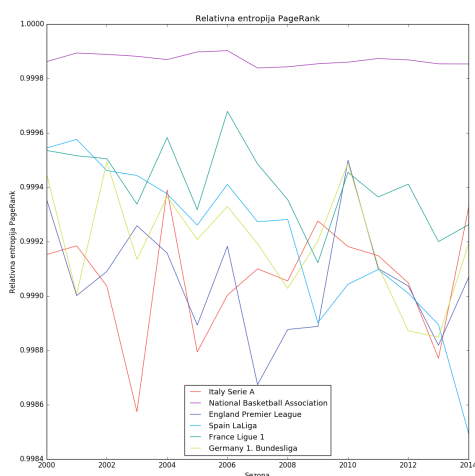
V tem pogledu je zanimiva španska prvi liga, kjer je po letu 2008 viden pa-

dec relativne entropije PageRank. To je posledica podpisa novih pogodb za televizijske pravice, ki je favoritoma Real Madridu in Barceloni prinesel še večji kos finančnih sredstev glede na ostale ekipe, ter finančne krize, ki je povzročila zlom ekip, ki so bile iz ozadja sposobne konkurirati nespornima favoritoma lige. Glede na razpored na samem vrhu lestvice je zanimivo, da je relativna entropija PageRank v letu 2011 (sezona 2011/2012) precej visoka, čeprav je bil Real Madrid v tisti sezoni dominanten. Vzrok za to je najverjetneje zelo zgoščena lestvica tik pod vrhom, kjer se v razmaku 10 točk zvrsti kar 9 ekip.

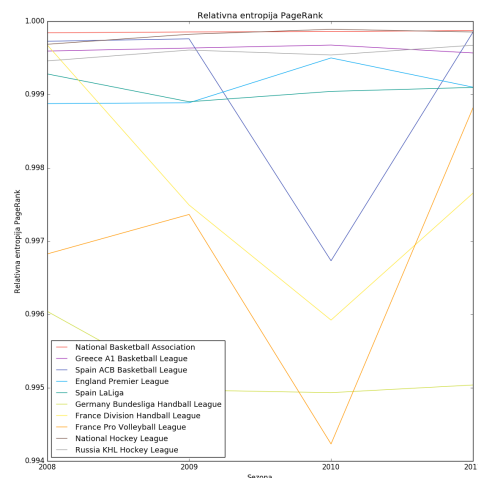
Pri italijanski ligi je viden padec relativne entropije PageRank v letu 2013 (sezona 2013/2014), kar je posledica močne dominance Juventusa v tej sezoni, ki je naslov prvaka osvojil s kar 17 točkami naskoka pred zasledovalci. Tudi leta 2003 (sezona 2003/2004) je bila relativna entropija PageRank precej nižja od povprečja, kar je najverjetneje posledica dominance AC Milana in nemoč zadnjeuvrščene Ancone, ki sta v jakosti zelo odstopala od glavnine.

Na sliki 4.12 so dodane še štiri nogometne evropske lige (skupaj 9). Vidimo, da po relativni entropiji PageRank v letu 2009 (sezona 2009/2010) navzdol izstopa nizozemska liga. V tej ligi v omenjeni sezoni težko govorimo o posebni dominanci katere izmed ekip, opazimo pa lahko, da so bile ekipe med seboj točkovno bolj oddaljene kot navadno. Manj je bilo torej ekip, ki bi bile točkovno zelo skupaj in bi se do zadnjih krogov neposredno borile za uvrstitev, kar je najverjetneje tudi vzrok padca relativne entropije PageRank. Prav tako je padec viden pri škotski ligi v letu 2010 (sezona 2010/2011). Ob pogledu na lestvico omenjene sezone lahko vidimo, da sta ekipi Celtic in Rangers dominirali skozi celotno sezono. Ti dve ekipi sta sicer bili izenačeni, dočim je bila razlika do sosednjih ekip izjemno velika.

Ko dodamo še prikaz relativne entropije PageRank najmočnejše košarkarske lige, t.j. ameriško ligo NBA, lahko vidimo, da je liga NBA vse sezone glede na relativno entropijo PageRank precej višje. To se sklada z začetno hipotezo, ki je predvidevala večjo izenačenost ameriških opazovanih tekmovanj (košarkarske lige NBA in hokejske lige NHL) zaradi številnih ukrepov vod-



Slika 4.13: Relativna entropija PageRank najmočnejših 5 evropskih nogometnih lig in košarkarske lige NBA.



Slika 4.14: Relativna entropija PageRank najmočnejših lig 5 različnih ekipnih športov.

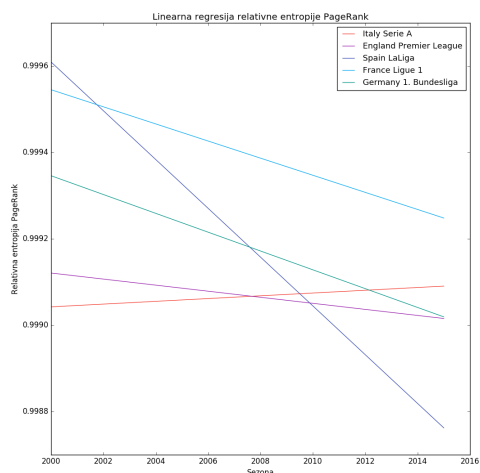
stva lige, ki pripomorejo k večji izenačenosti lige oziroma pomagajo slabšim ekipam v ligi. Prav tako je pri ligi NBA opaziti najmanjše nihanje v relativni entropiji PageRank v opazovanih sezonah.

Zanimiv je tudi pogled na relativno entropijo PageRank tekmovanj različnih ekipnih športov na sliki 4.14. Primerjali smo močnejše nogometne, košarkarske, hokejske, odbojkarske in rokometne lige. Izjemno visoka relativna entropija PageRank obeh ameriških lig (košarkarske NBA in hokejske NHL) tudi tukaj potrjuje začetno hipotezo, da bo zaradi ligaških pravil, ki spodbujajo izenačenost ekip, relativna entropija PageRank pri ameriških tekmovanjih višja.

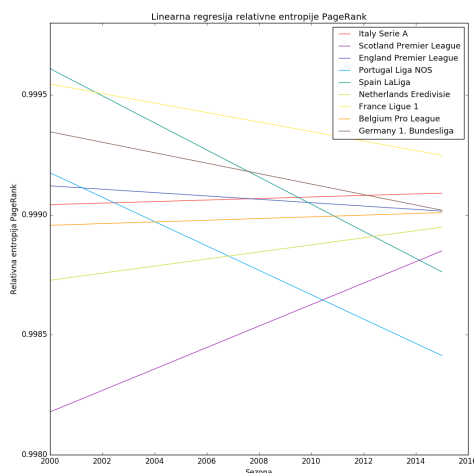
4.1.2.2 Linearna regresija relativne entropije PageRank

V začetni hipotezi smo predpostavili, da se bo predvsem pri ameriških športnih tekmovanjih z leti izenačenost povečevala, saj imajo posebna pravila, ki spodbujajo večjo izenačenost ekip. Za potrditev takšne hipoteze je primerjava relativne entropije PageRank manj primerna zaradi stalnih nihanj, ki se po-

javljajo iz sezone v sezono. Da bi bolje analizirali trend izenačenosti tekmovanja, ki ga enačimo s trendom relativne entropije PageRank tekmovanja, smo z uporabo linearne regresije izrisali trend relativne entropije PageRank.



Slika 4.15: Linearna regresija relativne entropije PageRank najmočnejših 5 evropskih nogometnih lig.

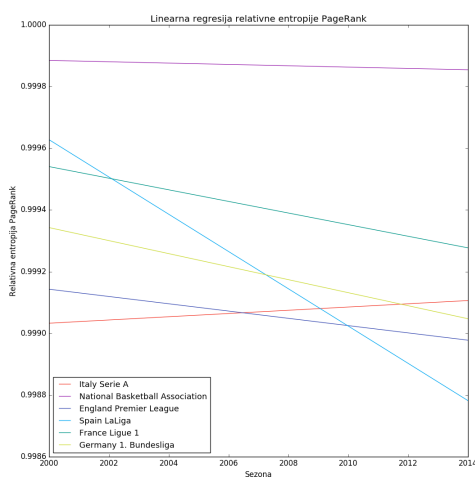


Slika 4.16: Linearna regresija relativne entropije PageRank najmočnejših 9 evropskih nogometnih lig.

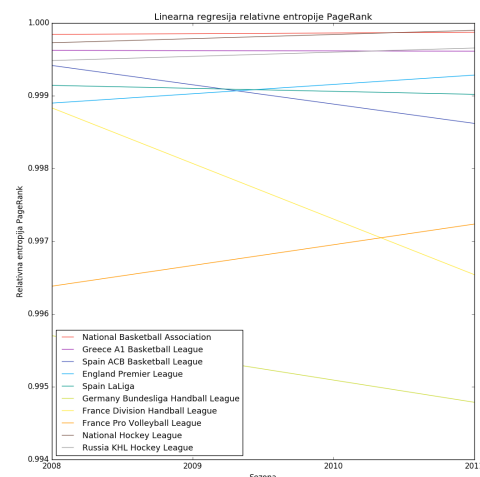
Prikaz linearne regresije relativne entropije PageRank na sliki 4.15 razkriva, da se je le-ta v najmočnejših petih evropskih nogometnih ligah, z izjemo italijanske v zadnjih 15 letih znižala. Največji padec je zaznati pri španski ligi predvsem zaradi zgoraj navedenih razlogov in posledične nesporne dominancje Real Madrida in Barcelone. Glede na prejšnje prikaze dejanskih vrednosti relativne entropije PageRank, je pričakovano najvišje francoska liga, najnižje pa italijanska liga. V splošnem bi lahko bila posledica padca relativne entropije PageRank v najmočnejših ligah znatno povečanje vložkov, ki jih ekipe namenjajo za nakupe in prestopne igralcev. Ekipe sredine lestvice z manj finančnimi sredstvi temu le težka sledijo, kar se na daljši rok odraža v njihovi nekonkurenčnosti.

Na sliki 4.16 je prikazan razširjeni nabor opazovanih evropskih nogometnih lig, kjer lahko opazimo, da izstopata predvsem portugalska in škotska liga.

Pri portugalski je opazen znaten padec, pri škotski pa znaten porast relativne entropije PageRank. Padec, ki je viden pri portugalski ligi, ne izvira iz posebne dominancne ene izmed ekip, saj iz lestvic zadnjih let lahko vidimo, da so se na prvih treh mestih večinoma menjavale tri različne ekipe. Padec linearne regresije je tako predvsem posledica večjih razlik med ekipami sredine lestvice. Pri škotski ligi je kljub dominanci Celtica relativna entropija PageRank visoka predvsem na račun zelo zgoščene lestvice, če izvzamemo najboljšo in najslabšo ekipo. Edini, ki je lahko konkuriral Celticu, je bila ekipa Rangers, ki pa je v sezoni 2011/2012 zaradi finančnih težav izpadla v drugo ligo. To je še bolj zaostribo boj za vsa mesta v ligi razen prvega, kjer kraljuje Celtic.



Slika 4.17: Linearna regresija relativne entropije PageRank najmočnejših 5 evropskih nogometnih lig in košarkarske lige NBA.



Slika 4.18: Linearna regresija relativne entropije PageRank najmočnejših 5 različnih ekipnih športov.

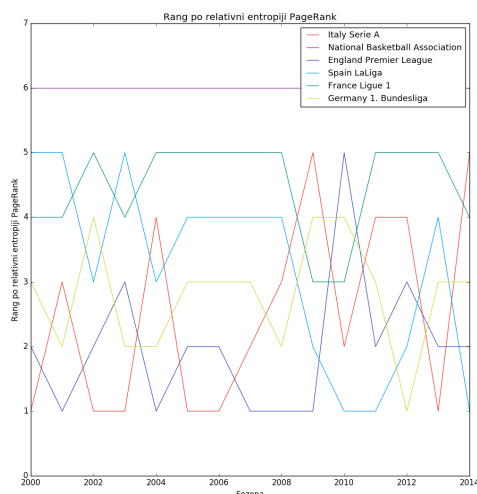
Pogled na linearno regresijo relativne entropije PageRank ob dodani ligi NBA na sliki 4.17 ni presenetljiv. Vidimo lahko, da je liga NBA najvišje in da je tu zaznati veliko nižji padec kot pri večini drugih nogometnih lig.

Na sliki 4.18 vidimo, da sta ameriški ligi NBA in NHL pričakovano visoko, prav tako tudi ruska hokejska liga KHL in grška košarkarska liga. Vse te

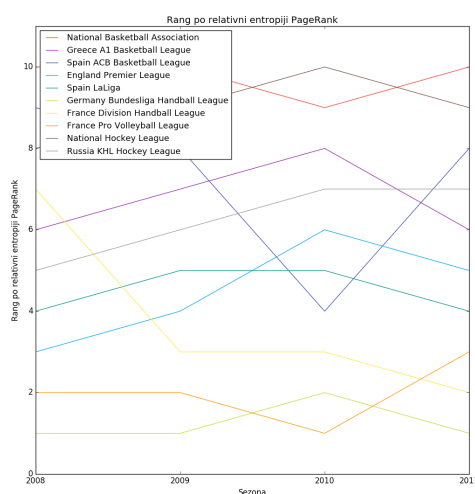
lige imajo tudi zelo rahlo padajočo ali celo naraščajočo linearno regresijo, kar nakazuje na precej konstantno oziroma nekoliko naraščajočo izenačenost lige.

4.1.2.3 Rang po relativni entropiji PageRank

Za prikaz razlik v izenačenosti med različnimi tekmovanji je v določenih primerih bolje uporabiti rangiranje tekmovanj glede na relativno entropijo PageRank namesto njihovih dejanskih vrednosti. Tako lahko enostavneje vidimo, kako so po izenačenosti razporejena različna tekmovanja, če se rangi ne spreminjajo veliko, ali pa vidimo, katera tekmovanja so si po izenačenosti morda bolj podobna od drugih.



Slika 4.19: Rang 5 najmočnejših evropskih nogometnih lig in košarkarske lige NBA po relativni entropiji PageRank.



Slika 4.20: Rang najmočnejših lig 5 različnih ekipnih športov po relativni entropiji PageRank.

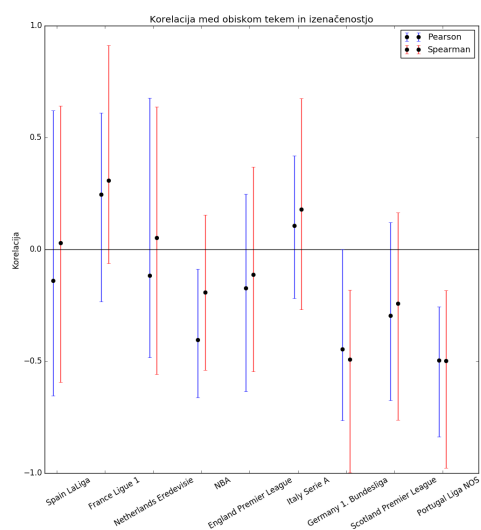
Pri prikazu ranga lig glede na relativno entropijo PageRank je pričakovano v ospredju ameriška košarkarska liga NBA. Iz lomljenke ranga te lige je lepo razvidno, da ima največjo izenačenost v vseh opazovanih sezonah, kar se odraža v najvišjem rangu po relativni entropiji PageRank. Tako kot že na grafu vrednosti, je tudi pri rangu viden padec španske lige po letu 2008.

Pri prikazu ranga vseh analiziranih športov na sliki 4.20 pričakovano zopet prevladujeta obe ameriški ligi.

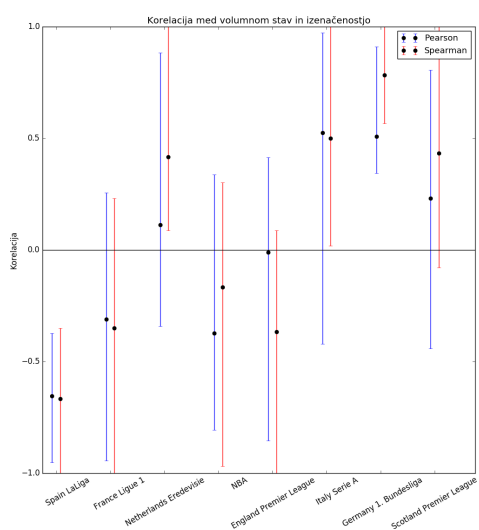
4.2 Analiza korelacije med zanimivostjo in izenačenostjo

Glavni namen in cilj magistrske naloge je bil primerjava izenačenosti športnih tekmovanj z njihovo tržno zanimivostjo. Korelacijo smo iskali med izenačenostjo, ki smo jo izračunali po metodi prilagojenega algoritma PageRank in relativne entropije ocen posameznih ekip znotraj tekmovanja ter obdelanimi podatki različnih metrik, ki naj bi predstavljale tržno zanimivost posameznega športnega tekmovanja. Za vsako metriko smo izračunali Pearsonov in Spearmanov koeficient korelacije med to metriko in izenačenostjo glede na relativno entropijo PageRank. Ker je pri izračunu korelacije izjemno pomemben tudi interval zaupanja, smo z vzorčenjem s pomočjo metode samovzorčenja korelacijo izračunali tisočkrat in izračunali 95% interval zaupanja.

Na slikah 4.21 in 4.22 lahko vidimo, da so 95% intervali zaupanja za korelacije na nivoju sezone izjemno široki. Zato je na podlagi teh rezultatov težko priti do trdnih zaključkov in potrditi ali zavreči katero izmed hipotez o korelaciji med izenačenostjo športnega tekmovanja in njegovo tržno zanimivostjo. Razlog za izjemno široke intervale zaupanja je predvsem v precej majhni količini podatkov, saj korelacijo za posamezna tekmovanja računamo na nivoju sezone, kar se glede na naš nabor opazovanih sezon prevede na 15 parov vrednosti. Vidimo tudi, da korelacija niha od tekmovanja do tekmovanja tako glede obiskanosti tekem kot tudi volumna stav. Korelacija izenačenosti in obiska tekem je pri veliki večini tekmovanj celo negativna. Vzrok za to bi lahko bila zvestoba navijačev ekip, ki predvsem v bolj znanih ligah zelo pogosto kupujejo sezonske vstopnice za tekme svojega priljubljenega kluba. Tako obiskujejo tekme ne glede na rezultatski uspeh ali neuspeh te ekipe. To je najbolj prisotno pri največjih klubih, ki imajo tudi največje kapacitete stadionov in posledično najbolj vplivajo na ligaško povprečje gledanosti. Tako je



Slika 4.21: Korelacija med obiskovanostjo tekem in relativno entropijo PageRank na nivoju sezone v najmočnejših nogometnih ligah in košarkarski ligi NBA.

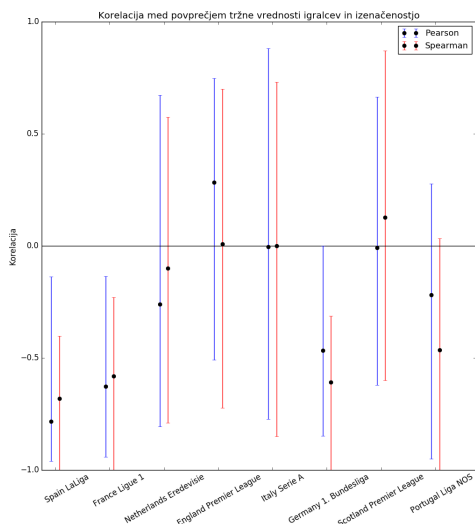


Slika 4.22: Korelacija med volumnom stav in relativno entropijo PageRank na nivoju sezone v najmočnejših nogometnih ligah in košarkarski ligi NBA.

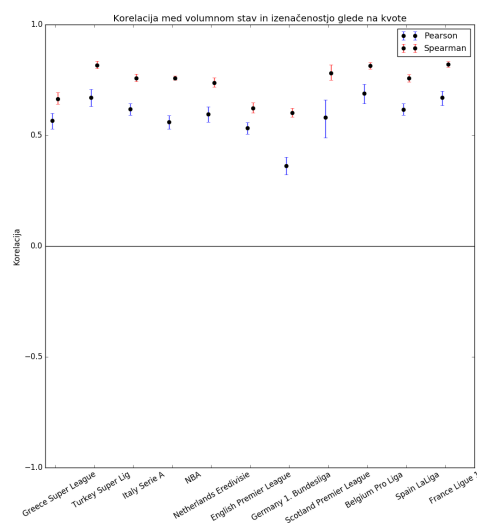
relativna zasedenost stadionov skozi opazovane sezone precej konstantna, absolutna gledanost pa se navadno viša, saj ekipe ogromno denarja investirajo v svoje stadione in dvorane [39] predvsem z namenom povečevanja njihovih kapacitet, kar se odraža v večjem številu prodanih vstopnic. V analizi iz prejšnjih poglavij smo videli, da izenačenost glede na relativno entropijo PageRank z leti rahlo pada v večini opazovanih lig. Oba omenjena pojava privedeta do negativne korelacije med gledanostjo in izenačenostjo v opazovanih tekmovanjih v opazovanih sezonah.

Nekoliko drugače je pri korelaciji med izenačenostjo glede na relativno entropijo PageRank in volumnov stav. Vrednosti korelacije so bolj pozitivne, intervali pa so zaradi istih razlogov še vedno zelo široki. Vzrok za višjo korelacijo je verjetno dejansko višja povezanost teh dveh metrik, saj je znano [40], da več ljudi stavi na tekme, ki so bolj izenačene, kar se na nivoju sezone odraža v večjem stavnem volumnu pri tekmovanjih, kjer so ekipe bolj

izenačene in posledično v posamezni sezoni postrežejo z več izenačenimi tekami.



Slika 4.23: Korelacija med povprečno tržno vrednostjo ekipe in relativno entropijo PageRank na nivoju sezone v najmočnejših nogometnih ligah in košarkarski ligo NBA.



Slika 4.24: Korelacija med izenačenostjo tekme glede na kvote in volumnom stav na nivoju tekme v najmočnejših nogometnih ligah in košarkarski ligo NBA.

Po izračunu in prikazu korelacije med povprečno tržno vrednostjo igralcev posamezne ekipe v ligi in relativno entropijo PageRank vidimo, da korelacija ni signifikantna. Zopet so zaradi prej omenjenih razlogov intervali zaupanja zelo široki, vrednosti korelacije pa nihajo med tekmovanji in so ponekod rahlo pozitivne, drugod pa rahlo negativne. Tako lahko zaključimo, da korelacije med izenačenostjo in povprečno tržno vrednostjo ekip v tekmovanju z uporabljenimi metodologijami ni bilo mogoče zaznati.

Ker je eden izmed problemov ugotavljanja korelacije na nivoju sezone premajhna količina podatkov, ki se odraža v izjemno širokem 95% intervalu zaupanja, smo korelacijo poskusili poiskati še z nekoliko drugačnim pristopom. Da bi pridobili večjo količino podatkov za računanje korelacije, smo se

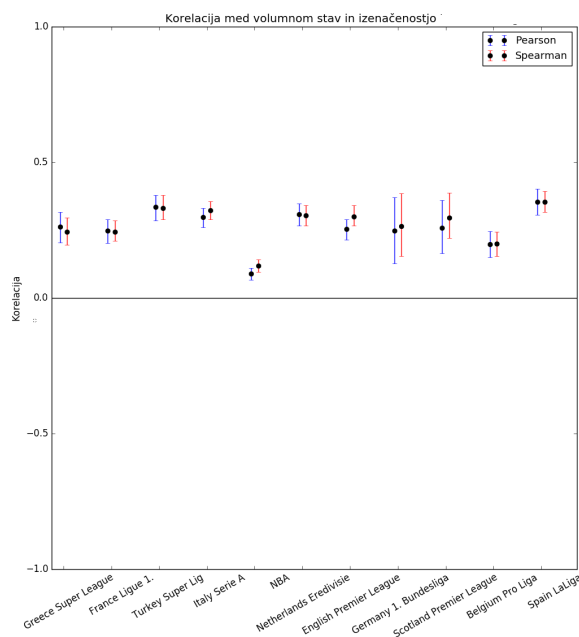
premaknili na nivo posamezne tekme. Najprej smo poizkusili z izračunom korelacije med izenačenostjo tekme, pridobljene z izračunom entropije verjetnosti iz stavnih podatkov po formuli (3.6) in volumnom stav na to tekmo. Rezultati na sliki 4.24 kažejo visoko korelacijo, kar potrjuje hipotezo, da na bolj izenačene tekme stavi več ljudi. Prav tako so zaradi analize na nivoju posamezne tekme in posledično večje količine podatkov 95% intervali zaupanja precej ožji. Zaključimo lahko, da je korelacija med neodločenostjo tekme glede na kvote stav in volumnom stav signifikantno pozitivna.

Ker je podatke o kvotah stav za daljša časovna obdobja težko pridobiti, smo poskusili izenačenost posamezne tekme izračunati še na podlagi v tem delu predlagane metode in ponoviti analizo korelacije. Izenačenost posamezne tekme smo dobili tako, da smo zgradili omrežje tekmovanja in sezone, kjer je bila odigrana posamezna opazovana tekma. Vozlišča v omrežju, ki predstavljajo ekipe, smo rangirali po oceni PageRank po algoritmu predstavljenem v poglavju 3.3.3. Končno oceno izenačenosti tekme med ekipama a in b smo izračunali po formuli

$$u = \frac{1}{|r_a - r_b|} \quad , \quad (4.1)$$

kjer, r_a predstavlja rang ekipe a glede na oceno PageRank, r_b pa rang ekipe b glede na oceno PageRank v sezoni, ko je bila odigrana opazovana tekma. Korelacijo smo izračunali po isti metodi kot v dosedanjih analizah.

Iz grafa korelacije med izračunano mero izenačenosti tekme in volumnov stav na tekmo je razvidno, da obstaja pozitivna korelacija med tema dvema metrikama. Korelacija je nekoliko manjša kot na sliki 4.24, kjer smo metriko izenačenosti osnovali na kvotah. Kvote so za določanje izenačenosti na nivoju posamezne tekme bolj natančne, saj v veliki meri upoštevajo trenutno formo nasproti si stoječih ekip, česar pa naša metoda ne upošteva, saj je bila zasnovana za analize na nivoju posamezne sezone. Tako je izenačenost izračunana po formuli (4.1) manj natančna kot tista pridobljena z izračunom entropije verjetnosti iz stavnih podatkov, ki jih izračunamo po formuli (3.6).



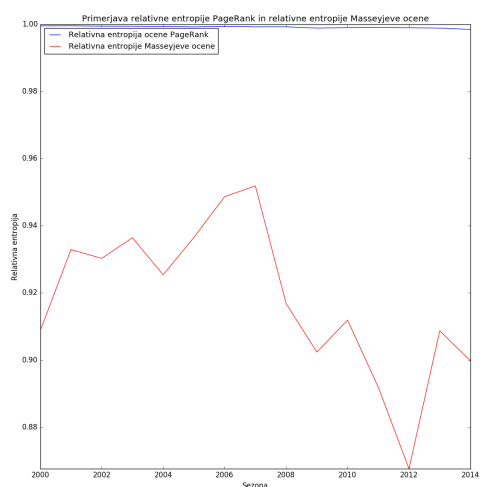
Slika 4.25: Korelacija med izenačenostjo tekme glede na razliko v rangu ekip po oceni PageRank in volumnom stav na nivoju tekme v najmočnejših nogometnih ligah in košarkarski ligi NBA.

4.3 Primerjava z Masseyjevo metodo

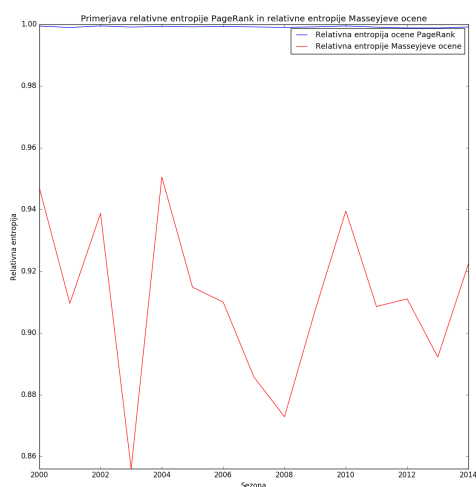
Predlagano metodo smo želeli primerjati z eno od že uveljavljenih metod za rangiranje ekip v športnih tekmovanjih. Izbrali smo Masseyjevo metodo in, tako kot pri predlagani metodi, izračunali relativno entropijo ocen ekip v opazovanem tekmovanju v opazovani sezoni. Relativna entropija predstavlja mero izenačenosti opazovanega tekmovanja. Primerjavo smo naredili na podatkih španske in nemške prve nogometne lige.

Na slikah 4.26 in 4.27 lahko vidimo, da so vrednosti relativne entropije PageRank veliko višje od vrednosti relativne entropije Masseyjeve ocene. Zato je zelo težko primerjati ti dve metriki izenačenosti, ko sta izrisani na istem grafu. Da bi lažje primerjali vrednosti metod, smo izračunali Pearsonov in Spearmanov koeficient korelacije med relativnimi entropijami za vsako ligo posebej in za združene relativne entropije obeh opazovanih lig.

Rezultati v tabeli 4.1 potrjujejo korelacijo med omenjenima metodama. To



Slika 4.26: Izenačenost skozi čas glede na relativno entropijo PageRank in relativno entropijo ocene po Masseyjevi metodi v španski prvi nogometni ligi.



Slika 4.27: Izenačenost skozi čas glede na relativno entropijo PageRank in relativno entropijo ocene po Masseyjevi metodi v nemški prvi nogometni ligi.

je pričakovano, saj metodi ocenjujeta isto metriko - izenačenost tekmovanja. Korelacija je podobna za obe opazovani tekmovanji, kar potrjuje pravilnost delovanja obeh metod.

Tekmovanje	Pearsonov koeficient	Spearmanov koeficient
Španska 1. nogometna liga	0.58	0.59
Nemška 1. nogometna liga	0.64	0.69
Skupaj	0.61	0.67

Tabela 4.1: Pearsonov in Spearmanov koeficient korelacije med relativno entropijo PageRank in relativno entropijo Masseyjeve ocene.

Poglavje 5

Sklepne ugotovitve

Omrežna analiza je v zadnjih letih postala zelo pogosta metoda reševanja problemov na različnih področjih. Tudi na področju športa se pojavlja vse več del, ki za osnovo uporabljajo omrežno analizo. V pričujočem delu smo s pomočjo podatkov rezultatov tekem različnih športov zgradili omrežja klubov v posameznih ligah, ki smo jih analizirali z metodami omrežne analize. Končni cilj analize je bilo ugotavljanje korelacije med izenačenostjo in tržno zanimivostjo opazovanih športnih tekmovanj.

Različna športna tekmovanja in lige imajo drugačna pravila in so drugače organizirana, vsem pa je skupen cilj uspešno poslovanje. Prihodke si zagotavljajo na različne načine prek različnih kanalov, osnova vseh prihodkov pa je gledanost. Zato se vsa tekmovanja trudijo privabiti čim več gledalcev in oboževalcev. Ker velja, da so bolj izenačena tekmovanja bolj zanimiva gledalcem, se vodstva lig poslužujejo različnih metod in prijemov, ki pripomorejo k večji izenačenosti ekip znotraj lige. To je predvsem prisotno v Severni Ameriki, kjer imajo lige vrsto pravil, katerih namen je favoriziranje in razvijanje slabših ekip. Že vrsto imajo pri naboru mladincev na začetku sezone slabše ekipe prednost pri izbiranju, prav tako vsako leto postavijo zgornjo mejo denarja, ki jih posamezna ekipa lahko nameni za plače svojih igralcev. Tako zagotavljajo večjo mero izenačenosti in si posledično obetajo bolj zanimivo tekmovanje in večje zanimanje gledalcev.

Da bi lahko potrdili ali ovrgli začetno hipotezo, da se večja izenačenost športnega tekmovanja odraža v večji tržni vrednosti, smo razvili metodo za določanje izenačenosti športnega tekmovanja. Vključuje gradnjo omrežij na podlagi omenjenih podatkov in analizo z uporabo prilagojenega algoritma PageRank. Po izračunu ocen PageRank ekip v opazovanem športnem tekmovanju izračunamo relativno entropijo PageRank, kar predstavlja izhod metode in je metrika za ocenjevanje izenačenosti športnega tekmovanja. Predlagana metoda je dovolj splošna, da jo je mogoče uporabiti za analize različnih ekipnih športov. Metoda kot vhod potrebuje le podatke o rezultatih odigranih tekem ekip v posameznem športnem tekmovanju, kar je tudi edini pogoj za uporabo metode. Zaradi preprostosti vhoda in dostopnosti omenjenih podatkov, lahko s predlagano metodo analizo izenačenosti tekmovanja izvajamo na zelo širokem naboru lig različnih ekipnih športov za daljša zgodovinska obdobja.

V delu smo analizo izvedli na različnih ligah petih ekipnih športov, t.j. nogometa, košarke, hokeja, odbojke in rokomet. Analiza izenačenosti je potrdila začetne hipoteze, da bo zaradi razlik v pravilih, ki pripomorejo k uravnoveženosti ekip, izenačenost severnoameriških lig večja. Tudi v daljšem opazovanem obdobju je vidna prevlada teh tekmovanj v pogledu izenačenosti po relativni entropiji PageRank. Po podrobnejšem pregledu rezultatov in analizi posameznih sezon določenih športnih tekmovanj, je mogoče najti vzporednice med anomalijami, ki so vidne v določenih sezonah in rezultati izenačenosti. To potrjuje pravilnost delovanja predlagane metode za ugotavljanje izenačenosti.

Predlagano metodo smo primerjali z Masseyjevo metodo, ki je prav tako namenjena ocenjevanju športnih ekip. Po izračunu Masseyjeve ocene za posamezne ekipe opazovanega tekmovanja smo izračunali relativno entropijo ocen in rezultate primerjali s tistimi, ki smo jih pridobili s predlagano metodo. Izračun Pearsonovega in Spearmanovega koeficienta potrjuje korelacijo med metodama. Korelacija med rezultati obeh metod je bila pričakovana, saj z njima ocenjujemo isto metriko. Rezultati metod niso enaki, kar je prav tako

pričakovano, saj metodi temeljita na različnih pristopih. Masseyjeva metoda je bolj primerna za rangiranje ekip kadar vse ekipe ne odigrajo medsebojnih tekem ali za rangiranje ekip tekom tekmovanja, ko še niso bile odigrane vse tekme.

Ugotavljanje izenačenosti športnih tekmovanj je le predpogoj za ugotavljanje korelacije med izenačenostjo in tržno zanimivostjo. Ker je tržna zanimivost zelo širok in nedefiniran pojem, smo v delu predstavili in predlagali nekaj različnih metrik, ki naj bi odražale tržno zanimivost športnega tekmovanja. Pridobili smo podatke o obiskanosti stadionov in dvoran v različnih ligah, povprečni tržni vrednosti igralcev ekip in volumnu stav na tekme posameznih ekip. Po preliminarni obdelavi podatkov metrik tržne zanimivosti in izračunu relativne entropije PageRank na nivoju sezone za opazovana športna tekmovanja smo izračunali Pearsonov in Spearmanov koeficient korelacije. Da bi pridobili intervale zaupanja izračunanih vrednosti koeficientov korelacije, smo uporabili metodo samovzorčenja. Ugotovili smo, da je podatkov, če se omejimo na posamezne sezone, za izračun korelacije premalo. Zato so intervali zaupanja preširoki, da bi rezultate lahko interpretirali. Da bi pridobili več podatkov za izračun korelacije, smo podoben postopek ponovili na nivoju posamezne tekme. Tako lahko drastično povečamo količino podatkov, nad katerimi v zaključni fazi računamo korelacijo. Z uporabo metode razlike v rangih smo pridobili mero izenačenosti posamezne tekme. Za metriko tržne zanimivosti na nivoju posamezne tekme smo vzeli volumen stav na posamezno opazovano tekmo in izračunali korelacijo. V tem primeru so intervali zaupanja veliko ožji. Rezultati koeficientov korelacije so potrdili začetne hipoteze, ki so predpostavljale pozitivno korelacijo med izenačenostjo in tržno zanimivostjo športnega tekmovanja.

Možno nadaljnje delo je nadaljevanje raziskovanja korelacije med izenačenostjo in tržno zanimivostjo tekmovanj na nivoju posamezne sezone. Za uspešno analizo bi potrebovali podatke o rezultatu tekem opazovanih tekmovanj v daljšem časovnem obdobju. Da bi lahko potrdili ali ovrgli hipotezo o pozitivni korelaciji izenačenosti in tržne zanimivosti športnih tekmovanj, bi bilo

najverjetneje potrebno pridobiti podatke o rezultatih tekem tekmovanj za zadnjih 40 ali 50 sezon. Ker smo na koncu korelacijo ugotavljali na nivoju posamezne tekme, metoda za ugotavljanje izenačenosti pa je bila sprva zasnovana za uporabo na nivoju sezone, obstajajo možnosti za izboljšave same metode. Predlagana metoda temelji na algoritmu PageRank in izračune opravlja na omrežju posameznega tekmovanja, ko so vanj vnešene vse povezave, ki predstavljajo vse tekme sezone. Zato je lahko na nivoju posamezne tekme pristranska. Določena tekma je bila lahko zelo izenačena zaradi trenutne forme, ko pa to tekmo ocenjujemo z vidika ranga ekip na koncu sezone, je lahko slika drugačna (in obratno). Zato bi bila verjetno primernejša različica Masseyjeve metoda, ki v zakup vzame tudi trenutno formo posamezne ekipe. Potrebno bi bilo tudi nekoliko spremeniti sam proces analize, saj bi bilo omrežja potrebno graditi v različnih časovnih obdobjih.

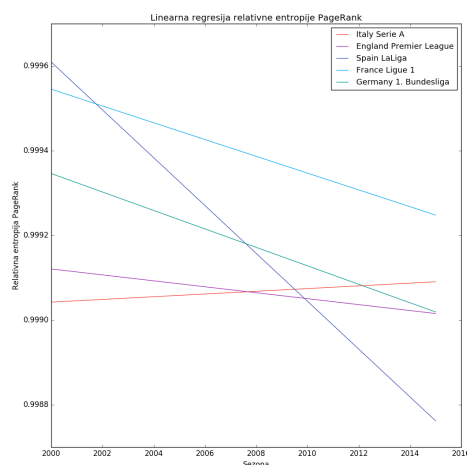
Dodatek A

Relativna entropija PageRank

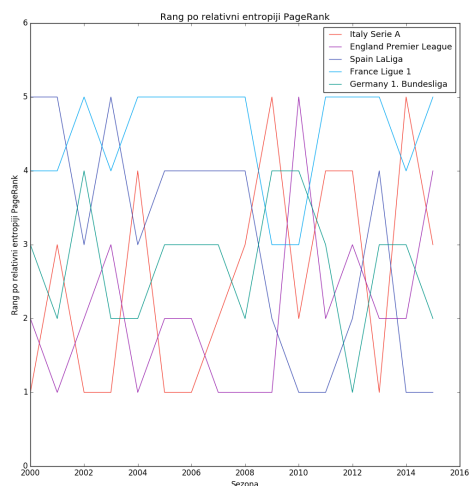
Spodaj so zbrani grafi relativne entropije PageRank, izračunane po metodi opisani v poglavju 3.3.3.2. Relativna entropija PageRank prikazuje mero izenačenosti tekmovanja oziroma lige v opazovani sezoni. V analizo smo vključili tekmovanja iz različnih celin in iz različnih ekipnih športov. Na grafih so zbrane ocene izenačenosti najmočnejših nogometnih, košarkaških, hokejskih, odbojcarskih in rokometnih lig v opazovanih sezonah. Za lažje primerjave med različnimi športi, so na določenih grafih skupaj prikazane lige različnih športov.



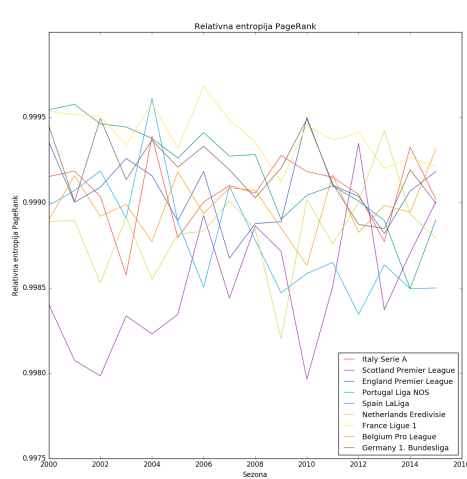
Slika A.1: Relativna entropija PageRank najboljših 5 evropskih nogometnih lig.



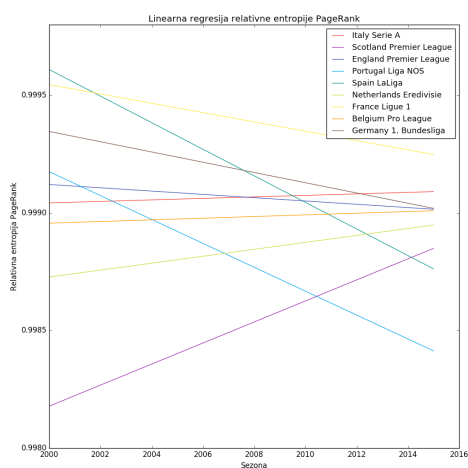
Slika A.2: Linearna regresija relativne entropije PageRank najboljših 5 evropskih nogometnih lig.



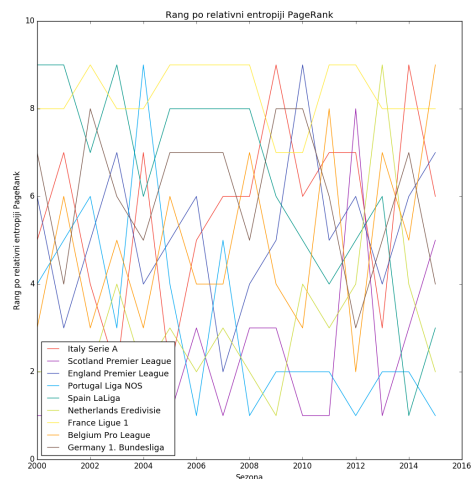
Slika A.3: Rang najboljših 5 evropskih nogometnih lig po relativni entropiji PageRank.



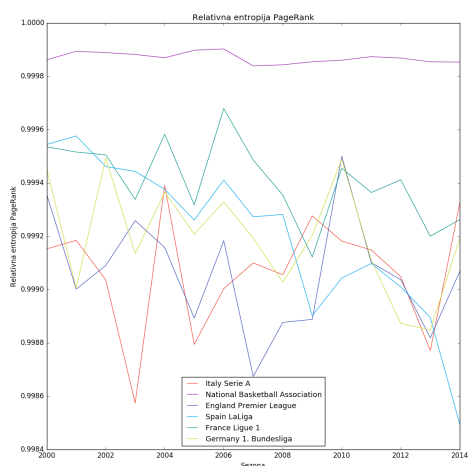
Slika A.4: Relativna entropija PageRank najboljših 9 evropskih nogometnih lig.



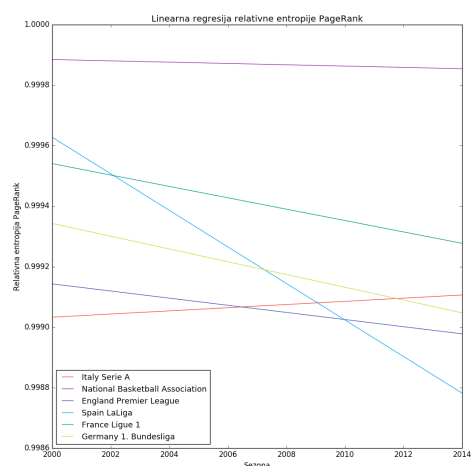
Slika A.5: Linearna regresija relativne entropije PageRank najboljših 9 evropskih nogometnih lig.



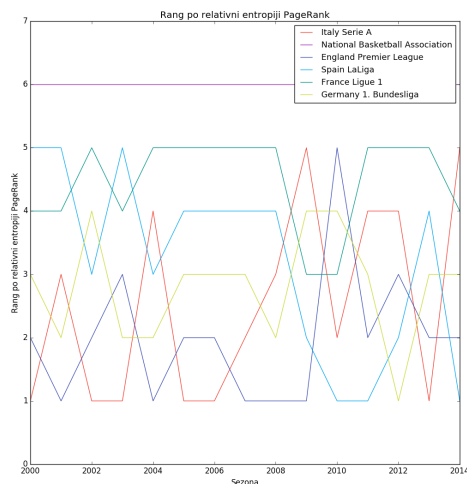
Slika A.6: Rang najboljših 9 evropskih nogometnih lig po relativni entropiji PageRank.



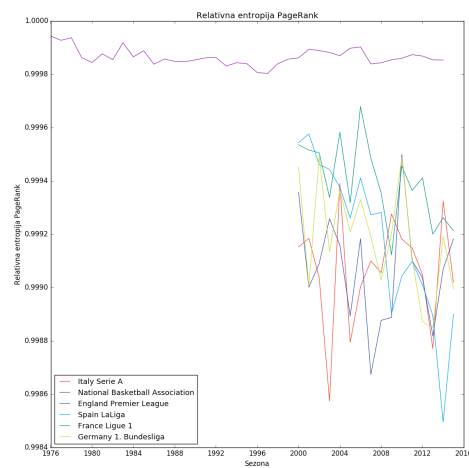
Slika A.7: Relativna entropija PageRank najboljših 5 evropskih nogometnih lig in košarkarske lige NBA.



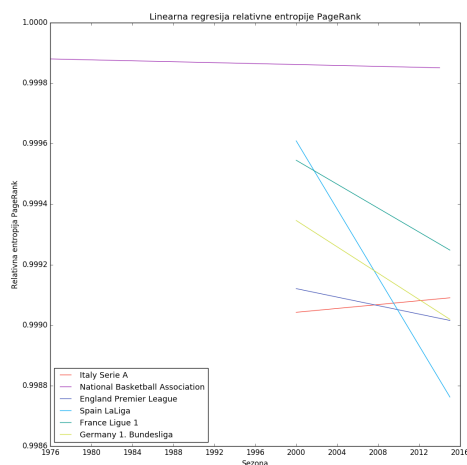
Slika A.8: Linearna regresija relativne entropije PageRank najboljših 5 evropskih nogometnih lig in košarkarske lige NBA.



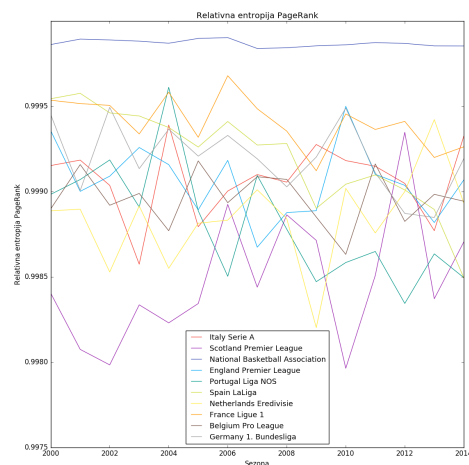
Slika A.9: Rang najboljših 5 evropskih nogometnih lig košarkarske lige NBA po relativni entropiji PageRank.



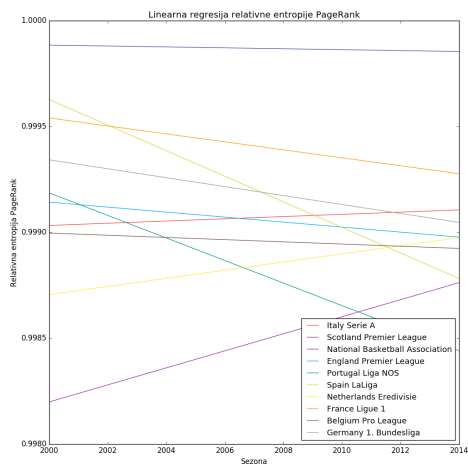
Slika A.10: Relativna entropija PageRank najboljših 5 evropskih nogometnih lig in košarkarske lige NBA.



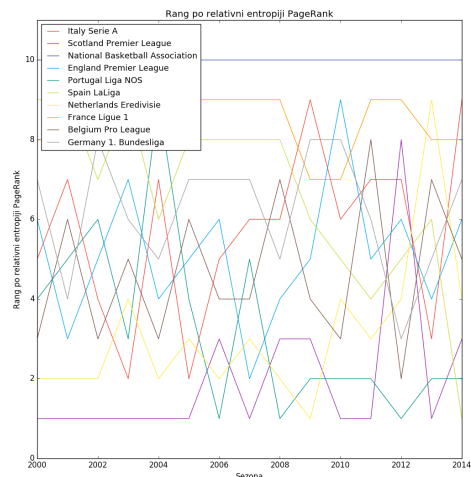
Slika A.11: Linearna regresija relativne entropije PageRank najboljših 5 evropskih nogometnih lig košarkarske lige NBA.



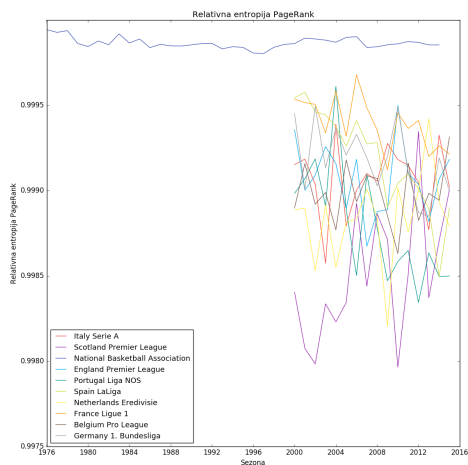
Slika A.12: Relativna entropija PageRank najboljših 9 evropskih nogometnih lig in košarkarske lige NBA.



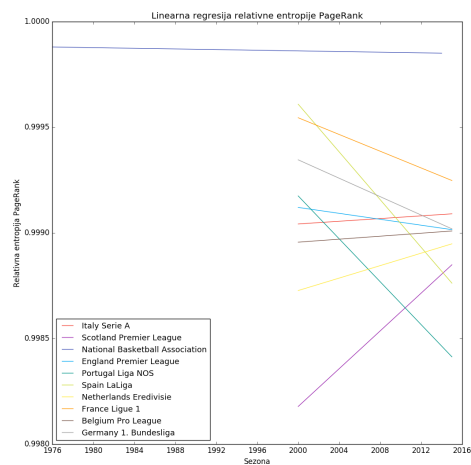
Slika A.13: Linearna regresija relativne entropije PageRank najboljših 9 evropskih nogometnih lig košarkarske lige NBA.



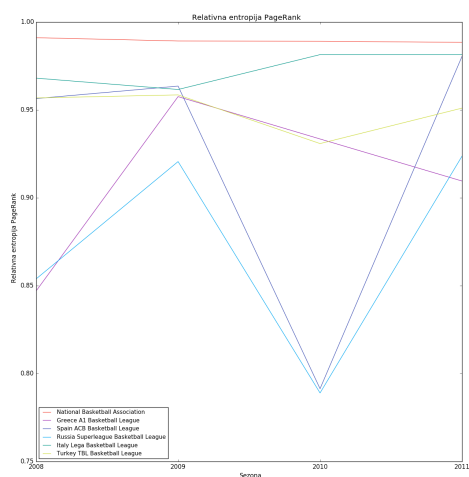
Slika A.14: Rang najboljših 9 evropskih nogometnih lig košarkarske lige NBA po relativni entropiji PageRank.



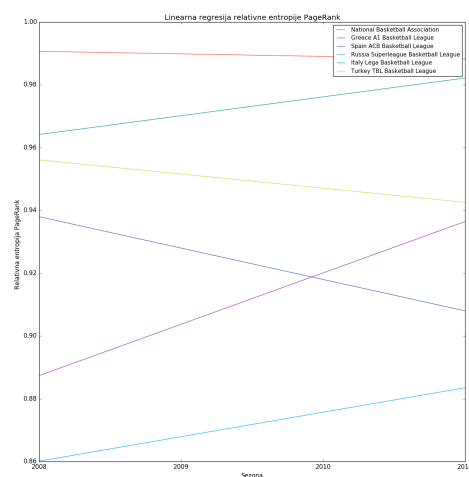
Slika A.15: Relativna entropija PageRank najboljših 9 evropskih nogometnih lig in košarkarske lige NBA.



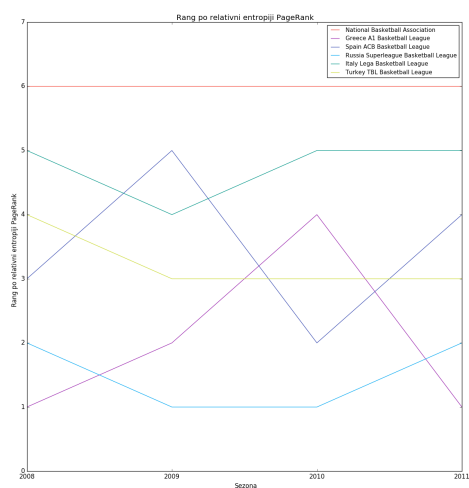
Slika A.16: Linearna regresija relativne entropije PageRank najboljših 9 evropskih nogometnih lig košarkarske lige NBA.



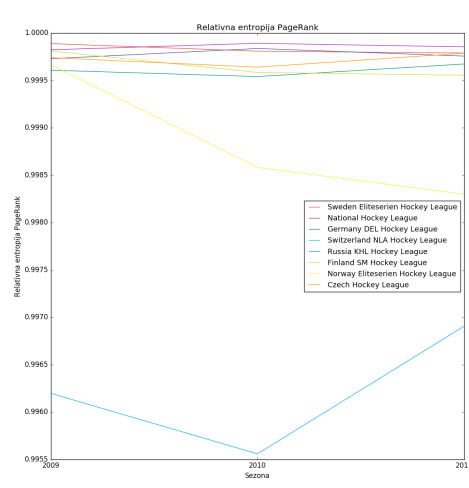
Slika A.17: Relativna entropija PageRank 6 opazovanih košarkarskih lig.



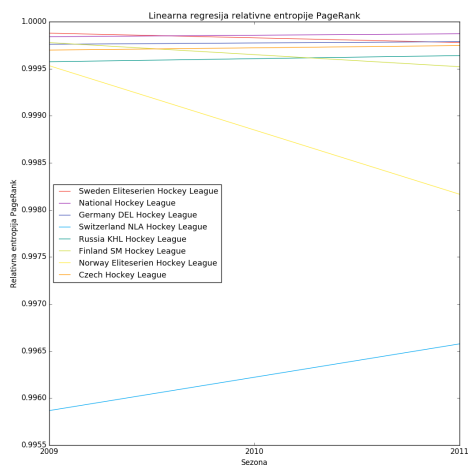
Slika A.18: Linearna regresija relativne entropije PageRank 6 opazovanih košarkarskih lig.



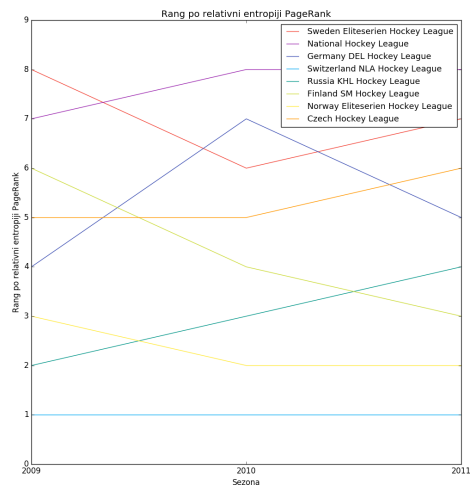
Slika A.19: Rang 6 opazovanih košarkarskih lig po relativni entropiji PageRank.



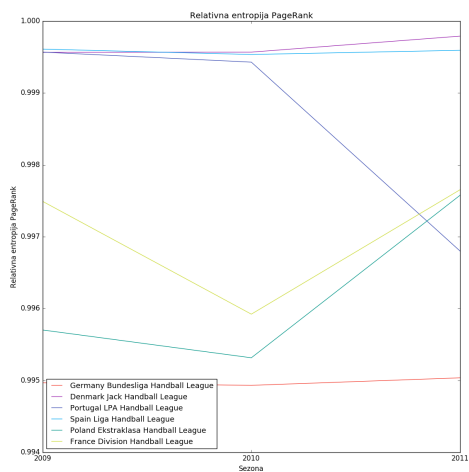
Slika A.20: Relativna entropija PageRank 8 opazovanih hokejskih lig.



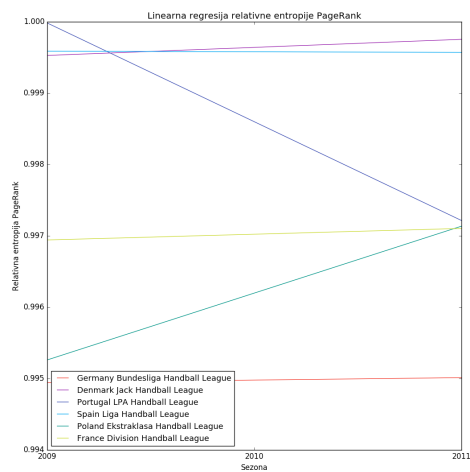
Slika A.21: Linearna regresija relativne entropije PageRank 8 opazovanih hokejskih lig.



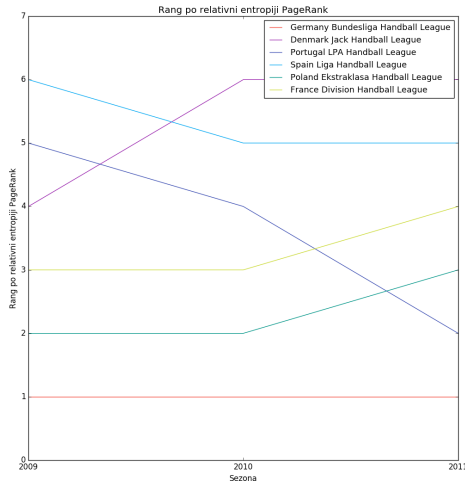
Slika A.22: Rang 8 opazovanih hokejskih lig po relativni entropiji PageRank.



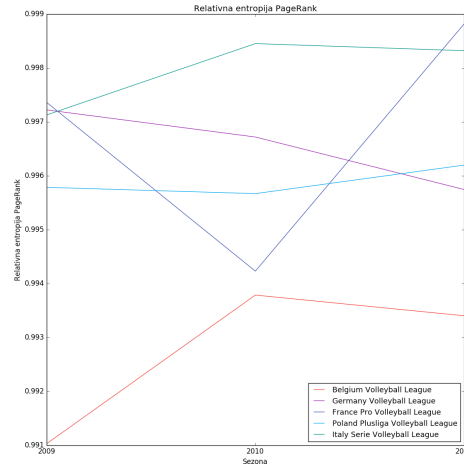
Slika A.23: Relativna entropija PageRank 6 opazovanih rokometnih lig.



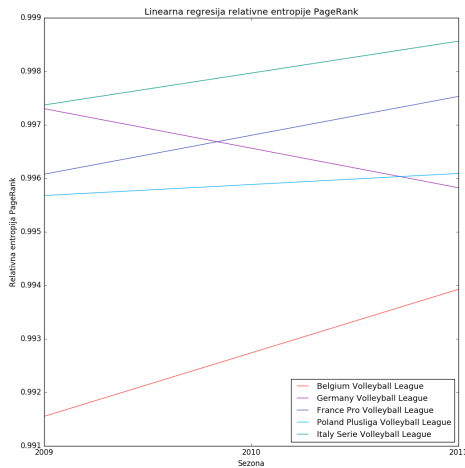
Slika A.24: Linearna regresija relativne entropije PageRank 6 opazovanih rokometnih lig.



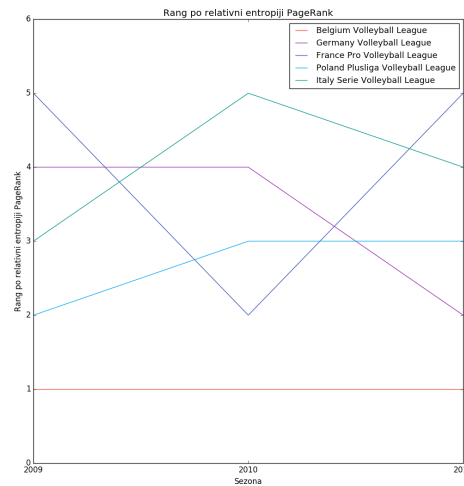
Slika A.25: Rang 6 opazovanih roketnih lig po relativni entropiji PageRank.



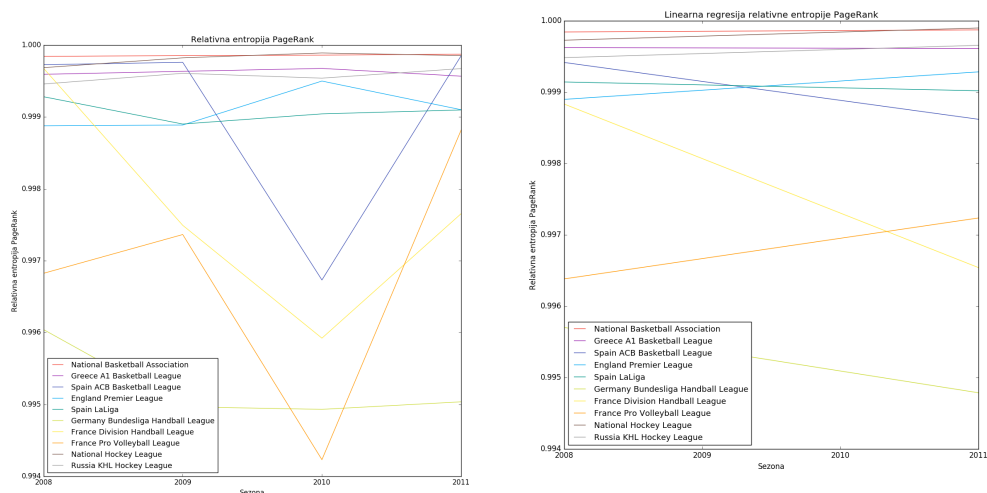
Slika A.26: Relativna entropija PageRank 5 opazovanih odbojgarskih lig.



Slika A.27: Linearna regresija relativne entropije PageRank 5 opazovanih odbojgarskih lig.

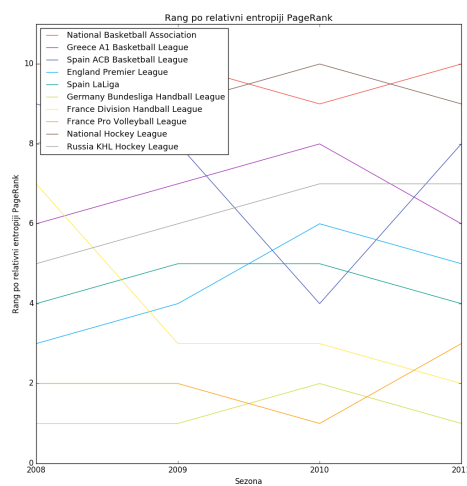


Slika A.28: Rang 5 opazovanih odbojgarskih lig po relativni entropiji PageRank.



Slika A.29: Relativna entropija PageRank nabora najboljših lig v vseh opazovanih športih.

Slika A.30: Linearna regresija relativne entropije PageRank nabora najboljših lig v vseh opazovanih športih.

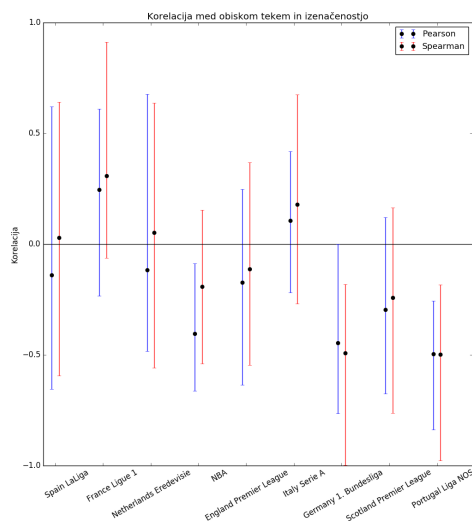


Slika A.31: Rang nabora najboljših lig v vseh opazovanih športih po relativni entropiji PageRank.

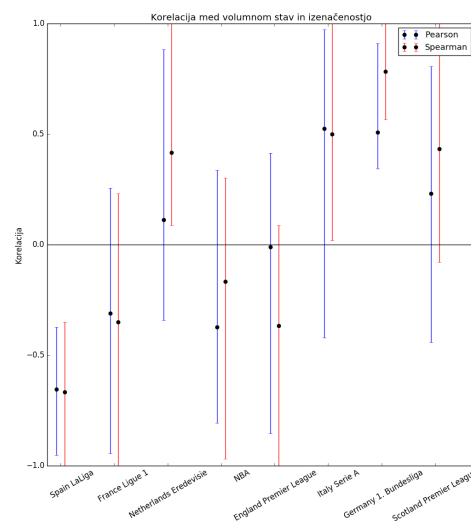
Dodatek B

Korelacija med izenačenostjo in zanimivostjo

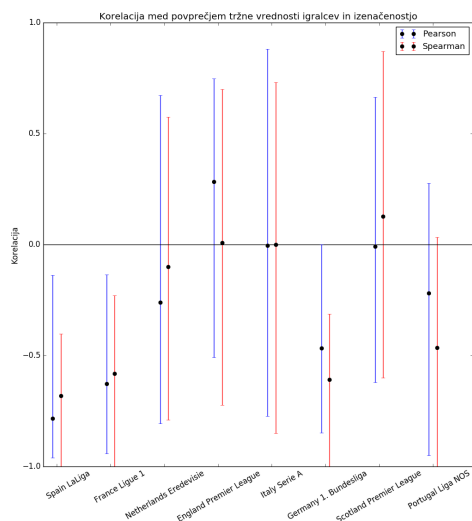
Spodnji grafi prikazujejo Pearsonov in Spearmanov koeficient korelacije med izenačenostjo in zanimivostjo tekmovanja oziroma lige. Zanimivost je definirana z različnimi metrikami, mero izenačenosti pa smo v večini primerov pridobili na podlagi predlagane metode za izračun relativne entropije Page-Rank tekmovanja.



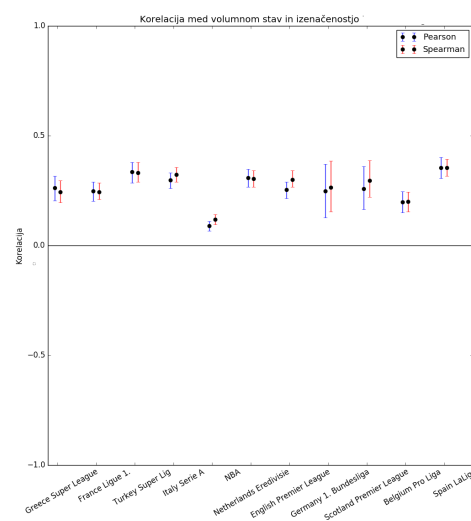
Slika B.1: Korelacija med relativno entropijo PageRank tekmovalna in obiskom na tekmah na nivoju sezone.



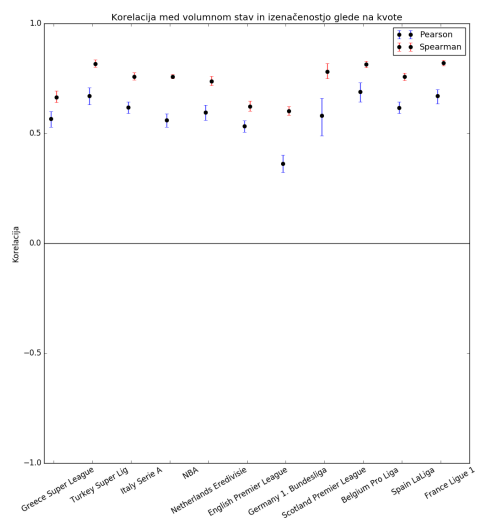
Slika B.2: Korelacija med relativno entropijo PageRank tekmovalna in volumnom stav na tekme tega tekmovanja na nivoju sezone.



Slika B.3: Korelacija med relativno entropijo PageRank in povprečno tržno vrednostjo ekipe tekmovanja na nivoju sezone.



Slika B.4: Korelacija med izenačenostjo glede na razliko v rangju po oceni PageRank in volumnom stav na nivoju tekme.



Slika B.5: Korelacija med izenačenostjo glede na stavne kvote in volumnom stav opazovanih tekmovalj na nivoju tekme.

Literatura

- [1] Sports Industry To Reach \$73.5 Billion By 2019. Dostopno na: <https://www.forbes.com/sites/darrenheitner/2015/10/19/sports-industry-to-reach-73-5-billion-by-2019> (pridobljeno 29.11.2017).
- [2] P. D. Staudohar, *Playing for dollars: Labor relations and the sports business*, Cornell University Press, 1996.
- [3] M. B. Callejo, F. J. Forcadell, Real Madrid football club: A new model of business organization for sports clubs in Spain, *Global Business and Organizational Excellence* 26 (1) (2006) 51–64.
- [4] J. P. Quirk, R. D. Fort, *Pay dirt: The business of professional team sports*, Princeton University Press, 1997.
- [5] D. Forrest, R. Simmons, Outcome uncertainty and attendance demand in sport: the case of English soccer, *Journal of the Royal Statistical Society: Series D (The Statistician)* 51 (2) (2002) 229–241.
- [6] J. L. Peña, H. Touchette, A network theory analysis of football strategies, arXiv preprint arXiv:1206.6904.
- [7] L. Page, S. Brin, R. Motwani, T. Winograd, *PageRank: Bringing order to the web*, Tech. rep., Stanford Digital Libraries Working Paper (1997).

-
- [8] NBA draft - Wikipedia. Dostopno na:
https://en.wikipedia.org/wiki/NBA_draft (pridobljeno 3.11.2017).
- [9] NBA salary cap - Wikipedia. Dostopno na:
https://en.wikipedia.org/wiki/NBA_salary_cap (pridobljeno 3.11.2017).
- [10] Luxury tax (sports) - Wikipedia. Dostopno na:
[https://en.wikipedia.org/wiki/Luxury_tax_\(sports\)](https://en.wikipedia.org/wiki/Luxury_tax_(sports)) (pridobljeno 3.11.2017).
- [11] F. Erdem, J. Ozen, Cognitive and affective dimensions of trust in developing team performance, *Team Performance Management: An International Journal* 9 (5/6) (2003) 131–135.
- [12] N. Masuda, M. A. Porter, R. Lambiotte, Random walks and diffusion on networks, arXiv preprint arXiv:1612.03281.
- [13] S. Shin, S. E. Ahnert, J. Park, Ranking Competitors Using Degree-Neutralized Random Walks, *PloS one* 9 (12) (2014) e113685.
- [14] A. Y. Govan, C. D. Meyer, Ranking national football league teams using google's pagerank, in: AA Markov Anniversary Meeting, 2006.
- [15] P. Rozenshtein, A. Gionis, Temporal PageRank (2016) 674–689.
- [16] (2016) Big Count. Dostopno na:
<http://www.fifa.com/worldfootball/bigcount/> (pridobljeno 2.2.2016).
- [17] R. Trequattrini, R. Lombardi, M. Battista, Network analysis and football team performance: a first application, *Team Performance Management* 21 (1/2) (2015) 85–110.
- [18] F. M. Clemente, M. S. Couceiro, F. M. L. Martins, R. S. Mendes, Using network metrics to investigate football team players' connections: A pilot study, *Motriz: Revista de Educação Física* 20 (3) (2014) 262–271.

-
- [19] J. Gama, P. Passos, K. Davids, H. Relvas, J. Ribeiro, V. Vaz, G. Dias, Network analysis and intra-team activity in attacking phases of professional football, *International Journal of Performance Analysis in Sport* 14 (3) (2014) 692–708.
- [20] M. Newman, *Networks: an introduction*, Oxford university press, 2010.
- [21] L. C. Freeman, A set of measures of centrality based on betweenness, *Sociometry* (1977) 35–41.
- [22] F. M. Clemente, F. M. L. Martins, M. S. Couceiro, R. S. Mendes, A. J. Figueiredo, A network approach to characterize the teammates' interactions on football: A single match analysis, *Cuadernos de Psicología del Deporte* 14 (3) (2014) 141–148.
- [23] Y. Yamamoto, K. Yokoyama, Common and unique network dynamics in football games, *PloS one* 6 (12) (2011) e29638.
- [24] C. Cotta, A. M. Mora, J. J. Merelo, C. Merelo-Molina, A network analysis of the 2010 FIFA world cup champion team play, *Journal of Systems Science and Complexity* 26 (1) (2013) 21–42.
- [25] F. M. Clemente, F. M. L. Martins, D. Kalamaras, P. D. Wong, R. S. Mendes, General network analysis of national soccer teams in FIFA World Cup 2014, *International Journal of Performance Analysis in Sport* 15 (1) (2015) 80–96.
- [26] S. Saavedra, S. Powers, T. McCotter, M. A. Porter, P. J. Mucha, Mutually-antagonistic interactions in baseball networks, *Physica A: Statistical Mechanics and its Applications* 389 (5) (2010) 1131–1141.
- [27] J. Park, M. E. Newman, A network-based ranking system for US college football, *Journal of Statistical Mechanics: Theory and Experiment* 2005 (10) (2005) P10014.

-
- [28] T. Callaghan, P. J. Mucha, M. A. Porter, Random walker ranking for NCAA division IA football, *American Mathematical Monthly* 114 (9) (2007) 761–777.
- [29] M. Tribušon, M. Lenič, Identifying top football players and springboard clubs from a football player collaboration and club transfer networks, arXiv preprint arXiv:1602.03664.
- [30] X. F. Liu, Y.-L. Liu, X.-H. Lu, Q.-X. Wang, T.-X. Wang, The anatomy of the global football player transfer network: Club functionalities versus network properties, *PloS one* 11 (6) (2016) e0156504.
- [31] K. Pelechrinis, E. Papalexakis, C. Faloutsos, Sportsnetrank: Network-based sports team ranking, in: *ACM SIGKDD Workshop on Large Scale Sports Analytics*, 2016.
- [32] K. Massey, *Statistical models applied to the rating of sports teams*, Bluefield College.
- [33] M. Franceschet, E. Bozzo, The Massey’s method for sport rating: a network science perspective, arXiv preprint arXiv:1701.03363.
- [34] E. Bozzo, M. Franceschet, The temporalized Massey’s method, arXiv preprint arXiv:1702.00585.
- [35] J. A. Morales, S. Sánchez, J. Flores, C. Pineda, C. Gershenson, G. Cocho, J. Zizumbo, R. F. Rodríguez, G. Iñiguez, Generic temporal features of performance rankings in sports and games, *EPJ Data Science* 5 (1) (2016) 33.
- [36] S. Mukherjee, Identifying the greatest team and captain—A complex network approach to cricket matches, *Physica A: Statistical Mechanics and its Applications* 391 (23) (2012) 6066–6076.
- [37] A. Aronson, "TeniRank": A new ranking of tennis players based on PageRank, arXiv preprint arXiv:1711.11122.

-
- [38] F. Radicchi, M. Perc, Who is the best player ever? A complex network analysis of the history of professional tennis, *PloS one* 6 (2) (2011) e17249.
- [39] J. Siegfried, A. Zimbalist, The economics of sports facilities and their communities, *The Journal of Economic Perspectives* 14 (3) (2000) 95–114.
- [40] M. B. Schmidt, D. J. Berri, Competitive balance and attendance: The case of Major League Baseball, *Journal of Sports Economics* 2 (2) (2001) 145–167.