

## METODAI SPORTO KOMANDOMS REITINGUOTI

Eimantas Palionis<sup>1</sup>, Viktoras Chadyšas<sup>2</sup>

Vilniaus Gedimino technikos universiteto matematinės statistikos katedra

Adresas: Saulėtekio al. 11, 10223 Vilnius, Lietuva.

El. paštas: <sup>1</sup> [ema.pal@gmail.com](mailto:ema.pal@gmail.com), <sup>2</sup> [viktoras.chadysas@vgtu.lt](mailto:viktoras.chadysas@vgtu.lt)

Gauta: 2015 m. liepa

Pataisyta: 2015 m. rugpjūtis

Paskelbta: 2015 lapkritis

**Santrauka.** Šiame straipsnyje nagrinėjami įvairūs metodai, kurie gali būti pritaikyti sporto komandoms reitinguoti. Naudojant realius Nacionalinės Krepšinio Asociacijos (NBA) 2014–2015 metų pirmos sezono pusės (nuo 2014-10-28 iki 2015-02-12) duomenis ir taikant skirtingus metodus, reitinguojamos rytų konferencijos komandos. Reitingavimo metodai palyginti tarpusavyje.

**Reikšminiai žodžiai:** sporto komandų reitingas, reitingavimo metodai, reitingavimo sąrašas.

### 1. Įvadas

Šiuolaikiniame pasaulyje žodis „reitingas“ yra labai svarbus. Kiekviena kompanija siekia turėti kuo aukštesnį reitingą, pirkėjai nori žinoti, kuris rinkos produktas yra įvertintas geriausiai, sporto fanai nori žinoti geriausias komandas. Dėl didelio skirtingų reitingų poreikio, buvo sukurta įvairiausių metodų kiekvienos srities reitingui nustatyti. Dažniausiai naudojami reitingavimo metodų pritaikymai: internetinių puslapių reitingavimas (metodai: *PageRank*, *HITS*, *SALSA* ir kt.), sporto komandų reitingavimas (metodai: *Massey*, *Colley*, *Keener*, *ELO* ir kt.), rekomendacinės sistemos filmams, knygoms (metodai: *Netflix movies*, *Amazon books* ir kt.).

Šiame straipsnyje apžvelgsime ir palyginsime metodus, kurie gali būti pritaikyti krepšinio komandoms reitinguoti (šie metodai taip pat gali būti pritaikyti ir kitų sporto šakų komandoms reitinguoti). Yra sukurta nemažai būdų sporto komandų reitingams skaičiuoti. Kiekvienas jų susideda iš įvairių komponentų, kurie padeda nuspręsti, kokią reitingavimo sistemą naudoti. Norint sudaryti efektyvų reitingavimo modelį, pirmiausia reikia nuspręsti, ką, konkrečiai, norima reitingavimu parodyti. Ar norima labiau akcentuoti komandų potencialą laimėti varžybas ateityje, ar parodyti jų situaciją, pavyzdžiui, įpusėjus sezonui. Vienas iš reitingavimui svarbių veiksnių yra namų arenos pranašumas. Kai susitinka dvi panašaus ar vienodo pajėgumo komandos, dažniau laimi ta komanda, kuri žaidžia namų arenoje. *Bleacherreport.com* [2] duomenimis, NBA lygoje kiekvienais metais namų arenose žaidžiančios komandos laimi maždaug 60 proc. rungtynių, šis rodiklis yra aukščiausias lyginant su kitomis didžiausiomis sporto lygomis Jungtinėse Amerikos Valstijose. Rungtynių grafiko sudėtingumo įtraukimas į reitingavimo sistemą yra labai naudingas reitinguojant komandas, kurios nėra žaidusios tarpusavyje. Pavyzdžiui, NCAA (Nacionalinėje Koledžų lygoje) dalyvauja 347 komandos, kurios gali sužaisti daugiausia 31 rungtynes (dažniausiai 27 arba 28). Todėl yra labai svarbu įvertinti, kokio pajėgumo komandos rungtyniavo tarpusavyje. Kiekvienų varžybų metu komandų pelnyti ir praleisti taškai kartais gali geriau atspindėti komandos pajėgumą, negu pergalės ir pralaimėjimai. Krepšinyje, ypač sezono pradžioje, galimos tokios situacijos, kad kelios komandos turi vienodą pergalių ir pralaimėjimų skaičių. Norint tiksliau įvertinti jų pajėgumą, reikia atsižvelgti į tai, koku skirtumu jie nugalėjo savo varžovus. Akivaizdu, kad komanda, kuri savo varžovus nugalėdama dideliais taškų skirtumais, turėtų reitingo lentelėje užimti aukštesnę vietą nei ta, kuri tik paskutinėmis sekundėmis įveikdavo oponentus.

Šio straipsnio tikslas yra išanalizuoti, realizuoti ir palyginti sporto reitingavimo metodus, naudojant realius 2014–2015 metų Nacionalinės Krepšinio Asociacijos (NBA) duomenis.

## 2. Reitingavimo metodai

Krepšinis – komandinis žaidimas, kuriame yra daugybė matematiškai labai sunkiai arba visai neaprašomų faktorių, pavyzdžiui, traumas, žaidėjų nuovargis, „komandinė dvasia“ ir t. t. [10]. Taip pat komandų vykdomi žaidėjų mainai yra sunkiai matematiškai interpretuojami, nes vieno žaidėjo perėjimas iš vienos komandos į kitą gali ir teigiamai ir neigiamai paveikti komandos rezultatus, priklausomai nuo to, kaip jis greitai įsilieja į komandos žaidimą ir ar apskritai pritampa prie komandos propaguojamo žaidimo stiliaus. Šiame darbe į pateiktus ir panašius faktorius nebus atsižvelgiama, o nagrinėjami tik tokie matematiškai vertinami kriterijai kaip pergalės, pralaimėjimai, įmestų ir praleistų taškų skaičius, rungtynių sudėtingumas ir t. t.

Colley metodas [1], komandoms reitinguoti, remiasi tik komandų pergalėmis ir pralaimėjimais darant prielaidą, kad lygiosios yra negalimos. Komandų reitingams, vektoriui  $\vec{r}$  rasti, sprendžiama lygtis

$$C\vec{r} = \vec{b}.$$

Čia  $i$ -tasis vektoriaus  $\vec{b}$  elementas apibrėžiamas

$$b_i = 1 + \frac{w_i - l_i}{2},$$

kur  $w_i$  ir  $l_i$  yra atitinkamai komandos  $T_i$  pergalių ir pralaimėjimų skaičius. Matrica  $C$  yra vadinama Colley matrica ir apibrėžiama

$$C_{ij} = \begin{cases} -n_{ji}, & \text{jei } i \neq j, \\ 2 + n_i, & \text{jei } i = j, \end{cases}$$

kur  $n_{ji}$  yra komandos  $T_i$  ir  $T_j$  tarpusavio rungtynių skaičius, o  $n_i$  – komandos  $T_i$  sužaistų rungtynių skaičius. Šio metodo privalumas yra tas, kad atsižvelgiant vien tik į pergalių ir pralaimėjimų skaičių, jį galima pritaikyti beveik bet kurio sporto komandoms reitinguoti. Metodas taip pat atsižvelgia į priešininkų komandos stiprumą (įveikus stiprią komandą, reitingas pakyla labiau, nei nugalėjus silpną).

Viena iš galimų šio metodo modifikacijų yra į modelį įtraukti komandos pergalės ar pralaimėjimo skirtumą. Tarkime, jeigu komanda  $T_i$  laimi prieš komandą  $T_j$  dešimties arba didesne taškų persvara, tai komandos  $T_i$  pergalės koeficientas būtų dauginamas iš  $c=1,5$ , priešingu atveju –  $c=1$ . Tokiu būdu labiau atsižvelgiama į komandos pajėgumą ir užtikrintai savo priešininkus nugalėjusi komanda gauna pranašumą prieš komandas, kurios oponentus įveikia vos kelių taškų skirtumu. Taigi  $i$ -tasis vektoriaus  $\vec{b}$  elementas apibrėžiamas

$$b_i = 1 + \frac{cw_i - l_i}{2}.$$

Jutos universiteto (*University of Utah*) matematikos profesorius James P. Keener pasiūlė dar vieną matricinį metodą komandoms reitinguoti. Keener matrica remiasi rezultatų santykiu, kuris gaunamas pasinaudojant Laplaso formule [13]:

$$\frac{S_{ij} + 1}{S_{ij} + S_{ji} + 2},$$

čia  $S_{ij}$  – komandos  $T_i$  pelnyti taškai prieš komandą  $T_j$ , atitinkamai  $S_{ji}$  – komandos  $T_j$  pelnyti taškai prieš komandą  $T_i$ . Šią formulę taip pat galime interpretuoti kaip tikimybę komandai  $T_i$  nugalėti komandą  $T_j$  kitame jų susitikime.

Kadangi skaičiavimams naudojami rungtynių rezultatai, norėdamas išvengti didelių reitingavimo paklaidų dėl galimo rungtynių rezultato manipuliavimo, Keener pasiūlė funkciją

$$h(x) = \frac{1}{2} + \frac{1}{2} \operatorname{sgn}\left(x - \frac{1}{2}\right) \sqrt{|2x - 1|},$$

kuri padeda adekvačiau žvelgti į komandų pelnomus taškus. Pastaroji taip pat labai naudinga tose situacijose, kai komanda paskutinėmis rungtynių minutėmis supratusi, kad jau nebeįveiks varžovų, į aikštę išleidžia mažiau žaidžiančius komandos žaidėjus, tokiu būdu tikėtina, kad komanda pralaimės didesniu skirtumu. Keener matricos  $K$  elementai apibrėžiami taip:

$$K_{ij} = \begin{cases} h\left(\frac{S_{ij} + 1}{S_{ij} + S_{ji} + 2}\right), & \text{jei komandos } T_i \text{ ir } T_j \text{ žaidė tarpusavyje} \\ 0, & \text{priešingu atveju} \end{cases}$$

Gauta matrica  $K$  yra neneigiama, o komandų reitingai, vektoriūs  $\vec{r}$ , randamas iš lygties

$$K\vec{r} = \lambda\vec{r},$$

kur  $\lambda$  – matricos  $K$  tikrinės reikšmės. Šis metodas yra puikus komandų palyginimui tiesiogiai tarpusavyje, nesunkiai skaičiuojami reitingai. Esminis metodo trūkumas yra tas, kad komandos turi būti sužaidusios tiek pat kartų, kas daugelyje sporto šakų, kol dar nesibaigęs sezonas, pasitaiko retai, todėl skaičiuojant reitingus naudojami rungtynių rezultatų vidurkiai.

Dar vienas, pakankamai dažnai reitingavime naudojamas, ELO metodas sezono pradžioje kiekvienai komandai priskiria reitingą 1500. Komandų reitingai kinta intervale [100; 2800]. Kiekvienoms rungtynėms yra skaičiuojamas tikėtinas procentinis komandos pergalės koeficientas. Tarkime, turime komandas  $A$  ir  $B$ , kurių reitingai yra  $\tau_A$  ir  $\tau_B$ , o rungtynių baigtys (komandų pelnyti taškai) yra  $s_A$  ir  $s_B$ , tada tikėtinas komandos  $A$  rezultatas skaičiuojamas taip:

$$E_A = \frac{1}{1 + 10^{(\tau_B - \tau_A)/400}}$$

Tada perskaičiuojamas komandos reitingas:

$$\tau_{Anaujas} = \tau_{Asenas} + K \cdot (s_A - E_A),$$

kur  $K$  yra konstanta, kuri parenkama pagal tai, kiek norima pabrėžti pergalės ar pralaimėjimo svarbą, dažniausiai ši konstanta yra sveikasis skaičius intervale [10;40]. Šis reitingas nėra labai tikslus, kol komandos reitingas nėra perskaičiuotas pakankamai kartų. NBA reguliariojo sezono atveju, reitingas tikslesnius rezultatus komandos ateities perspektyvai duotų įpusėjus sezonui. Šis metodas ypač tinkamas reitinguoti komandas arba žaidėjus turnyruose, kuriuose yra labai didelis dalyvių skaičius.

Dar vienas metodas, kuris gali būti naudojamas sporto komandoms reitinguoti, tai Massey mažiausių kvadratų metodas. Sakykime, kad  $Y$  yra atsitiktinis dydis, apibrėžiantis dviejų komandų  $A$  ir  $B$  susitikimo rezultatų skirtumą. Tarkime, kad komandų pergalės persvara yra tiesiogiai proporcinga jų reitingams. Tada vidurkis

$$E[Y] = r_a - r_b,$$

kur  $r_a$  – komandos  $A$  reitingas,  $r_b$  – komandos  $B$  reitingas. Rezultatas, gautas naudojant pastarąją formulę parodys vidutinį varžybų rezultato skirtumą, jei rungtynės būtų žaidžiamos daug kartų.

Akivaizdu, kad neturint komandų reitingų šių rezultatų apskaičiuoti nepavyks. Taigi Massey pateikė būdą šiems reitingams apskaičiuoti [4]. Tarkime, kad kiekvienose sužaistose rungtynėse tarp komandų  $A$  ir  $B$  yra apskaičiuojamas rezultato skirtumas  $s$ :

$$s = a - b,$$

kur  $a$  ir  $b$  yra komandų pelnytų taškų skaičius. Jeigu iš viso sezono dalyvauja  $n$  komandų ir jos sužaidžia  $m$  varžybų, kurių kiekvienai yra skaičiuojamas rezultatų skirtumas, galima sudaryti lygčių sistemą su  $m$  lygčių ir  $n$  nežinomųjų. Šiai lygčių sistemai spręsti sudaroma matrica  $X_{m \times n}$ . Šios matricos eilutėse yra nurodomos komandų  $A$  ir  $B$  varžybų baigtys. Kiekvienoje eilutėje 1 žymi nugalėjusią komandą, o  $-1$  – pralaimėjusią. Visur kitur rašomas 0. Matricos  $X$  elementai apibrėžiami atitinkamai

$$x_{ki} = \begin{cases} 1, & \text{jei komanda } T_i \text{ laimėjo } k - \text{tąsias rungtynes,} \\ -1, & \text{jei komanda } T_i \text{ pralaimėjo } k - \text{tąsias rungtynes,} \\ 0, & \text{priešingu atveju.} \end{cases}$$

Komandų reitingų vektorius  $\vec{r}$  randamas išsprendus lygtį

$$X\vec{r} = Y,$$

kur  $Y$  –  $m \times 1$  eilės, kiekvienų varžybų rezultato skirtumo  $s$ , matrica stulpelis. Akivaizdu, kad visų komandų tarpusavyje sužaistų rungtynių skaičius bus didesnis, nei yra dalyvaujančių komandų skaičius. Norint gauti sprendinius, abi lygties pusės dauginamos iš transponuotos matricos  $X^T$ . Gaunama  $X^T X r = X^T Y$ . Pažymėjus  $X^T X = M$ , o dešinę lygties pusę  $p$ , ieškodami  $\vec{r}$  sprendžiame lygtį

$$M\vec{r} = \vec{p},$$

kur  $M$  yra  $n \times n$  eilės matrica,  $\vec{p}$  –  $n \times 1$  dydžio visų varžybų taškų skirtumų sumos vektorius. Matricos  $M$  diagonalėje skaičius  $M_{ii}$  nurodo  $i$ -tosios komandos sužaistų rungtynių skaičių, o neigiamas skaičius  $M_{ij}$ ,  $i \neq j$  nurodo, kiek kartų  $i$ -toji komanda žaidė prieš  $j$ -tąją komandą. Tačiau bet kurios matricos eilutės ir bet kurio stulpelio reikšmių suma yra lygi 0, tad stulpeliai yra priklausomi ir matrica turės be galo daug sprendinių. Šiai problemai pašalinti paskutinė matricos  $M$  eilutė pakeičiama vienetais, o paskutinis vektorius  $\vec{p}$  elementas – nuliu, kad matricos rangas nebūtų mažesnis už  $n$ . Atlikus šią transformaciją, įmanoma išspręsti sistemą ir gauti komandų reitingų vektorių  $\vec{r}$ . Šis metodas yra paremtas

mažiausių kvadratų metodu paklaidoms minimizuoti, todėl yra matematiškai pagrįstas ir gan nesudėtingas. Taip pat norint nustatyti komandų reitingus užtenka žinoti, tik kiek kartų komandos susitiko ir kokių rezultatu baigėsi varžybos.

Procentinio indekso modelis (RPI) [11] buvo sukurtas Nacionalinės Koledžų Asociacijos (NCAA) 1981 m., kad būtų galima nustatyti, kokiose pozicijose išrikiuoja komandos ir kurios iš jų patenka į sezono pabaigoje vykstančias atkrentamąsias varžybas. Šis modelis naudoja komandos pergalių procentą reitingo indeksui nustatyti [9]. Tarkime,  $pp$ ,  $ppp$  ir  $pppp$  yra atitinkamai komandos pergalių procentas, tos komandos priešininkų pergalių procentas ir tos komandos priešininkų priešininkų pergalių procentas. Pergalių procentas ( $pp$ ) yra skaičiuojamas kaip santykis laimėtų rungtynių ir sužaistų rungtynių. Priešininkų pergalės procentas ( $ppp$ ) nagrinėjama komandai randamas skaičiuojant kiekvienos priešininkų komandos pergalių ir pralaimėjimų santykio vidurkį, į skaičiavimus neįtraukiant nagrinėjamos komandos. Komandos priešininkų priešininkų pergalių procentas ( $pppp$ ) randamas imant vidurkį kiekvienos iš priešininkų komandų  $ppp$ . Kiekvienai komandai RPI indeksas skaičiuojamas taip:

$$RPI = 0,25 * pp + 0,5 * ppp + 0,25 * pppp.$$

Kadangi NCAA lygoje, priešingai nei NBA, visos komandos tarpusavyje neturi galimybių sužaisti kiekviena su kiekviena, komandos pagal šį matematinį modelį yra išrenkamos į atkrentamąsias varžybas. Dėl šio metodo paprastumo ir svorių paskirstymo, jis sulaukia nemažai kritikos. Koledžų krepšinio analitikas Ken Pomeroy [7] pastebi, kad indeksas ne visai sąžiningai vertina rungtynių grafiko sudėtingumą. Komandos rungtynių grafiko sudėtingumas šiame indekse sudaro net 75 %, o atsižvelgiant į tai, kad komandos nėra suskirstytos į atskiras grupes pagal pajėgumą, įmanomas ir toks variantas, kad geriausia lygos komanda rungtyniaus vien tik prieš pačias silpniausias komandas ir jo reitingas bus mažesnis, už prastesnės komandos, kuri turi mažiau pergalių, tačiau žaidė prieš stiprius oponentus. Vadinas, kartais komandos priešininkų pajėgumas nulemia jų vietą reitinge labiau, nei jų pačių pergalės ar pralaimėjimai. Taip pat indeksas neatsižvelgia į komandų pelnomų ir praleistų taškų skaičių, nors akivaizdu, kad komanda nugalėjusi varžovus dideliu taškų skirtumu yra stipresnė, už komandą, kuri tuos pačius varžovus nurungė vos keliais taškais. Nors šis reitingavimo metodas neturi pagrįstų statistika besiremiančių įrodymų, tačiau jis tebėra naudojamas. Viena iš priežasčių yra komandų tarpusavio rezultatų manipuliavimas, t. y. abipusis susitarimas dėl palankios rungtynių baigties abiem komandoms.

Šiuo metu viena populiariausių ir labiausiai nagrinėjamų bei komentuojamų yra buvusio ESPN.com žurnalisto (dabartinio Memfio „Grizzlies“ krepšinio operacijų viceprezidento) John Hollinger pristatyta Power Rankings sistema. Šis reitingavimas vadovaujasi keliais, daugelio ekspertų manymu svarbiausiais, krepšinio aspektais: vidutinis komandos pelnytų ir praleistų taškų skirtumas ( $MARG$ ); vidutinis komandos pelnytų ir praleistų taškų skirtumas per paskutines 10 sužaistų rungtynių ( $MARGL10$ ); rungtynių grafiko sudėtingumas (vidutinis priešininkų pergalių koeficientas) ( $SOS$ ); rungtynių grafiko sudėtingumas (vidutinis priešininkų pergalių koeficientas) per paskutines 10 sužaistų rungtynių ( $SOSL10$ ); kiek rungtynių komanda yra sužaidusi namų ar priešininkų arenoje per sezoną ( $HOME$ ,  $ROAD$ ); kiek rungtynių komanda yra sužaidusi namų ar priešininkų arenoje per paskutines 10 sužaistų rungtynių ( $HOME10$ ,  $ROAD10$ ). Modelis, kurį pasiūlė Hollinger, komandoms reitinguoti yra:

$$r = \left( \left( \frac{SOS - 0,5}{0,037} \right) * 0,67 \right) + \left( \left( \frac{SOSL10 - 0,5}{0,037} \right) * 0,33 \right) + 100 \\ + 0,67 * \left( MARG + \left( \frac{(ROAD - HOME) * 0,35}{GAMES} \right) \right) + \left( 0,33 * \left( MARGL10 + \left( \frac{(ROAD10 - HOME10) * 3,5}{10} \right) \right) \right).$$

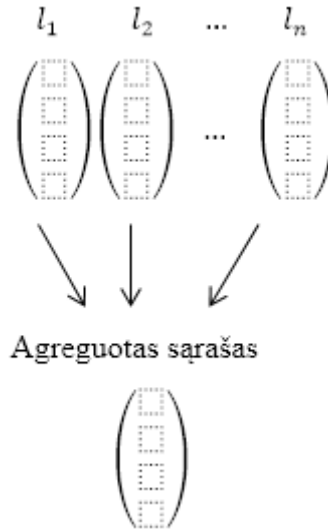
Šį modelį galima interpretuoti taip: pergalės namų arenoje lemia mažiau, nei pergalės išvykoje, taip pat komandos pergalė namuose mažesniu nei 3 taškų skirtumu komandos reitingui turi neigiamą įtaką; 1/3 viso komandos reitingo sudaro paskutinių 10 sužaistų varžybų rezultatai; formulė stipriai remiasi komandos įmestų ir praleistų taškų skaičiumi, t. y. komandos gebėjimu pulti ir gintis. Perfrazuojant paties Hollinger komentarą apie modelį, jis remiasi įmestais ir praleistais taškais, nes toks vertinimas geriau atspindi komandos pajėgumą nei įprastas pergalių ir pralaimėjimų santykis.

Tačiau, kaip pastebėjo vienas reitingavimų sistemas analizuojantis sporto analitikas Kevin Rossen [8], šis metodas turi ir keletą trūkumų. Jis teigia, kad pagal šį metodą, komandos pergalės namų arenoje turi labai mažai įtakos, jos kilimui reitinguose.

Komandų reitingams skaičiuoti ir reitingavimo sąrašams sudaryti galime naudoti vieną ar kitą metodą, tačiau kaip palyginti taikytais metodais gautus rezultatus ir kuris iš jų yra geriausias, aptarsime kitame skyriuje.

### 3. Geresni reitingai ir reitingavimo sąrašų palyginimas

Siekiant gauti geresnius reitingus yra naudojama reitingavimo sąrašų agregacija. Šis būdas apjungia visais metodais gautus sąrašus į vieną jungtinį sąrašą, kuris gaunamas apjungiant gautus rezultatus pagal kokį nors požymį (pvz., vidurkį). Agregaciją grafiškai galima pavaizduoti taip:



1 pav. Agreguoto sąrašo sudarymas

Sakykime, turimi reitingavimo sąrašai  $l_1, l_2, \dots, l_n$ , reitinguojamos komandos  $x_1, x_2, \dots, x_m$ , o  $r_{mn}$  – komandos užimama vieta reitingavimo sąraše. Agreguotas reitingavimo sąrašas sudarytas iš vidurkio komandų pozicijų, kurios gaunamos realizavus kiekvieną iš analizuojamų metodų komandų reitingams sudaryti.

Turint jau sudarytus komandų reitingavimo sąrašus yra aktualus jų palyginimas. Yra nemažai būdų, kuriais galima būtų palyginti reitingavimo sąrašus [12]. Šiame straipsnyje tam tikslui pasirinktas vienas iš dažniausiai naudojamų – Kendalo  $\tau$  koreliacijos koeficientas [3]. Jis parodo priklausomybę tarp dviejų reitingavimo sąrašų pagal komandų išsidėstymą juose. Sakykime, turimi du skirtingi reitingavimo sąrašai  $X$  ir  $Y$ , kurių elementai yra sugrupuoti  $(x_1; y_1), (x_2; y_2), \dots, (x_n; y_n)$ . Skaičiuojamas Kendalo koeficientas  $\tau$ :

$$\tau = \frac{n_c - n_d}{C_n^2},$$

kur  $n_c$  – sutampančių porų skaičius,  $n_d$  – nesutampančių porų skaičius. Sutampančios poros bus tada, kai  $x_i < x_j$  ir  $y_i < y_j$ , arba  $x_i > x_j$  ir  $y_i > y_j$ , nesutampančios, kai  $x_i < x_j$  ir  $y_i > y_j$ , arba  $x_i > x_j$  ir  $y_i < y_j$ , nei sutampančio nei nesutampančios, kai  $x_i = x_j$  ir  $y_i = y_j$ . Koeficientas  $\tau$  kinta intervale  $-1 \leq \tau \leq 1$ , kai  $\tau = 1$  reitingų sąrašai visiškai sutampa, o kai  $\tau = -1$ , sąrašai – atvirkštiniai.

### 4. Statistinis tyrimas

Anksčiau pristatytų metodų sporto komandoms reitinguoti realizacijai ir palyginimui naudojame realius duomenis. Reitinguojamos Nacionalinės Krepšinio Asociacijos (NBA) rytų konferencijos komandos. Jų tarpusavio rungtynės imamos nuo 2014-10-28 (sezono pradžios) iki 2015-02-12 (sezono vidurio) imtinai [6]. Komandos ir jų rezultatai pateikiami 1 lentelėje. Realizuojami visi 2 skyriuje pateikti reitingavimo metodai. Imant visais metodais gautų vietų reitinge vidurkį gaunama kiekvienos komandos agreguota vieta reitingavimo sąraše (2 lentelė). Kiekvienas reitingavimo metodas lyginamas su agreguotu ir apskaičiuojamas Kendalo koeficientas  $\tau$ . Laikoma, kad agreguotas reitingavimo sąrašas geriausiai atspindi komandos pajėgumą, todėl ieškoma, kuris metodas geriausiai koreliuoja su agreguotu pagal vietą sąraše (3 lentelė).

**1 lentelė.** NBA rytų konferencijos komandos, jų rezultatai

Komanda	Sužaistų rungtynių skaičius	Pergalės	Pralaimėjimai
„Atlanta Hawks“	35	28	7
„Boston Celtics“	31	12	19
„Brooklyn Nets“	34	14	20
„Charlotte Hornets“	31	16	15
„Chicago Bulls“	31	20	11
„Cleveland Cavaliers“	31	20	11
„Detroit Pistons“	31	13	18
„Indiana Pacers“	30	14	16
„Miami Heat“	31	15	16
„Milwaukee Bucks“	33	20	13
„New York Knicks“	33	6	27
„Orlando Magic“	35	11	24
„Philadelphia 76ers“	31	8	23
„Toronto Raptors“	34	24	10
„Washington Wizards“	35	22	13

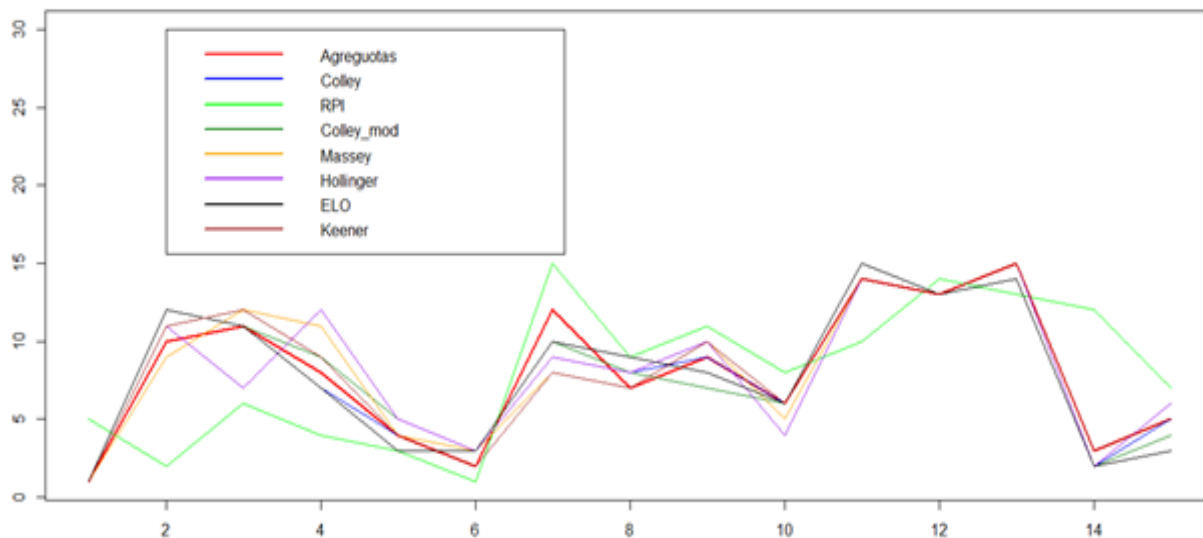
**2 lentelė.** Komandų reitingai

Komanda	RPI	Colley	Colley modifikuotas	Hollingers modelis	Massey	ELO	Keener	Agreguotas sąrašas
„Atlanta Hawks“	5	1	1	1	1	1	1	1
„Boston Celtics“	2	12	12	11	9	12	11	10
„Brooklyn Nets“	6	11	11	7	12	11	12	11
„Charlotte Hornets“	4	7	9	12	11	7	9	8
„Chicago Bulls“	3	4	5	5	4	3	4	4
„Cleveland Cavaliers“	1	3	3	3	3	3	2	2
„Detroit Pistons“	15	10	10	9	8	10	8	12
„Indiana Pacers“	9	8	8	8	7	9	7	7
„Miami Heat“	11	9	7	10	10	8	10	9
„Milwaukee Bucks“	8	6	6	4	5	6	6	6
„New York Knicks“	10	15	15	14	14	15	14	14
„Orlando Magic“	14	13	13	13	13	13	13	13
„Philadelphia 76ers“	13	14	14	15	15	14	15	15
„Toronto Raptors“	12	2	2	2	2	2	3	3
„Washington Wizards“	7	5	4	6	6	3	5	5

**3 lentelė.** Kendalo koeficientas  $\tau$  apskaičiuotas rangams agreguoto sąrašo su gautais kiekvienu kitu metodu

	RPI	Colley	Colley modifikuotas	Hollingers modelis	Massey	ELO	Keener
Agreguotas sąrašas	0,39	0,89	0,85	0,77	0,83	0,84	0,92

Iš rezultatų matyti, kad geriausiai agreguotą reitingų sąrašą atspindi Keenerio metodas, prasčiausiai – RPI. Tai parodo ir grafinis komandų vietų reitinguose vaizdas (2 pav.), kuriame aiškiai matyti, kad RPI metodo kreivė labiausiai nukrypsta nuo kitų metodų kreivių.



2 pav. Komandų rangai gauti naudojant skirtingus metodus

Norint patikrinti kiekvieno metodo prognozuojamąjį potencialą, kiekvieno metodo reitingų sąrašai, naudojant Kendalo koeficientą  $\tau$ , lyginami su sezono pabaigos galutiniais (žinomais) reitingais [5] (4 lentelė).

4 lentelė. Kendalo koeficientas  $\tau$ , apskaičiuotas rangams sezono pabaigos su gautais kiekvienu kitu metodu

	RPI	Colley	Colley modifikuotas	Hollingers modelis	Massey	ELO	Keener	Agreguotas sąrašas
Sezono pabaiga	0,45	0,75	0,75	0,79	0,77	0,70	0,75	0,83

Geriausiai sezono pabaigą atspindi agreguotas sąrašas. Pastarasis rezultatas nėra netikėtas, nes agreguotas sąrašas sudarytas iš visų taikytų metodų rezultatų, t. y. juose yra daugiausia įvairių faktorių. O iš visų metodų geriausiai sezono pabaigą prognozuoja *Hollinger Power Rankings* metodas, prasčiausiai – RPI metodas.

## 5. Išvados

Reitingavimas yra skirtas kuo tiksliau atspindėti komandos dabartinę pajėgumą, skirtingi metodai atsižvelgia į skirtingus kriterijus, priklausomai nuo to, ką reitinguotojas laiko svarbiausiais faktoriais. Vieniems tai pergalės ir pralaimėjimai, kitiems – pelnyti taškai, tretiems – kokio sudėtingumo komandos grafikas. Lažybų kontoroms nuolatiniai reitingų perskaičiavimai yra labai svarbūs koeficientų sudarymams, kad būtų kuo tiksliau įvertintas komandos pajėgumas ar kokia persvara galima tikėtis vienos ar kitos komandos pergalės būsimose rungtynėse. Reitingai neprognozuoja tolimes ateities, bet gerai atspindi dabartinę komandos situaciją lygoje.

Reitingavimas neturi nustatytų griežtų ribų, kaip ir kada jis turi būti atliekamas, todėl reitinguojant komandas svarbiausia yra atsižvelgti į tai, ką norima juo pasiekti. Norint pažiūrėti, kaip šiuo metu komandoms sekasi kokioje nors lygoje, užtenka paskaičiuoti jų pergalių ir pralaimėjimų santykį, jeigu norima sužinoti, kaip komandai sekasi atsižvelgiant į tai prieš kokio pajėgumo komandas jai teko žaisti, galima naudoti RPI metodą arba Colley matricinį metodą, kuris gali būti modifikuojamas ir pagal dar daugiau kriterijų, pavyzdžiui, kokia persvara buvo pasiekta pergalė. ELO metodą reikėtų rinktis reitinguojant turnyrus, kuriuose dalyvauja daug komandų / žaidėjų. Massey metodas tinka ne tik reitinguoti komandas – komandų reitingų skirtumai prognozuoja vidutinę jų tarpusavio rungtynių rezultatų baigtį. Keener metodas remiasi tik komandų tarpusavio sužaistų rungtynių rezultatais, tačiau taip pat naudoja funkciją, kuri panaikina rezultatų manipuliavimo galimybę ir, kaip parodė tyrimo rezultatai, labai gerai atspindi komandų pajėgumą. Netikėtą rezultatą parodė *Hollingers Power Rankings* metodas, kuris sezono viduryje nerodė gerų rezultatų atspindint komandų pajėgumą, tačiau pasirodė esąs vienas geriausių nuspėjant komandų išsidėstymą sezono pabaigoje.

Labiausiai iš nagrinėtų reitingavimo modelių rezultatais išsiskyrė RPI, kuris buvo labiausiai atitolęs nuo agreguoto sąrašo, todėl bent jau šio tyrimo atveju rezultatai gauti šiuo metodu yra nepatikimi, o Keener metodu gautas reitingavimo sąrašas ir agreguotas sąrašas beveik identiški. Norint pateikti išsamesnes išvadas reiktų atlikti tolesnius tyrimus, pavyzdžiui, palyginti skirtingais metodais gautus reitingus keliose skirtingose sporto šakose, palyginimui naudoti kitas galimas priemones.

### Literatūra

1. Colley W. 2002: Colley's Bias Free College Football Ranking Method: The Colley Matrix Explained. *Colleyrankings* [interaktyvus]. [žiūrėta 2015-05-31]. Prieiga per internetą: <<http://www.colleyrankings.com/matrate.pdf>>.
2. How important is home court advantage in NBA? *Bleacherreport* [interaktyvus]. 2013 [žiūrėta 2015-05-31]. Prieiga per internetą: <<http://bleacherreport.com/articles/1520496-how-important-is-home-court-advantage-in-the-nba>>.
3. Kendall tau rank correlation coefficient. *Wikipedia* [interaktyvus]. [žiūrėta 2015-05-31]. Prieiga per internetą: <[http://en.wikipedia.org/wiki/Kendall\\_tau\\_rank\\_correlation\\_coefficient](http://en.wikipedia.org/wiki/Kendall_tau_rank_correlation_coefficient)>.
4. Massey K. 1997: Statistical Models Applied to the Rating of Sports Teams. *Masseyratings* [interaktyvus]. 1 [žiūrėta 2015-05-31]. Prieiga per internetą: <<http://www.masseyratings.com/theory/massey97.pdf>>.
5. NBA rezultatų statistika. *NBA* [interaktyvus]. 1997 [žiūrėta 2015-05-31]. Prieiga per internetą: <[http://www.nba.com/standings/team\\_record\\_comparison/conferenceNew\\_Std\\_Cnf.html?ls=iref:nba:gnav](http://www.nba.com/standings/team_record_comparison/conferenceNew_Std_Cnf.html?ls=iref:nba:gnav)>.
6. NBA 2014-2015 metų sezono rungtynių tvarkaraštis ir rezultatai. *Basketball-reference* [interaktyvus]. 2015 [žiūrėta 2015-05-31]. Prieiga per internetą: <[http://www.basketball-reference.com/leagues/NBA\\_2015\\_games.html?lid=header\\_seasons](http://www.basketball-reference.com/leagues/NBA_2015_games.html?lid=header_seasons)>.
7. Pomeroy K. Ratings madness. *Slate* [interaktyvus]. [žiūrėta 2015-05-31]. Prieiga per internetą: <[http://www.slate.com/articles/sports/sports\\_nut/2011/03/ratings\\_madness.html](http://www.slate.com/articles/sports/sports_nut/2011/03/ratings_madness.html)>.
8. Rossen K. The Biggest Flaw in John Hollinger's ESPN NBA Power Rankings. *Kevinrossen* [interaktyvus]. 2010 [žiūrėta 2015-05-31]. Prieiga per internetą: <<http://kevinrossen.com/the-biggest-flaw-in-john-hollingers-espn-nba-power-rankings/>>.
9. RPI metodas. *Bigbluhistory* [interaktyvus]. [žiūrėta 2015-05-31]. Prieiga per internetą: <<http://www.bigbluhistory.net/bb/rpi.html>>.
10. Sports rating system. *Wikipedia* [interaktyvus]. [žiūrėta 2015-05-31]. Prieiga per internetą: <[http://en.wikipedia.org/wiki/Sports\\_rating\\_system](http://en.wikipedia.org/wiki/Sports_rating_system)>.
11. What is the RPI? *Rpiratings* [interaktyvus]. [žiūrėta 2015-05-31]. Prieiga per internetą: <<http://rpiratings.com/WhatisRPI.php>>.
12. Webber W., Moffat A., Zobel J. 2010: A similarity measure for indefinite rankings, *ACM Transactions on Information Systems*, 28(4): 20.1 – 20.38.
13. Zabell L. S. 1989: The rule of succession. *Erkenntnis*, 31(2-3): p. 283-321.

### METHODS FOR RANKING SPORTS TEAMS

Eimantas Palionis, Viktoras Chadyšas

**Abstract.** The paper describes various methods that can be applied to the ranking of sports teams. Using the real data of the National Basketball Association (NBA) from the first half of the 2014–2015 season (from 2014-10-28 to 2015-02-12), teams are ranked using different ranking methods. All methods are compared with each other.

**Keywords:** sports teams ranks, ranking methods, ranking list.