NEM TELJES PÁROS ÖSSZEHASONLÍTÁSOK: A NŐI TENISZEZŐK VILÁGRANGSORÁNAK PÉLDÁJA[1]

TEMESI JÓZSEF – SZÁDOCZKI ZSOMBOR – BOZÓKI SÁNDOR
BCE – ELKH SZTAKI, BCE – ELKH SZTAKI, BCE

A páros összehasonlítás a döntési módszertanban gyakran alkalmazott eljárás rangsorok készítésére. Tanulmányunkban a nemzetközi női tenisz éljátékosainak összehasonlítására alkalmazzuk, a győzelem/vereség arányok felhasználásának segítségével. Ezekből az arányokból nem teljesen kitöltött páros összehasonlítás mátrixok képezhetők. A mátrixok azon elemei nem ismertek, ahol a két játékos nem mérkőzött egymással. Mivel az 1973 és 2020 közötti adatokat használjuk, ez az eset azért is bekövetkezhet, mert két játékos hivatásos tenisz karrierje különböző időszakokra esett. A módszertan megfelelő feltételek mellett alkalmas arra, hogy a mátrix alapadataival a logaritmikus legkisebb négyzetek módszerével egyfajta örökrangsort képezzünk. A cikk első részében kitérünk a tenisz rangsorok gyakorlati és elméleti kérdéseire, majd a felhasznált módszertan alapfogalmait és tételeit ismertetjük. A második részben az említett időszakban a hivatásos játékosok szervezeteinek hivatalos adataiból felépített adatrendszert mutatjuk be, majd azoknak a női éljátékosoknak a rangsorait elemezzük többféle bontásban, akik a megadott időszakban bármennyi ideig vezették a hivatalos ranglistát. Ugyanitt visszatérünk egy már megjelent tanulmányunkra a férfi teniszezőkről, az ott használt adatrendszert 2020-ig kibővítve. Eredményeink egybecsengenek a teniszszakértők által várt rangsorokkal, és azt mutatják, hogy mindkét adatrendszer elég robusztus ahhoz, hogy egyes változtatások (elemek elhagyása, beemelése, csoportokra bontás) keveset változtassanak a rangsorokon, illetve a rangsorok változásai jól magyarázhatók az adatok és a módszer jellemzőivel. A konkrét rangsor-számításokon túl elemzéseinket kiterjesztettük a nem teljes páros összehasonlítási mátrixok gráfreprezentációinak vizsgálatára, ezáltal mélyebb betekintést kapva a rangsorok hátteréről.

1 Bevezetés

Éppen 10 évvel ezelőtt készült el az a tanulmány, amelyik a legjobb férfi teniszezők rangsorkészítését nem teljesen kitöltött páros összehasonlítási mátrixok módszerével elemezte (Temesi, Csató, Bozóki 2012). A magyar nyelvű tanulmány példája a professzionális férfi teniszezők top-listájából kiválasztott 34 éljátékos rangsorolása volt, illetve egy 23 fős rangsor leszűkítve azokra, akik 1973 és 2012 között az ATP (a férfi teniszezők hivatalos szervezete) honlapjá-

[1] Beérkezett 2021. október 8. E-mail: `temesi@hungarnet.hu`, `bozoki.sandor@sztaki.hu`, `szadoczki.zsombor@sztaki.hu`.

nak adatai szerint a hivatalos ranglistát vezették. A 2016-ban megjelent angol nyelvű cikk a módszertani eredmények és a tenisz alkalmazás összefoglalásaként íródott (Bozóki, Csató, Temesi 2016). Az adatbázist ekkor kiegészítettük 2015-ig, s így 25 listavezetőt rangsoroltunk.

Miért volt ez érdekes módszertani szempontból? A rangsorképzés a legkülönbözőbb területeken jelent meg, különösen divatossá vált például a felsőoktatási intézmények, szakok kiválóságának bemutatására. Mivel a Corvinus Egyetem és az MTA SZTAKI kutatócsoportja több kutatási projektet is indított a páros összehasonlítások tulajdonságainak feltárására (ezek témáiból jó összefoglalást ad Temesi, 2017) utánanéztünk a szakirodalomban annak, hol merül fel természetes módon ez a modell rangsorok készítésénél. Ezt a lehetőséget nem a felsőoktatásban, hanem a sportban találtuk meg.

A bajnokot az egyéni sportok (néhányat megemlítve: sakk, vívás, tenisz, ökölvívás) vagy a csapatsportok (például kosárlabda, futball, kézilabda) egy részében az egymás közötti mérkőzések eredményei alapján hirdetik ki. Az egyes sportágak igen változatos, a hagyományokon alapuló pontszámítást vagy lebonyolítási rendszert követnek. Nem arról van szó azonban, hogy valamely aktuális bajnokság helyezettjeit a meglévőnél ,,igazságosabb" módon állapítsuk meg. A páronkénti eredményeket egyfajta erősorrend kialakítására kívánjuk felhasználni. Örök kérdés, hogy egy sportágban ki a legjobb egy bizonyos időszakban, vagy akár a sportág megjelenése óta. Azokban a sportágakban, ahol a statisztikák (szinte) hiánytalanul rendelkezésre állnak (pl. az USA baseball vagy kosárlabda mérkőzései), akár adatokkal is meg lehet támogatni a vitát. A teniszben sem azt a kérdést tesszük fel, hogy az éppen aktív játékosok ranglistája hogyan építendő fel. Ehelyett azt szeretnénk az egymás elleni eredmények tükrében megmondani, hogy a mai és régebbi teniszezőket összevetve ki a legjobb. Ám hogyan lehet áthidalni azt a nehézséget, hogy a nagy teniszezők egy része soha nem találkozott egymással a pályán? Hogyan hasonlítható össze Borg és Nadal, vagy a nőknél Navratilova és Hingis? Beleütközünk a nem teljes páros összehasonlítások problémájába.

Az angol nyelvű cikk idézettsége azt mutatja, hogy sokakat érdekelt a teniszrangsorokra történt alkalmazás, valamint általában a sport alkalmazások is igen népszerűek (Csató, 2021). Mivel továbbra is több irányból foglalkoztunk a nem teljesen kitöltött páros összehasonlítás mátrixokkal, ezért úgy gondoltuk, érdemes visszatérni a témára. Annál is inkább, mert a férfi teniszezők előtérbe helyezését semmi nem indokolta, a női teniszezők ugyanolyan érdeklődésre tarthatnak számot. Nekiláttunk tehát, hogy a WTA (a női teniszezők hivatalos szervezete) adatainak felhasználásával a hölgyjátékosok ,,örökrangsorát" is megalkossuk.

Természetesen nem arról van szó, hogy egyszerűen megismételjük mindazt, amit az ATP versenyzők esetében már elvégeztünk. Erősen támaszkodunk ugyan az említett tanulmányra és cikkre, s ahol szükséges, ott a hivatkozást formálisan is megtesszük, ám arra törekszünk, hogy a 2012-es számításokból csak a releváns eredmények mentén haladjunk, azokat nem megismételve. Találtunk viszont újabb elemzés tárgyává tehető kérdéseket, melyeket részletesebben mutatunk be. A férfi teniszezőkre is visszatérünk néhány számítás erejéig, 2020-ig kibővítve adatbázisunkat.[2]

A sporttal foglalkozó cikkek, szakkönyvek száma nagy és egyre gyarapodik. Ezen belül a sportbajnokságokkal, azok lebonyolításával és az eredmények sorba rendezésével (ki a bajnok, kik a helyezettek, kik jutnak tovább egyes selejtezőkből) kapcsolatos elemzések kiemelten érdekesek számunkra. Az általánosabb megközelítések közül érdemes megemlíteni Csató és Tóth (2020) cikkét, amelynek szemlélete eltér a szokásos súlyozott tényezős szemlélettől és ötleteket adhat a sportbeli kiválóság mérésének kezelésére is. Csató (2017) sakkcsapatok rangsorolására viszont már felhasználja a nem teljes páros összehasonlításokat.

Az egyéni sportolók és a sportcsapatok egymás elleni eredményein alapuló rangsorokkal, illetve a sportbeli kiválóság mérésével foglalkozó cikkek általában vagy a játékosok (csapatok) teljesítményének statisztikai elemzéseken alapuló összehasonlításával foglalkoznak, vagy – legtöbbször éppen ezeket az eredményeket felhasználva – előrejelzéseket próbálnak adni eljövendő sportesemények eredményeire. Kovalchik (2016) három kategóriába sorolja az előrejelző modellekkel foglalkozó szakirodalmat. A regressziós modellek – leginkább logit vagy probit modellek – egy-egy mérkőzés győztesét igyekeznek megmondani. Jó példa erre Lisi és Zanella (2017) cikke, akik egy logisztikus regressziós modellel becslik a mérkőzések győzelmi valószínűségét. A példa a 2013-as négy Grand Slam férfi bajnoksága, a változók között az ATP pontok és helyezések, a játékosok életkora, a helyszín és a fogadóirodák oddsai szerepelnek. A pontbecslések a mérkőzés győztesét egy-egy nyertes pont (kiemelten a szerva) valószínűségeiből rakják össze. A páros összehasonlítások Bradley-Terry típusú modelljeiben (Bradley, Terry 1952) a játékosok képességei alapján számolnak.

Kovalchik (2016) összehasonlító elemzésének egyik érdekes eredménye, hogy az Élő-modellt (Elo, 1978) felhasználó módszerek a legversenyképesebbek az előrejelzési „versenyben”. Különösen jól szerepelnek az éljátékosok mérkőzéseinek előrejelzésében, míg a hátrébb rangsoroltaknál ez nem működik olyan jól. Ezt megerősítik Vaughan-Williams et al. (2019) számításai a 2018-as wimbledoni füves teniszbajnokság női eredményeit illetően. Főleg bizonyos típusú sportfogadások esetében lehet érdekes a Kovalchik és Reid (2018) cikkében leírt dinamikus módszer a teniszmérkőzések győztesének menet közbeni előrejelzésére vonatkozóan.

A felhasznált módszerek változatosak. Zhou et al. (2020) tanulmánya adott számú csapat egymás elleni eredményeit értékeli egy irányított élekkel ellátott hálózatban. Három rangsoroló módszert hasonlítottak össze: a győzelmi arányt (W/L win/loss hányados), a PageRank értékeket és a PageRank általuk javasolt módosítását, ahol megkülönböztetik a jó és a rossz ellenfelek elleni győzelmeket. Számunkra is érdekes következtetésük valós kosárlabda, jégkorong, baseball és futball eredményekre az, hogy azokban az esetekben, ha egy bajnokságban nem mindenki játszik mindenkivel – ilyen a tenisz is –,

[2] Az élet úgy hozta, hogy mind a férfi, mind a női esetben 2020 eleje egy természetes szakaszhatár lett, mivel ezután a Covid-19 járvány miatt bő egy éven keresztül gyakorlatilag szüneteltek, illetve igencsak foghíjasak voltak a nagy tornák.

akkor a PageRank modellnek vannak előnyei a többivel szemben.

Ramón et al. (2012) a férfi teniszezők rangsorából kiválasztottak 53 játékost, akik legalább 40 mérkőzést játszottak egy adott évben. Egy olyan outputorientált DEA (Data Envelopment Analysis) modellt írtak fel, ahol a következő százalékos értékeket használták outputként a DEA hatékonysági elemzésnél: az első adogatás beütése; az első és a második szerva utáni pont megnyerése; a saját és az ellenfél adogatójátékainak megnyerése; a break labdák megnyerése, illetve hárítása; az ellenfél első és második szervájának visszaadása utáni pontok megnyerése. A nominális input minden játékos esetében 1 volt. A DEA modell súlyait (különböző távolságmértékeket megvizsgálva) a játékosok rangsorolására használták.

Gu és Saaty (2019) leíró mutatókkal (pl. kor, bal- vagy jobbkezesség, rangsor-helyezés) és teljesítmény-mutatókkal (pl. ászok száma, megnyert szervajátékok, break pointok) dolgozik. Mintegy 80 ezer férfi és 35 ezer női mérkőzés hivatalos adatait dolgozták fel. Először egy logisztikus modellel vizsgálták a 44 mutató jelentőségét a mérkőzés kimenetelére vonatkozóan. Ezután az Analytic Network Process (ANP) modellt alkalmazták, ahol a kulcsmutatók faktoranalízissel kapott hét csoportja volt az elemzés alapja. A modellt a US Open 2015-ös mérkőzésein tesztelve 85%-os egyezést kaptak, ami jobb a szokásos 70% körüli előrejelzési eredményeknél.

Közelebb esik rangsorolási módszertanunkhoz Chao et al. (2018) cikke, akik a nem teljes PCM (a továbbiakban az angol pairwise comparison matrix elnevezést használva rövidítésre) megközelítést alkalmazták fuzzy preferencia reláció felhasználásával Go játékosok rangsorolására, és hozzánk hasonlóan a győzelem/vereség arányokkal dolgoztak. A fuzzy megközelítés segített az adatok korrekciójában.

Baker és McHale (2014) sem értettek egyet azzal a sportújságírók által kedvelt módszerrel, hogy minden idők legjobbját a teniszben egyszerűen a megnyert Grand Slam versenyek számával határozzák meg. Elmélyedtek az 1968 és 2012 között rendezett Grand Slam mérkőzések részleteiben (több mint 20 ezer mérkőzés adatait felhasználva), és a páros összehasonlítások valószínűségi modelljeinek segítségével dinamikus módon becsülték az egyes versenyzők játékerejét, majd a játékosokat a karrierjük során elért legnagyobb játékerő értékek alapján rangsorolták. A vizsgált 1163 játékos közül az első öt: Federer, Borg, Connors, Laver, Nadal. Bár Temesi, Csató, Bozóki (2012) tanulmányunk is 1973 és 2012 között dolgozta fel az ATP adatokat, egészen más játékoskört és eredményeket vett figyelembe: míg Bakeréknél csak a négy nagy verseny minden mérkőzése, addig nálunk a 23 listavezető egymás közötti eredményei számítottak. És mégis, érdekes a hasonlóság, hiszen Federer, Nadal, Sampras, Borg, Lendl a 23 listavezető közötti sorrendünk – az ötből három név azonos. Megcsinálták a számításokat a női játékosokra is (Baker, McHale 2017): az időszak 1968–2016, a négy nagy bajnokság 1123 játékosa és közel 22 ezer mérkőzése. Az első öt: Graf, Serena Williams, Evert, Navratilova és Court. Itt is látni fogjuk a hasonlóságot jelenlegi cikkünk rangsoraival.

Orbán-Mihálykó et al. (2019) a WTA Head to Head eredményeket használta fel arra, hogy a női játékosok között rangsort állítson fel, a Thurstone modellt alkalmazva, a paraméterek maximum likelihood becslésével. A valószínűségi modellel végzett számítások a következő sorrendet generálták: 1. Navratilova, 2-3. Graf és Serena Williams, 4. Evert.

Cikkünk első része röviden összefoglalja az alkalmazási hátteret, illetve a felhasznált modellezési eszközöket. A második rész a WTA-rangsorokról szól, kitérve azokra a nem teljes páros összehasonlítási mátrix tulajdonságokra, amelyeket az alkalmazás kapcsán vizsgáltunk. Végezetül visszatérünk az ATP-rangsorokra, majd az elemzések fontosabb következtetéseit foglaljuk össze, illetve kitekintünk a nyitott kérdésekre.

2 Az adatbázis

A professzionális tenisz alapjai 1973 óta keveset változtak. Ebben az évben alapították a hivatásos férfi játékosok (ATP) és a hivatásos női játékosok (WTA) szervezetét. Gondosan felépített versenyrendszer biztosítja, hogy a különböző típusú és erősségű tornákon kiválasztódjanak a legeredményesebb játékosok. Ezeken a versenyeken (gyakorlatilag a pénzdíjakkal arányos) pontrendszer révén lehet előrehaladni és a legjobbak számára kiírt tornákra bekerülni. A játékosoknak előre meghatározott szabályok szerint, megfelelő számú versenyen kell részt venniük. Ha ,,magasabb osztályba" akarnak lépni, akkor pontszámuknak elegendőnek kell lennie ahhoz, hogy a selejtezőbe vagy egyenesen a főtáblára kerüljenek. Az első 2-300 legmagasabban rangsorolt teniszezőnek van csak esélye arra, hogy a legnagyobb pénzdíjú versenyek (általában 32 vagy 64 létszámú) főtábláján kieséses rendszerben minél előbbre jusson. A legrangosabb versenyek a Grand Slam tornák: ez a négy bajnokság kiemelkedik a többi közül pénzdíjai, nézettsége és a szerezhető pontszámok tekintetében is. Itt 128-as főtábla van, a nevezéseket a ranglista szerint fogadják – bár kivételesen mindig fenntartanak néhány helyet feltörekvő tehetségeknek vagy sérülések miatt kihagyó éljátékosoknak: ezekről a tornák szervezőbizottsága dönt.

Az ATP és a WTA számontartja mindegyik hivatalos verseny eredményeit és ezek egy adatbázisban kereshetők a tornák és egyes játékosok neve szerint is[3]. A tájékozódást megkönnyíti, hogy létezik egy a honlapon elérhető H2H (head to head) statisztika, ahol két megnevezett játékos egymás elleni összes eredménye megtekinthető. Ugyancsak a honlapon látható az aktuális pontszámok alapján összeállított ranglista[4]. A már említett nevezés-elfogadási rendszer mellett a ranglista fontos funkciója a kiemelési sorrend. Elkerülendő azt, hogy a tornára benevezett legmagasabb rangú játékosok azonnal a verseny elején összekerülve az egyenes kieséses rendszerben kiejtsék egymást, illetve biztosítandó, hogy a gyengébben rangsoroltak csak a magasabban rangsoroltak elleni győzelmekkel bizonyítva haladhassanak előre

[3] https://www.atptour.com, illetve https://www.wtatennis.com

[4] https://www.atptour.com/en/rankings/singles, illetve https://www.wtatennis.com /rankings/singles

(ezzel a verseny nézettségét is növelve), a játékosok beosztása a táblán előre meghatározott. Egy 128-as táblán az első és második kiemelt az 1. és 128. helyre kerül, a harmadik és negyedik a 64. és 65. helyre, és így tovább, mind a 32 kiemelt helyet betöltve. A többieket véletlenszerűen hozzájuk sorsolják.

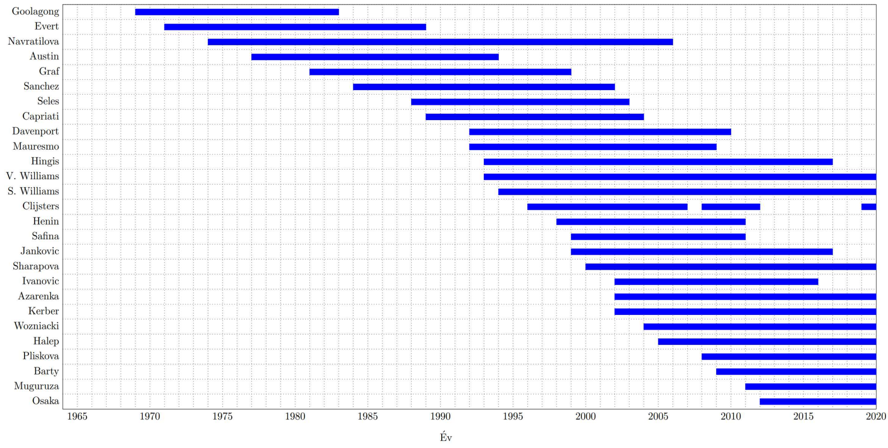
Minden versenyen előre közlik a helyezettek számára kiosztandó ranglista pontokat. A versenyzőkre különböző előírások vonatkoznak versenynaptáruk elkészítésekor: hány versenyen kell minimálisan részt venniük, hogyan szerezhetik meg a pontokat egy már az előző évben is látogatott versenyen és így tovább. Nem megyünk azonban bele a részletekbe, a lényeg az, hogy ennek a speciális pontrendszernek és az így kialakuló rangsornak természetesen nagy szerepe van a modern tenisz versenyrendszerében, és bizonyos értelemben tükröz is egyfajta erősorrendet. Azonban az mindig vita tárgya marad, hogy ki a ,,legjobb" teniszező, vagy az egyik teniszező ,,jobb-e" a másiknál. Akár azt is mondhatjuk persze, hogy a ranglista alapján ezeket a kérdéseket megválaszoltnak tekintjük: a legjobb az aktuális WTA (ATP) ranglistavezető, két játékos esetében pedig az a jobb, aki előrébb áll a rangsorban. A tenisz szerelmesei azonban mindig vitatják ezeket a rangsor szerint adott válaszokat, és van is igazság abban, hogy az üzleti szempontok alapján (is) kreált rangsor nem feltétlenül azt tükrözi, amit egy teniszszakértő megfelelőnek gondol. A szakértők például nem szeretik azt, hogy a rangsor egyformán kezeli a különböző borításokon lejátszott mérkőzéseket. Jobb lenne salakon, kemény pályán, füvön külön rangsort hirdetni – mondják.

Nem akarunk a hasonló vitákban igazságot tenni. Nem szeretnénk a hivatalos szervezetek ranglistáit egy másik, általunk jobbnak tartott rangsorral felváltani. Bármit is javasolnánk, lennének ádáz kritikusaink – valószínűleg joggal. Ezért egy ártatlanabbnak tűnő kérdést tettünk fel. Mi lenne akkor, ha két versenyző rangsorban elfoglalt helyét az egymás elleni eredmények és a más versenyzők ellen elért eredmények kölcsönösen határoznák meg? Ez a megközelítés igazságosnak tűnik, de bonyodalmakat is vet fel. Indítsunk azzal, hogy az A játékos ,,jobb" a B játékosnál, ha többször nyert ellene, mint ahányszor veszített (milyen jó, hogy a teniszben nincs döntetlen!). Egyformán jók, ha azonos a nyertes és vesztes meccsek száma. Készítsünk egy olyan mátrixot, ahol az egymás elleni W/L hányadosok vannak, és tekintsük ezeket az arányokat a ,,jobb" kifejezés mérőszámának, ha nagyobbak 1-nél (később az esetleges 0-val való osztás problémáját kezelni fogjuk). A ,,nem jobb" mérőszáma legyen ennek a számnak a reciproka. Ezáltal az egymás elleni eredményekből egy páros összehasonlítási mátrixot kapunk.

Kitöltve a mátrixot több problémánk is akadhat. Ha bevonjuk a C játékost, akkor a számok alapján kiderülhet, hogy A és B W/L hányadosa > 1, B és C W/L hányadosa is > 1, s ugyanakkor A és C W/L hányadosa < 1. Mint ahogyan páronként egyetlen mérkőzést játszva is előfordulhat, hogy A megveri B-t, B megveri C-t, de A nem veri meg C-t (ahogyan azt az első két mérkőzés alapján várnánk), hanem kikap tőle, ugyanígy a W/L hányadosok szerint is előfordulhat ,,körbeverés". A másik szembetűnő gond az lehet, hogy már aránylag kevés számú játékos eredményeit feltüntetve láthatóvá válik, hogy bizonyos párok soha nem játszottak egymással, pályafutásuk során soha nem futottak össze, azaz nincs W/L hányadosuk. Mivel mi azt az ambiciózus célt tűztük ki magunk elé, hogy 1973 és 2020 közötti topjátékosokat hasonlítsunk össze és próbáljunk meg rangsorolni, szinte biztos, hogy mind a két probléma előkerül.

De kik legyenek a rangsorolandó teniszezők? A férfi teniszezők esetében végül a szubjektív válogatást feladva azt a szelekciós elvet alkalmaztuk (Temesi, Bozóki, Csató 2012), hogy az ATP rangsorok első helyezettjeinek eredményeiből építettük fel adatmátrixunkat. Természetesen egyéb lehetőségeink is lettek volna: Grand Slam győztes játékosok összehasonlítása, ATP rangsorok legjobb 5-ben lévő játékosai, minden nemzet legjobbjai, pályaborítások szerinti rangsorok, és így tovább. Alkalmazási szempontból végül a fenti, jól értelmezhető változat mellett döntöttünk, aminek az az előnye is megvolt, hogy az ebben szereplő nevek még a tenisz iránt kevésbé érdeklődőknek is mondanak valamit.

A hölgyjátékosok esetében sem tértünk el ettől az elvtől. Összegyűjtöttük a WTA rangsoraiból azokat a teniszezőket, akik bármilyen időszakaszon világelsők voltak[5]. 27 ilyen hölgyet találtunk. Nevük[6] és profi pályafutásuk hossza az *1. ábrán* látható. Az ábrából leolvasható például, hogy a leghosszabb aktív pályafutást Navratilova és a Williams testvérek mondhatják a magukénak, bár többen is vannak még, akik a 20 évhez közelítenek. Clijsters többször is visszavonult, majd újrakezdte. 11-en vannak, akik ma is aktív versenyzők, köztük a több mint 25 éve pályán és az élvonalban lévő Williams testvérekkel.



1. ábra. WTA ranglistavezetők hivatásos pályafutásának hossza

A WTA honlapjának H2H rovatából összeállítottuk ezeknek a teniszezőknek az egymás elleni eredményeit, vagyis a W/L hányadosokat. Ez a mátrix az *1. táblázatban* található.

[5]https://en.wikipedia.org/wiki/List_of_WTA_number_1_ranked_tennis_players

[6]A teniszezők nevének írásakor azoknak a WTA, illetve az ATP honlapokon szereplő formáját használjuk.

	Goolagong	Evert	Navratilova	Austin	Graf	Sanchez	Seles	Capriati	Davenport
Goolagong		12/26	12/15	4/4					
Evert	26/12		37/43	9/8	6/7	1/1	2/1		
Navratilova	15/12	43/37		20/13	9/9	12/3	7/10	1/1	1/0
Austin	4/4	8/9	13/20		1/1				
Graf		7/6	9/9	1/1		28/8	10/5	10/1	8/6
Sanchez		1/1	3/12		8/28		3/20	6/4	7/5
Seles		1/2	10/7		5/10	20/3		9/5	3/10
Capriati			1/1		1/10	4/6	5/9		3/9
Davenport			0/1		6/8	5/7	10/3	9/3	
Mauresmo					0/1	2/1	2/3	7/4	4/12
Hingis					2/7	18/2	15/5	5/4	11/14
V.Williams					2/3	6/3	9/1	4/0	13/14
S.Williams					1/1	3/4	4/1	10/7	10/4
Clijsters					0/1	4/0	0/1	3/3	9/8
Henin						0/1	3/4	5/2	7/5
Safina									1/2
Jankovic									2/4
Sharapova							0/1	0/1	5/1
Ivanovic									0/1
Azarenka									
Kerber									0/1
Wozniacki									0/1
Halep									
Pliskova									
Barty									
Muguruza									
Osaka									

1. táblázat. A WTA ranglistavezetők egymás elleni eredményei, a W/L hányadosok

	Mauresmo	Hingis	V.Williams	S.Williams	Clijsters	Henin	Safina	Jankovic	Sharapova
Goolagong									
Evert									
Navratilova									
Austin									
Graf	1/0	7/2	3/2	1/1	1/0				
Sanchez	1/2	2/18	3/6	4/3	0/4	1/0			
Seles	3/2	5/15	1/9	1/4	1/0	4/3			1/0
Capriati	4/7	4/5	0/4	7/10	3/3	2/5			1/0
Davenport	12/4	14/11	14/13	4/10	8/9	5/7	2/1	4/2	1/5
Mauresmo		7/8	3/5	2/10	7/8	6/8	4/3	6/1	3/1
Hingis	8/7		11/10	6/7	4/5	2/2	2/1	0/2	1/2
V.Williams	5/3	10/11		12/19	6/7	7/2	3/1	7/7	3/5
S.Williams	10/2	7/6	19/12		7/2	8/6	6/1	10/4	20/2
Clijsters	8/7	5/4	7/6	2/7		13/12	8/2	8/1	5/4
Henin	8/6	2/2	2/7	6/8	12/13		5/1	10/0	7/3
Safina	3/4	1/2	1/3	1/6	2/8	1/5		3/4	3/4
Jankovic	1/6	2/0	7/7	4/10	1/8	0/10	4/3		1/8
Sharapova	1/3	2/1	5/3	2/20	4/5	3/7	4/3	8/1	
Ivanovic	2/6	1/1	3/9	1/9	0/6	0/5	3/1	9/3	4/10
Azarenka	2/0	1/0	3/6	5/18	3/4		2/4	7/4	7/8
Kerber			6/3	3/6	1/0	0/2		4/2	5/4
Wozniacki	1/0	0/2	1/7	1/10	0/3	0/1	0/1	6/5	4/7
Halep			4/3	2/9	0/1		0/1	7/1	2/7
Pliskova			2/1	2/2				1/1	0/2
Barty			2/0	0/2					2/1
Muguruza			2/4	3/3	1/0			3/2	1/3
Osaka			1/1	2/1					1/0

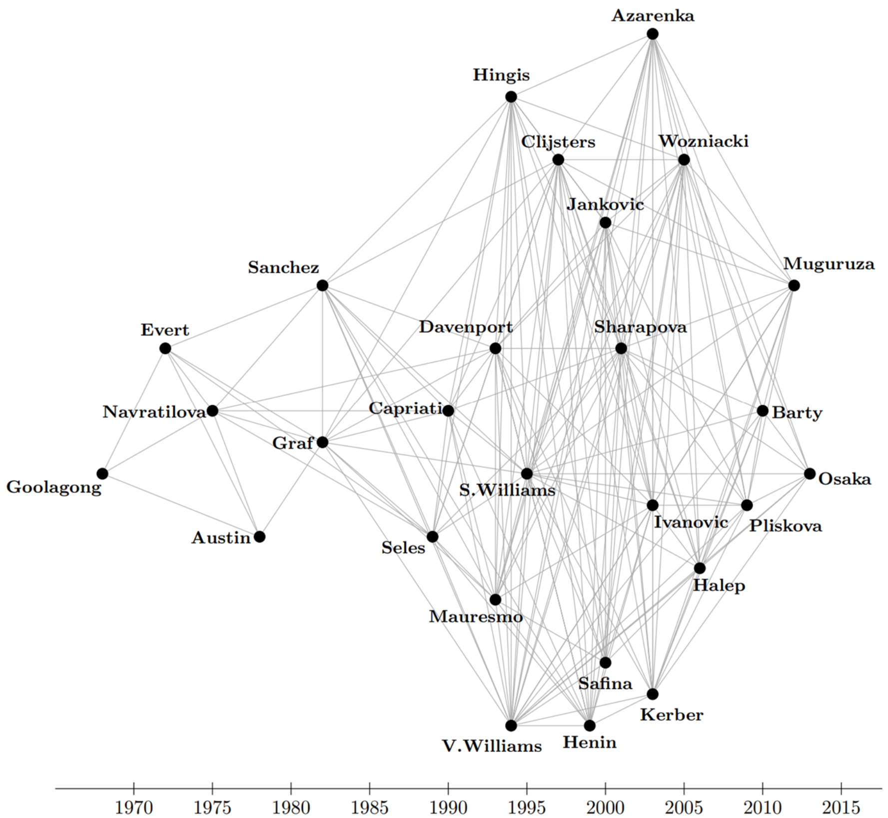
1. táblázat. A WTA ranglistavezetők egymás elleni eredményei, a W/L hányadosok *(folyt.)*

	Ivanovic	Azarenka	Kerber	Wozniacki	Halep	Pliskova	Barty	Muguruza	Osaka
Goolagong									
Evert									
Navratilova									
Austin									
Graf									
Sanchez									
Seles									
Capriati									
Davenport	1/0		1/0	1/0					
Mauresmo	6/2	0/2		0/1					
Hingis	1/1	0/1		2/0					
V.Williams	9/3	6/3	3/6	7/1	3/4	1/2	0/2	4/2	1/1
S.Williams	9/1	18/5	6/3	10/1	9/2	2/2	2/0	3/3	1/2
Clijsters	6/0	4/3	0/1	3/0	1/0			0/1	
Henin	5/0		2/0	1/0					
Safina	1/3	4/2		1/0	1/0				
Jankovic	3/9	4/7	2/4	5/6	1/7	1/1		2/3	
Sharapova	10/4	8/7	4/5	7/4	7/2	2/0	1/2	3/1	0/1
Ivanovic		3/5	5/2	5/2	3/2	0/5		0/1	
Azarenka	5/3		8/1	7/4	2/3	4/4	1/1	2/2	1/3
Kerber	2/5	1/8		8/7	5/6	7/5	2/2	3/5	4/1
Wozniacki	2/5	4/7	7/8		5/2	6/4	3/0	3/3	2/1
Halep	2/3	3/2	6/5	2/5		8/4	3/1	3/4	4/1
Pliskova	5/0	4/4	5/7	4/6	4/8		2/4	8/2	3/2
Barty		1/1	2/2	0/3	1/3	4/2		2/1	2/2
Muguruza	1/0	2/2	5/3	3/3	4/3	2/8	1/2		
Osaka		3/1	1/4	1/2	1/4	2/3	2/2		

1. táblázat. A WTA ranglistavezetők egymás elleni eredményei, a W/L hányadosok *(folyt.)*

A mátrix nyilvánvalóan nem teljes, hiszen akinek a pályafutási szakasza az 1. ábrán egyetlen pontban sem közös egy másik játékoséval, annak a két teniszezőnek eleve nem lehet W/L hányadosa. Ha jobban megnézzük az elemeket, könnyen találunk példát a ,,körbeverésre" is, például az Henin – Davenport – Venus Williams trió esetében. Az 1. ábra és az 1. táblázat információinak megjelenítését szolgálja a *2. ábra* gráfja.

Itt most inkább a szemléltetés a célja, az elemzéseknél viszont vissza fogunk térni ennek a gráfnak a tulajdonságaira. Egyelőre annyit érdemes az ábráról megjegyezni, hogy láthatóan elkülönülnek az időtengely elején és az időtengely végén található játékosok, és találunk ,,összekötő" egyéniségeket, mint például Sharapova vagy a Williams-testvérek. Navratilováról azt gondolnánk, hogy ő is ebbe a csoportba fog tartozni hosszú pályafutása okán, de ő éppen annak a példája, hogy a karrierutak hossza némileg csalóka, mert az egyesre vonatkozó pályafutás összekapcsolódik a hivatásos páros szereplésekkel – ez Navratilovánál jóval hosszabb és több versenyt jelent, mint az egyes szereplés. Navratilova 8 játékossal találkozott a másik 26-ból, Sharapova 20 másik versenyzővel játszott legalább egy mérkőzést. A gráfban ez a csúcsok fokszámának felel meg. Ezek a fokszámok 3 és 22 között mozognak. Ránézésre nem könnyű ellenőrizni, de az is igaz, hogy a 2. ábra gráfja összefüggő, tenisznyelven tehát nem találunk olyan egy vagy több főből álló csoportot, akik ne játszottak volna legalább egy játékossal a csoporton kívüliek közül.



2. ábra. A WTA ranglistát vezető női teniszezők egymás elleni mérkőzéseinek gráfja

3 Páros összehasonlítások tulajdonságai

Mielőtt továbbmennénk, röviden összefoglalunk néhány definíciót és tételt. Ezekre a leírás és az elemzés során hivatkozni fogunk. Az itt tárgyalt fogalmakat sokszor nem a konkrét alkalmazás szempontjából, hanem attól általánosabban fogalmazzuk meg, azonban jelöléseinket a későbbi tárgyaláshoz igazítjuk.

Jelölje $P_1, P_2, \ldots, P_n$ az általunk vizsgált alternatívákat (szempontokat, szavazóerőket vagy esetünkben játékosokat)!

1. Definíció. *Egy $n \times n$-es $\mathbf{P} = [p_{ij}]$ mátrixot páros összehasonlítás mátrixnak (PCM) nevezünk, ha teljesíti az alábbi tulajdonságokat:*

(i) $p_{ij} > 0$, $i, j = 1, 2, \ldots, n$ *(pozitivitás)*

(ii) $p_{ji} = 1/p_{ij}$, $i, j = 1, 2, \ldots, n$ *(reciprocitás), így* $p_{ii} = 1$, $i = 1, 2, \ldots, n$ *(homogenitás).*

A mátrix általános eleme, p_{ij} azt mutatja meg, hogy hányszor jobb/nagyobb/erősebb/fontosabb a P_i alternatíva, mint a P_j alternatíva. Gyakorlati szempontból rendkívül fontos kérdés a mátrix inkonzisztenciája (ellentmondásossága).

2. Definíció. *Egy PCM-et konzisztensnek nevezünk, ha minden alternatíva-hármasra (triádra) teljesül, hogy*

(iii) $p_{ik} = p_{ij} \cdot p_{jk}, \;\; i, j, k = 1, 2, \ldots, n$ *(kardinális tranzitivitás).*

Ha egyetlen triád esetén is sérül az egyenlőség, akkor a PCM inkonzisztens.

Ezen felül egy triádot ordinálisan intranzitívnak nevezünk, ha a megfelelő mátrixelemek által meghatározott sorrend is ellentmondásos. Azaz példaképp $p_{ik} > 1$, $p_{jk} < 1$ és $p_{ij} < 1$, vagyis P_i alternatíva jobb, mint P_k alternatíva, míg P_k jobb, mint P_j, azonban P_j jobb, mint P_i. A mi vizsgálatainkban az ordinálisan intranzitív triádok rendkívül jól interpretálhatóak a játékos-hármasok egymás közti körbeverése segítségével. Az inkonzisztencia a különböző PCM-ek esetén természetesen eltérő erősségű lehet, melynek mérése hatalmas szakirodalommal rendelkezik (Brunelli 2018). Ennek ellenére a mai napig a Saaty (1977) által javasolt CR (Consistency Ratio) inkonzisztencia index maradt a legnépszerűbb a gyakorlati alkalmazások terén, melyre a $CR < 0{,}1$ elfogadhatósági szabály a legelterjedtebb.

3. Definíció. *A* $\mathbf{P}$ $n \times n$*-es PCM* CR *inkonzisztencia indexe:*

$$CR = \frac{CI}{RI}$$

ahol CI *(Consistency Index) az alábbi módon számolható:*

$$CI = \frac{\lambda_{\max} - n}{n - 1}$$

ahol $\lambda_{\max}$ *a* $\mathbf{P}$ *mátrix domináns sajátértéke, míg RI (Random Index) egy elegendően nagy, véletlenszerűen generált* $n \times n$*-es PCM-ekből álló mintán számított átlagos* CI*.*

Egy PCM-ből különböző módszerek segítségével egy súlyvektort tudunk kiszámítani, amely megadja az alternatívák jóságát (erősségét, fontosságát). A két leggyakrabban alkalmazott súlyszámítási technika a sajátvektor módszer (SV) (Saaty 1977) és a logaritmikus legkisebb négyzetek módszere (LLNM) (Crawford, Williams 1985), melyek az alábbi formulákkal adottak:

a) SV: $$\mathbf{P}\mathbf{w} = \lambda_{\max} \cdot \mathbf{w},$$

b) LLNM: $$\sum_{i=1}^{n} \sum_{j=1}^{n} \left(\ln p_{ij} - \ln \frac{w_i}{w_j} \right)^2 \underset{\mathbf{w}}{\to} \min ,$$

ahol $\mathbf{w}$ jelöli a kiszámolt súlyvektort, melynek elemei w_i ($i = 1, \ldots, n$), $\lambda_{\max}$ pedig, mint korábban, a $\mathbf{P}$ mátrix domináns sajátértéke. Ha egy PCM konzisztens, akkor semmilyen ellentmondást nem tartalmaz, így elemei pontosan fölírhatóak $p_{ij} = w_i / w_j$ alakban, ezáltal bármely súlyszámítási technikát alkalmazzuk is, mindegyik ugyanazon súlyvektorhoz vezet. Emiatt gyakori, hogy egy inkonzisztens mátrixot egy konzisztenssel közelítünk különböző módszerek mentén (Anholcer, Fülöp 2019).

Előfordulhat azonban, hogy a PCM néhány eleme hiányzik. Ennek hátterében számos ok állhat: lehetséges, hogy az összehasonlítások egyszerűen nem elvégezhetőek (mint az általunk vizsgált példákban), adatvesztés történt, vagy esetleg a döntéshozónak nincs elég ideje az összes elem meghatározásához. Ekkor nem teljesen kitöltött páros összehasonlítás mátrixokról beszélünk. A korábban említett súlyszámítási technikák könnyen általánosíthatók a nem teljesen kitöltött esetre is. A sajátvektor módszer alkalmazása a minimális CR értéket eredményező kitöltésen alapszik (Shiraishi, Obata, Daigo 1998), míg az LLNM esetében az optimalizálási feladatot csak az ismert elemek alapján végezzük el.

A nem teljesen kitöltött eset vizsgálatát nagyban megkönnyíti a gráfreprezentációra való áttérés.

4. Definíció. *A $G = (V, E)$ irányítatlan gráfot, ahol V a gráf csúcsainak, E pedig az éleinek halmaza, a* $\mathbf{P}$ *nem teljesen kitöltött PCM reprezentáló gráfjának nevezzük, ha V megfeleltethető a* $\mathbf{P}$ *mátrix alternatíváinak, és egy adott él pontosan akkor eleme E-nek, ha a* $\mathbf{P}$ *mátrix megfelelő eleme ismert.*

A reprezentáló gráf néhány tulajdonságának segítségével számos páros összehasonlításokkal kapcsolatos probléma és eredmény fogalmazható meg.

1. Tétel. (Bozóki, Fülöp, Rónyai 2010). *Az SV és LLNM nem teljesen kitöltött esetre általánosított súlyszámítási módszereknek pontosan akkor egyértelmű a megoldása, ha a PCM-et reprezentáló gráf összefüggő.*

Egy gráfot összefüggőnek nevezünk, ha bármely két csúcsa között vezet út. Amennyiben két elem között ez nem teljesül, akkor a súlyaik (erőviszonyuk) kapcsolatát sem tudjuk egyértelműen megállapítani. Azonban az összefüggőség szigorúbb változatait is érdemes lehet megvizsgálni a feladatunk kapcsán: ezekre ki is térünk majd elemzésünkben.

5. Definíció. *a) A $G = (V, E)$ gráfot k-szorosan élösszefüggőnek nevezzük, ha bárhogyan távolítunk el kevesebb, mint k darab élt a gráfból, az összefüggő marad, azaz bármely $G' = (V, E \backslash H)$ gráf összefüggő, ahol $H \subset E$ és $|H| < k$. A G gráf élösszefüggősége a maximális k, melyre G k-szorosan élösszefüggő.*

b) A $G = (V, E)$ gráfot k-szorosan csúcsösszefüggőnek nevezzük, ha bárhogyan távolítunk el kevesebb, mint k darab csúcsot a gráfból, az összefüggő marad, azaz bármely $G' = (V \setminus L, E)$ gráf összefüggő, ahol $L \subset V$ és $|L| < k$. A G gráf csúcsösszefüggősége a maximális k, melyre G k-szorosan csúcsösszefüggő.

Szintén kérdéses lehet két elem viszonyának megbízhatósága a súlyokat tekintve, amennyiben közöttük csak egy hosszú, sok összehasonlítást tartalmazó indirekt út vezet. Ennek a problémának egy természetes mérőszáma a gráf átmérője.

6. Definíció. *Egy $G = (V, E)$ összefüggő gráf átmérője (d) a leghosszabb legrövidebb út a gráf bármely két csúcsa között:*

$$d = \max_{u,v \in V} l(u, v)$$

ahol $l(\cdot, \cdot)$ a gráftávolságot, vagyis a legrövidebb utat jelöli két csúcs között (esetünkben minden él súlya 1).

4 Az adatok előkészítése

A döntési feladatoknál alkalmazott páros összehasonlítások általában a legjobb elem kiválasztását, vagy ezen túlmenően egy részleges vagy teljes rangsor felállítását tűzik ki célul. A döntéshozó minden elempárt összehasonlít az előre megadott arányskálán (vagy összehasonlításait egy ilyen skálára transzformálják), s az így generált PCM szolgál a további számítások alapjául. Mint az előző alfejezetben láttuk, a rangsor előállítása akkor a legegyszerűbb, ha a PCM konzisztens. Valós feladatoknál azonban általában nem biztosítható a döntéshozó tökéletes konzisztenciája, a PCM magától értetődő módon megsértheti az arányokra vonatkozó konzisztencia-szabályt, sőt, akár nem tranzitív körhármasok is előfordulhatnak. A döntéshozatallal foglalkozó kutatók és gyakorlati szakemberek erőfeszítései ezért többirányúak. Egyrészt az inkonzisztencia-indexek és küszöbök meghatározásával igyekeznek kiszűrni az ítéleteikben megbízhatatlan vagy egyszerűen hanyag döntéshozókat (akik például egy-egy elemet rosszul jegyeznek fel), majd vagy a döntéshozókat szembesítik következetlenségeikkel, rábírva őket az esetleges változtatásra, vagy pedig automatikus módszereket fejlesztenek ki (ha a döntéshozó nem elérhető) a PCM elemeinek korrekciójára, a közel-konzisztens vagy konzisztens állapot elérésére. Másrészt olyan kérdezési technikákat fejlesztenek ki, amelyek ezeket a hibákat a minimumra csökkenthetik. A szakirodalom mindegyik esetben rendkívül szerteágazó megoldásokat javasol (lásd például Pereira, Costa 2018). Esetünkben nincs szükség ezekre a módszerekre, hiszen nincs ,,döntéshozó", a PCM objektív adatokból épül fel, a mátrixban detektálható inkonzisztencia és intranzitivitás nem korrigálandó – viszont érdemes lesz majd tudnunk ezek mértékéről.

Ugyanakkor a PCM valós háttere a módszertan alkalmazása kapcsán egyéb problémákat vet fel. Ezeket az ATP-rangsoraink előkészítése során alaposan megvizsgáltuk, itt most azért térünk ki rájuk röviden, mert ugyanazokat a korrekciós és transzformációs elveket fogjuk a WTA-rangsoroknál is alkalmazni. A részletek a Temesi, Bozóki, Csató (2012) tanulmányban érhetők el.

a) Valamelyik játékospár esetében a nyert vagy veszített mérkőzések száma 0. Ekkor a W/L hányados (vagy reciproka) nevezőjében 0 jelenne meg. Ezek a párok kihagyhatók lennének ebből a technikai okból kifolyólag, ez azonban méltánytalan lenne a győztesekkel szemben, és eltorzítaná az összehasonlítást. Számítássorozatokat végeztünk több adatkorrekciós megoldással, ezek közül kettő jelenik meg a Bozóki, Csató, Temesi (2016) cikkben. Minden számításunk arra vezetett, hogy a W/L hányadosok korrekciója (a később tárgyalandó terjedelem-transzformációval együtt) nem okoz jelentős eltérést a rangsorokban, így a továbbiakban a (3)-ban leírt korrekciót alkalmaztuk a PCM alapadatok W/L alapú kitöltésekor. Formálisan legyenek az alapadatok:

z_{ij} $(i, j = 1, \dots, n, i \neq j)$: a P_i és P_j játékosok közötti mérkőzések száma, $(z_{ij} = z_{ji})$;

x_{ij} $(i \neq j)$: a P_i és P_j játékosok közötti azon mérkőzések száma, ahol P_i volt a győztes; (1)

$y_{ij} = z_{ij} - x_{ij}$ $(i \neq j)$: a P_i és P_j játékosok közötti azon mérkőzések száma, ahol P_i veszített.

A PCM mátrix legyen n elemű és jelölje a P_i és P_j játékosok közötti W/L hányadosokat p_{ij}, $i, j = 1, \dots, n$. Ekkor a teniszezők $\mathbf{P}$ páros összehasonlítási mátrixának elemei:

$$p_{ij} = \begin{cases} x_{ij}/y_{ij}, & \text{ha } i > j \text{ és } x_{ij} \neq 0,\ y_{ij} \neq 0; \\ y_{ji}/x_{ji} = 1/p_{ji}, & \text{ha } i < j \text{ és } x_{ji} \neq 0,\ y_{ji} \neq 0; \\ 1, & \text{ha } i = j. \end{cases} \tag{2}$$

p_{ij} és p_{ji} minden más esetben hiányzik. Végezzük el a következő adatkorrekciót:

$$p_{ij} = \begin{cases} x_{ij} + 2, & \text{ha } x_{ij} \neq 0,\ y_{ij} = 0; \\ 1/(y_{ij} + 2) = 1/p_{ji}, & \text{ha } x_{ij} = 0,\ y_{ij} \neq 0\ . \end{cases} \tag{3}$$

Jegyezzük meg, hogy ennek a felírásnak az értelmében a $\mathbf{P}$ mátrix nem teljes. Ha a két játékos nem találkozott egymással a pályán, akkor a megfelelő elemet üresen hagytuk. A mátrix p_{ij} általános elemének értelmezése: hányszor erősebb a P_i játékos, mint a P_j játékos.

b) Az alapadatokból a fenti módon generált PCM – mint ahogyan az az 1. táblázatból is könnyen látható – a z_{ij} összes mérkőzésszámban szélsőséges különbségeket mutat, s ezáltal a p_{ij} értékek terjedelme is nagy. Ennek a terjedelemnek az összehúzását hivatott elvégezni az alábbi transzformáció:

$$\begin{aligned} t_{ij} &= p_{ij}^{\min(z_{ij}/K, 1)} \\ K &= \max_{i,j} z_{ij} \end{aligned} \tag{4}$$

ahol a transzformációs kitevő K fenti választásával az egymás elleni mérkőzések számának és az összes párra vonatkozó mérkőzések száma maximumának a hányadosa. Ha mindenki ugyanannyi mérkőzést játszott, akkor $t_{ij} = p_{ij}$. Ha két játékos nagyon keveset játszott egymással, akkor az adott t_{ij} elem – érthetően – az 1-hez közelít, és mivel a többi hányadost is összehúzza a transzformáció, elkerüljük az erőfölényt jellemző t_{ij} értékek szélsőségeit, kiegyensúlyozottabbá téve az induló $\mathbf{T}$ mátrixot.

c) Az egymás elleni mérkőzések számbavételével készített rangsor több észrevételt is generált, ezekre érdemes röviden kitérni:

- Miért nem korrigáljuk az eredményeket az egyes játékosok pályafutásának korai, legsikeresebb és késői szakaszának megfelelően? Felmerülhet az, hogy nem méltányos egy későbbi nagy játékos korai mérkőzéseit az éppen karrierjük csúcsán lévő játékosokkal szemben ugyanolyan súllyal figyelembe venni, mintha pályájuk azonos szakaszában lennének. Ezt két okból is elvetettük. Egyrészt nem könnyű a korai és késői szakaszok meghatározása, mind ez, mind pedig a súlyok nagysága eléggé erős szubjektív tényezőt jelentene az elemzések számára. Másrészt mindig vannak olyan teniszezők, akik már pályájuk elején nagyon jók bárki ellen (Austin, Osaka), egyesek hullámzó teljesítményt nyújtanak (Muguruza), mások mindig jók (Serena Williams), míg vannak olyanok is, akik hamar visszavonulva el sem érik a késői szakaszt (Henin), a nagy visszatérőkről nem is beszélve (Clijsters). Véleményünk szerint tehát az időtényező bevétele több torzítást vinne a rangsorokba, mintha ezzel nem foglalkozunk[7].

- Különbséget lehetne tenni a különböző borításokon elért eredmények között. Ez ismét ingoványos terepre vezetne bennünket, akkor is, ha a borításokat látnánk el súlyokkal, és akkor is, ha egyénileg tennénk különbséget. Az első megoldás ismét túlságosan szubjektív lenne (például salak 1, fű 2, kemény pálya 0,5) – a szakértők gyanakvását váltanánk ki, ha bevezetnénk egy ilyen ötletet. A második esetben újra csak az eldönthetetlen viták terepére tévednénk: hogy jövünk ahhoz, hogy Nadal salakos eredményeit vagy Federer füves eredményeit kisebb súllyal vegyük figyelembe (mert ők „túl jók” ezeken a felületeken), vagy éppen ellenkezőleg: Nadal füves eredményeit vagy Federer salakos eredményeit átsúlyozzuk – miközben persze akkor ezt mindenki mással is egyenként meg kellene tenni. Az egyetlen jó megoldás az, ha a pályaborítások szerint külön rangsorokat készítünk. Ez valóban megtehető, az adatok szétszedhetők, azonban mi most ezt a változatot nem vállaltuk.

- Elvileg figyelembe lehetne venni, hogy egy-egy győzelem 3 nyert játszmára menő küzdelemben dőlt el, vagy 2 nyert játszmában. Az ATP elemzésben ezzel nem foglalkoztunk. A nők esetében ilyen probléma nincs, mert mindenütt, a négy nagy versenyen is két nyert játszma dönt. Viszont felvetődhet az, hogy – ha már erősorrendről van szó – a játszmaarányokat is megnézzük. Sőt, a W/L hányadost két játékos között

[7] Ugyanakkor az életkor és a teljesítmény összevetése nem lehetetlen. Kovalchik et al. (2017) 250 top 100 női teniszező teljesítményét vizsgálta az 1990 és 2015 közötti időszakban. Erős összefüggést találtak a játékosok évenkénti legjobb helyezéseiből alkotott görbe alakja és a teljes karrier alatti legjobb helyezés között. A legmagasabban rangsorolt versenyzők voltak a legfiatalabbak az első rangsorbeli megjelenésük idején (a top 10 átlagban 15,5 évesen lépett be a rangsorba). Ők voltak azok is, akik a legtovább a rangsorban maradtak (29,0 év, ellentétben az 51-100. helyen lévők 24,4 évével).

a játszmákból is számolhatnánk! A férfiak esetében végeztünk számításokat a játszmaarányok segítségével, azonban a kapott rangsorok alig különböztek a mérkőzésarányokkal számított rangsoroktól, így nem látjuk értelmét egy ilyen mesterséges konstrukciónak. Viszont megvalósíthatónak látjuk azt, hogy a játszmaarányok segítségével ,,eltüntessük" a döntetleneket a **T** mátrixból. Ahol két játékos W/L hányadosa 1, ott a játszmaaránnyal korrigálva ,,eldönthetjük", hogy kit tekintünk ,,jobbnak". Ennek az az előnye, hogy a körbeverések kezelése módszertanilag egyszerűbben megoldható a döntetlenek nélküli páros összehasonlítási mátrixok esetében. Mivel elemzéseink erre az aspektusra is kitérnek, a 7. fejezetben bevezetünk egy ilyen korrekciót.

Végül érdemes megemlíteni azt, hogy számításaink eredményei hogyan interpretálhatók. A cél ugyan a nem teljes PC mátrixokra vonatkozó alkalmazási lehetőség bemutatása egy olyan példán keresztül, ahol egyben jól szemléltethető a közvetett kapcsolatokon keresztül történő rangsor kialakítása egymással közvetlenül össze nem hasonlított elemekre vonatkozóan, ám eljátszhatunk azzal a gondolattal, hogy egyben egyfajta ,,örökrangsort" kaptunk: számításaink megmutatják, hogy ki a legjobb (második legjobb, és így tovább) az elmúlt majd' ötven év női teniszében. Bár – mint látni fogjuk, illetve az ATP-rangsorokban a férfiakra vonatkozóan előző publikációinkban bemutattuk – a rangsoraink szakértői szemmel is reálisak, óvnánk attól, hogy túlértékeljük azokat. Ezek a rangsorok nyilván magukon viselik az alkalmazott módszertan kisebb-nagyobb torzító tényezőinek hatását és az adatbázis korlátait. Utolsó megjegyzésünk erről szól. A rangsorolt teniszezők kiválasztása jól indokolható, de mégis arról van szó, hogy egymás között rangsorolva őket nem vesszük figyelembe más teniszezőkkel elért eredményeiket. Vajon ugyanez jönne ki, ha minden év első 10 (vagy több) versenyzőjének eredményeit vesszük figyelembe? Nem tudjuk, ezt a munkát nem végeztük el – de elvileg megoldható. Saját eredményeink védelmében annyit azonban elmondanánk, hogy az éltenisz meglepően belterjes olyan szempontból, hogy a legjobbak a mérkőzéseik aránylag nagy százalékában egymás ellen játszanak. Különösen, ha már ,,befutottak" – erre magyarázatot adnak a versenyekről és a kiemelésekről szóló információk a második fejezet elején. A 27 játékos 22 310 mérkőzést játszott teljes pályafutása alatt összesen. Kiválóságukat mutatja, hogy ebből csak 5280-at veszítettek el (24%). Az egymás elleni mérkőzések száma (adatbázisunk összes mérkőzése) 3292, azaz a lejátszott mérkőzéseiknek mintegy 15%-a egymás ellen történt – ez igen sok, ha figyelembe vesszük, hogy egymást át nem fedő karrierpályáik miatt eleve más játékosok alkotják a generációs többséget, akikkel versenyezhettek. Nem véletlen, hogy ezen adatok kapcsán kiemelkedik Serena Williams, aki az általunk vizsgált 25 évében (1995–2020) mérkőzései 24,4%-át játszotta 22 ellen a többi játékos közül.

5 A női teniszezők ranglistáinak empirikus elemzése

A **T** mátrix alapadataira alkalmazhatjuk a logaritmikus legkisebb négyzetek becslési módszerét (LLNM), mivel a 2. ábra gráfja összefüggő, vagyis az 1. Tétel feltétele teljesül. Az így kapott rangsor a *2. táblázatban* látható. A táblázatban a teniszezőket az LLNM becslésből kapott súlyok szerint raktuk sorrendbe, majd a súlyok melletti második oszlopba egy „naiv mutatóként" az összesített W/L szerinti rangsor-helyezéseket írtuk be.

	LLNM	W/L	$\text{LLNM}_{K=30}$	SV
S. Williams	1	1	1	1
Graf	2	2	2	2
Navratilova	3	5	3	3
Hingis	4	6	4	4
Clijsters	5	4	5	5
Henin	6	3	6	6
V. Williams	7	7	7	7
Davenport	8	8	9	8
Evert	9	9	8	9
Barty	10	13	10	11
Seles	11	15	11	10
Pliskova	12	11	12	13
Sharapova	13	10	13	12
Osaka	14	19	14	14
Halep	15	12	15	15
Kerber	16	16	16	16
Azarenka	17	14	17	17
Muguruza	18	17	18	18
Mauresmo	19	18	19	19
Wozniacki	20	21	21	20
Safina	21	24	22	21
Austin	22	20	20	22
Ivanovic	23	23	23	23
Capriati	24	25	24	24
Jankovic	25	26	25	25
Sanchez	26	27	27	26
Goolagong	27	22	26	27

2. táblázat. A női élteniszezők rangsorai

Az eredmény valószínűleg nem lepi meg a teniszrajongókat. Az első három helyen Serena Williams, Graf és Navratilova állnak, majd Hingis, Clijsters és Henin követi őket: sokan mindenféle számolgatás nélkül is megadnának egy ilyen rangsort. Sharapovát talán előbbre vártuk volna, ám hosszú karrierje végén túl sok vereség csúszott be neki. Megnyugtató, hogy a régebbi nagyok és a Williams testvérek mellé újak is felzárkóztak, és hogy éppen Barty helyezése a legjobb, aki a legutóbbi idők legstabilabb teniszezőnője. Osaka rövid idő alatt ért el kiemelkedő eredményeket nagy versenyeken. Ha átgondoljuk azt, hogy egymással a pályán nem mérkőző játékosokat is összevetettünk, s a rangsor a páros összehasonlítások révén közvetett módon, egy triviálisnak nem mondható súlymeghatározás alapján alakult ki, akkor az eredmény jól példázza módszerünk erejét.

A W/L rangsor statisztikai szóhasználattal élve nagyon jó proxynak bizonyul. A Spearman-rangkorreláció 0,957. Sejtésünk szerint arról van szó, hogy az aggregált mérkőzésarányokkal számolt, játékosonként egyetlen W/L hányadosok és azok „szétbontása" az adathalmazunkban lévő páronkénti W/L hányadosokra erős kapcsolatban van egymással a rangsorkészítő módszer révén – ez azonban egyáltalán nem magától értetődő, és a PCM modell mélyébe történő további betekintést igényel.

A férfi játékosok esetében az idézett publikációkban ugyan megnyugtató volt a transzformáció hatásának és a becslési módszer hatásának vizsgálata a rangsor változására, de ez lehetett akár az ATP-adatokból adódó, csak a férfi éljátékosokra vonatkozó eredmény. Mivel a női teniszezők transzformációs paramétere (az aktuális maximum, $K = 80$) egy kiugró érték, kiszámoltuk a rangsort egy átlagosnak mondható $K = 30$ értékkel is. A lényeg a PCM elemeinek kevésbé erős összehúzása volt. A 2. táblázat W/L utáni oszlopába a $K = 30$ paraméterrel transzformált PCM LLNM becsléssel kialakított sorrendjét írtuk be. Csekély változást tapasztalunk: Evert és Davenport a 8-9. pozíciókban helyet cserél, a 20–22. helyen változik Austin, Wozniacki és Safina sorrendje, illetve az utolsó helyeken Sanchez és Goolagong is helyet cserél. A Spearman-rangkorreláció 0,997. Más K értékeket is használva a számításokban azt láttuk, hogy a transzformáció paraméterének hatása a nőknél is elhanyagolható. Emellett belátható, hogy adott PCM-re a maximális mérkőzésszámnál nagyobb transzformációs paraméter használata pontosan ugyanazt a sorrendet eredményezi, mint a maximális érték.

Az eredeti PCM adataival elkészítettük a sajátvektor módszerrel is a becslést, az eredményeket beírva a 2. táblázat következő oszlopába, SV fejléccel. A rangsor a 10-11. helyen cserélődött meg Barty és Seles, illetve mögöttük Pliskova és Sharapova között. A Spearman-rangkorreláció 0,999.

Tapasztalataink tehát megegyeznek a férfi teniszezőknél részletesebben megvizsgált és leírt helyzettel. Ez ugyanazt a következtetést sugallja, ami egyben a tenisz rangsorokra vonatkozó alkalmazásaink egyik lényeges empirikus eredménye: mind a férfi, mind a női tenisz top játékosoknál a W/L hányadossal jellemzett egymás elleni eredmények rendkívül robusztusak abban a tekintetben, hogy a nem teljes páros összehasonlítási mátrixokból kapott rangsorok az adatmátrix észszerű mértékű korrekciójára, transzformációjára nem érzékenyek, és a becslési módszerek sincsenek rájuk szignifikáns hatással. A módszer tehát az alkalmazásokban bemutatott célra ajánlható.

Nézzük meg azonban az eredményeket más szemszögből is. A döntéselmélet nagy hangsúlyt fektet arra, hogy rangsorkészítésnél megvizsgálja kimaradó vagy újonnan bevett elemek, elemhalmazok hatását, hátha könnyen felborítható a szép sorrend. Mit látunk, ha a játékosok egy-egy csoportját kiemelve újabb (rész)rangsorokat készítünk?

Első részmátrixunk sorrendje, a *3. táblázat* PCM1 blokkja, a 2. táblázat LLNM rangsorának első 8 helyezettjének kihagyásával készült sorrendet mutatja. A helyezési számokat összevetve az eredeti rangsorral azt látjuk, hogy csak Jankovic és Goolagong maradt a helyén. Seles, Sharapova és Evert állnak az első 3 helyen, őket Mauresmo, Azarenka és Halep követi. A PCM2 blokkban az eredeti sorrend utolsó 8 helyezettjét elhagyva kaptunk egy új rangsort az első 19 játékosra vonatkozóan. Serena Williams és Graf ugyan megmaradt az első két helyen, utánuk viszont Barty, Davenport, Pliskova és Venus Williams a sorrend – erősen eltérve a 2. táblázat első oszlopának rangsorától. A legérdekesebb talán Seles helyzete: ha az első nyolcat hagytuk ki, akkor ő volt az első, ha az utolsó nyolcat, akkor ő az utolsó!

	PCM1	PCM2	PCM3	PCM4	PCM5
S. Williams	-	1	-	1	-
Graf	-	2	-	-	1
Navratilova	-	9	1	-	2
Hingis	-	7	-	6	3
Clijsters	-	11	-	2	4
Henin	-	12	-	3	5
V. Williams	-	6	-	4	-
Davenport	-	4	-	5	6
Evert	3	18	2	-	-
Barty	7	3	-	-	13
Seles	1	19	-	-	-
Pliskova	10	5	-	-	12
Sharapova	2	16	-	7	8
Osaka	12	10	-	-	15
Halep	6	8	-	-	14
Kerber	15	13	-	-	-
Azarenka	5	15	-	-	-
Muguruza	14	14	-	-	16
Mauresmo	4	17	-	8	10
Wozniacki	9	-	-	11	11
Safina	8	-	-	9	-
Austin	18	-	3	-	7
Ivanovic	11	-	-	10	17
Capriati	13	-	-	-	18
Jankovic	17	-	-	12	-
Sanchez	16	-	-	-	19
Goolagong	19	-	4	-	9

3. táblázat. Különböző módokon generált játékoscsoportok rangsorai

Megváltoztak az egyes játékosok mérkőzésszámai: van, akinek sokkal kevesebb lett, másoknál ez a csökkenés nem olyan jelentős. A kihagyott játékosokkal együtt kihagytuk az eredményeiket is. Van, akin ez nagyot lendített (kimaradtak, akik nagyon megverték), van, akinél ez nem számított (többségükkel nem is játszott), és van, akinél eltűntek, akiknél ő jobb volt. És persze eltűntek a közvetett hatások is. Nem várhattuk tehát, hogy az eredmények hasonló sorrendet adjanak – az lett volna a meglepő és magyarázatra szoruló jelenség, ha így történt volna. A klasszikus kérdésfelvetés szerinti sorrendek a kilépő alternatívák szerinti függetlenséget cáfolják.

Generáltunk még néhány más logikájú részhalmazt is. A PCM3 blokk az 1970-es években indult négy nagy teniszezőé. Azt látjuk, hogy a négy játékos sorrendje tökéletesen követi a teljes mátrixból számított helyezéseket. A blokkon kívüli eredményeik ezek szerint követik a blokkon belüli mintát.

A PCM4 blokk az előző négy játékos után következő korszakból választ ki 12 játékost. Ennek a részhalmaznak az a jellemzője, hogy az általa felölelt időszak olyan, hogy ezek a játékosok karrierjük zömét benne töltötték. Az első hat játékosból 5-nek az egymás utáni helye az eredeti rangsor szerint következik, Hingis az, aki Clijsters, Henin, Venus Williams és Davenport mögé csúszik; Wozniacki pedig Safina és Ivanovic mögé kerül a többi kilenc játékosnak a teljes rangsorhoz képest változatlan sorrendje mellett. Az eredmény arra utal, hogy kiválaszthatók a teljes időszak helyezéseinek csekély mértékben ellentmondó helyezésű játékosok kisebb-nagyobb részhalmazai. Az azonos időszakban lejátszódott versengés (eredmények) gyakorlatilag annyira meghatározza a játékosok relatív helyét, hogy az még az időszak tágításával sem változik.

Elméleti szempontból érdekes, hogy a PCM3 és PCM4 blokkban vizsgált játékosok egy-egy teljes részmátrixot (részgráfot) határoznak meg, azaz a blokkokon belül minden teniszezőnő játszott mindenki mással. Így lehetővé válik a teljes mátrixok inkonzisztenciájának vizsgálata is, ami a robusztus eredményeket alátámasztva szintén igen alacsony (0,02 alatti CR indexek).

Végül válasszunk ki véletlen módon 19 játékost (az első vagy utolsó 8 elhagyásával egyező elemszámot kapva) és helyezzük el a sorrendjüket a PCM5 blokkba. Kimaradt a két Williams, Evert, Seles, Kerber, Azarenka, Safina és Jankovic. Először arra érdemes felfigyelnünk, hogy az első hat helyezett sorrendje megegyezik a teljes sorrendbeli helyezésekkel. Valószínűleg azért, mert a Williams nővérek és Evert elleni hasonló irányú eredmények nem befolyásolták az egymás elleni eredményekből kialakuló erősorrendet. Sharapova és (főleg) Austin előbbre került – jól jártak a kimaradók elleni eredményeik törlésével –, míg a többiek az eredeti súlyok szerint is igen közel lévén egymáshoz, most is együtt maradtak, miközben rangsorfordulások is történtek. A véletlen kiválasztással más esetekben is hasonló helyzet várható: az új eredménymátrixban maradt elemek segítenek előre menni vagy hátrébb sorolnak egyes játékosokat.

6 Kitekintés a férfi játékosok sorrendjére

A férfi ranglistavezetők 2015-ig kitöltött adatbázisát (Bozóki, Csató, Temesi 2016) kiegészítettük 2020-ig, illetve egyetlen új ranglistavezetővel bővítettük: Andy Murray lett a 26. játékos. Az 5 év új eredményeinek hozzáadása természetesen csak az ebben az időszakban aktív játékosoknál jelentett változást a PCM elemeiben. Kiszámoltuk ennek a hatását, azaz a Murray nélküli ranglistát. A *4. táblázat* első oszlopában a 2020-as, a második oszlopban a 2015-ös rangsor látható. A Spearman-rangkorreláció 0,988. A változások csekélyek. Legfeltűnőbb Nadal és Federer helycseréje az első helyen. Ez leginkább annak köszönhető, hogy kettőjük mérlegén Federer javítani tudott (10/22-ről 16/24-re változott az arány), miközben a többi játékos ellen Nadal a sérülései miatt kevésbé nyújtott olyan stabil teljesítményt, mint Federer.

	2020	2015	Változás
Federer	1	2	↑1
Nadal	2	1	↓1
Sampras	3	3	-
Djokovic	4	7	↑3
Lendl	5	4	↓1
Becker	6	6	-
Borg	7	5	↓2
Hewitt	8	9	↑1
Agassi	9	8	↓1
Kuerten	10	10	-
Safin	11	11	-
Nastase	12	13	↑1
McEnroe	13	12	↓1
Ferrero	14	14	-
Roddick	15	15	-
Wilander	16	16	-
Rios	17	17	-
Rafter	18	18	-
Kafelnikov	19	20	↑1
Moya	20	21	↑1
Edberg	21	22	↑1
Newcombe	22	19	↓3
Courier	23	23	-
Muster	24	24	-
Connors	25	25	-

4. táblázat. A férfi teniszezők rangsorának változása, 2015-2020

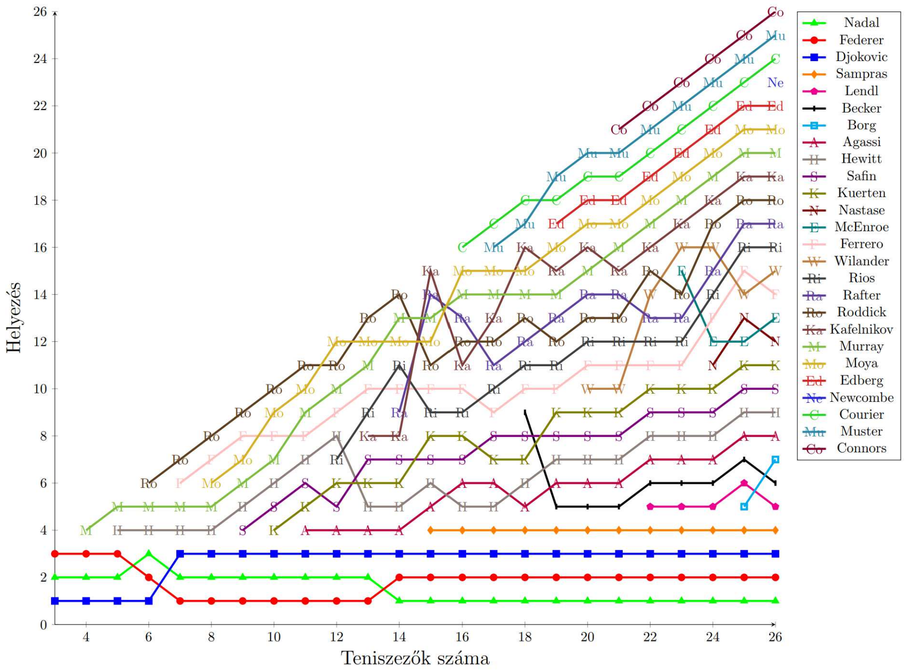
Az *5. táblázat* már Murray-t is tartalmazza: az első oszlop 2020-as rangsorában beékelődött a 20. helyre. Éppen neki ,,köszönhető", hogy Djokovic immár a harmadik helyen áll. Míg Samprasnak (természetesen) nem állt módjában Murray-vel játszani, ezáltal ,,megvédve" harmadik helyezését, addig Djokovic 25/11-es W/L aránya elegendő lehetett a feljebb lépéshez. A táblázat második oszlopában a W/L arányok által meghatározott rangsort láthatjuk. Bár nagyobb különbségeket tapasztalunk az LLNM rangsorhoz képest, mint a nők esetében, a Spearman-rangkorreláció így is igen erős, 0,904.

Az 5. táblázat további oszlopai a 3. táblázathoz hasonlóak, mert a férfiaknál is elvégeztünk néhány részmátrixra bontást. A táblázat PCM1 és PCM2 blokkjában – a nőkhöz hasonlóan – az első nyolc és az utolsó nyolc helyezett elhagyásával készült rangsorok láthatók. Nagy meglepetésnek látszik, hogy a PCM1 blokkban Murray az első helyezett! Egy pillanatra előre ugorva a PCM4 sorrendhez, ahol szintén elöl, a második helyen áll, láthatjuk, hogy ez az a két eset, amikor hiányoznak a Nadal (7/17), Federer (11/14) és Djokovic (11/25) elleni rossz eredményei, de beszámít a többiek elleni általában pozitív mérleg (csak Safintól kapott ki egyetlen mérkőzésükön). Újra látjuk tehát, hogy az elemszám változása megváltoztatja a sorrendet, ám a változások a W/L arányok logikája mentén magyarázhatóak. Ha a PCM1 oszlopban 8-at hozzáadunk a rangszámokhoz, akkor láthatóvá válik az elmozdulás a teljes rangsorhoz képest. Nagyot lépett még előre Roddick és Edberg, erősebben hátracsúszott Safin, Nastase, Rios – a magyarázat hasonló az előzőhöz. A PCM2-nél Becker és Nastase mozdult hátrébb. Itt azt érdemes megjegyezni, hogy az eredeti rangsorban az utolsó helyeken szereplő játékosok elhagyása – intuícióinkkal egybevágóan – az első kétharmadra csekély hatású.

	2020	W/L	PCM1	PCM2	PCM3	PCM4	PCM5
Nadal	1	2	-	2	2	-	1
Federer	2	5	-	1	3	-	-
Djokovic	3	7	-	3	1	-	-
Sampras	4	1	-	4	-	1	2
Lendl	5	6	-	5	-	3	5
Becker	6	4	-	11	-	5	4
Borg	7	3	-	6	-	6	3
Agassi	8	8	-	7	-	8	-
Hewitt	9	10	2	9	-	4	-
Safin	10	14	8	8	-	9	7
Kuerten	11	9	7	10	-	11	9
Nastase	12	12	10	17	-	12	-
McEnroe	13	11	4	16	-	13	14
Ferrero	14	20	9	13	-	10	-
Wilander	15	15	5	18	-	14	-
Rios	16	23	15	12	-	15	12
Rafter	17	17	12	14	-	17	11
Roddick	18	19	3	15	-	7	6
Kafelnikov	19	16	11	-	-	19	13
Murray	20	18	1	-	-	2	8
Moya	21	24	14	-	-	16	10
Edberg	22	13	6	-	-	18	15
Newcombe	23	22	17	-	-	20	-
Courier	24	21	13	-	-	21	16
Muster	25	26	16	-	-	22	17
Connors	26	25	18	-	-	23	18

5. táblázat. A különböző játékoscsoportok rangsorai a férfiak esetében

A rangsorokat, illetve azok változását megvizsgáltuk abban az esetben is, ha az elemeket (a teniszezőket) egyesével vonjuk be az adathalmazba. A jelenleg is aktív nagy hármassal (Nadal, Federer, Djokovic) kezdtük, megnéztük, hogy milyen rangsort kapnánk, ha csak az ő egymás elleni eredményeiket vennénk figyelembe, ez a PCM3 rangsor. Ezt követően időrendi sorrendben haladtunk visszafelé aszerint, hogy ki fejezte be legkésőbb a karrierjét (döntetlen esetén ki kezdte el később profi pályafutását). Egyesével vontuk be a teniszezőket, amíg eredeti rangsorunk mind a 26 játékosa be nem került az adatbázisunkba. Ezzel összesen 24 darab rangsort határoztunk meg, melyet az egyes teniszezők bevonása nyomán megfigyelhető változásokkal együtt a *3. ábra* mutat be. Fontos megjegyezni, hogy pontosan annyi rangsorfordulást figyelhetünk meg egy adott játékos bevonása nyomán (ahol elindul az őt reprezentáló görbe), ahány görbe metszi egymást eggyel a bevonás előtti és a bevonáskori értékek közt. Láthatjuk, hogy a rangsor igen stabil, egy-egy játékos bevonása nem sokat változtat, illetve, ha változtat is, akkor is általában egy, maximum két helyet mozdulnak föl vagy le a különböző teniszezők.



3. ábra. Az egyesével beléptetett teniszezők rangsorainak változása

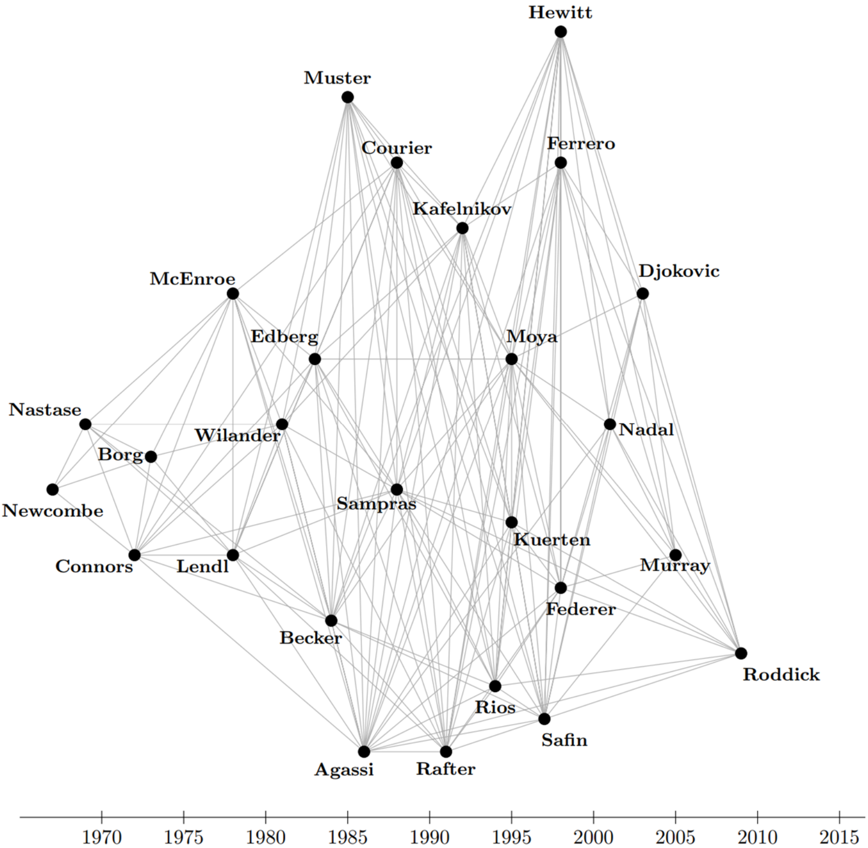
Azok a ritka esetek, amikor több helyet is változtatnak, általában annak köszönhetők, hogy egy már bekerült teniszező az eddig bent lévő többi játékossal valójában alig játszott (például Becker), és a most bevont új elem (Edberg) az, akivel valójában megbízható az összehasonlítása. Valamint az is előfordul persze, hogy egy játékos egy újonnan bevont elemtől nagyon sokszor kikapott (Kafelnikov vereségei Samprastól), vagy éppen olyan játékost sikerült többször legyőznie, aki a többi eredménye alapján jobban szerepelt, mint ő (Roddick és Sampras viszonya). Érdemes azt is megjegyezni, hogy az új játékosok bevonásának általában azokra van nagyobb hatása, akikkel valóban sokat játszottak pályafutásuk során. A rangsor eleje és vége igen robusztusnak tűnik, középen láthatunk több rangsorfordulást az elemek bevonása során, azonban itt is egyfajta klaszterek alakulnak ki, és csak ezeken belül változik az egyes játékosok helyezése.

Mivel az első 3 helyezett játékos hosszú pályafutása során rendkívül erős hatást gyakorolt a rangsorokra, őket (és csak őket) kihagyva is készült egy rangsor, ez az 5. táblázat PCM4 blokkjában látható. Érdekesség, hogy a teniszről szóló szakértői véleményekkel összhangban számításaink megerősítik azt, hogy a nagy hármas két legnagyobb kárvallottja Roddick (akinek a pályafutását Federer keserítette meg a legjobban: 3/21-es arány és legendás vereségek nagy döntőkben), illetve Murray (akinek a hármassal vívott csatáinak eredményeit már tárgyaltuk).

Végül, a női teniszezőkkel végzett számításokhoz hasonlóan, itt is készült egy 18 elemű véletlen mintára vonatkozó rangsor, ahol a tanulság is hasonló: a kiesők csak egy-két kiugró változást okoznak eredményeik kihagyása révén, ilyen Roddick, Moya és Murray javulása, illetve McEnroe rosszabb helyezése.

7 A (nem teljes) PCM tulajdonságainak vizsgálata a konkrét alkalmazás keretei között

Ahhoz, hogy még jobban megértsük a kapott eredményeket, adataink struktúrájának mélyebb elemzésére van szükségünk, amelyben elsősorban a PCM-ek reprezentáló gráfjai lesznek a segítségünkre. Ezt a nők esetében a 2. ábra szemlélteti, míg a férfiaknál a Bozóki, Csató, Temesi (2016) tanulmányban közölt gráfhoz egyetlen csúcsot adunk hozzá, ez a gráf a *4. ábrán* látható.



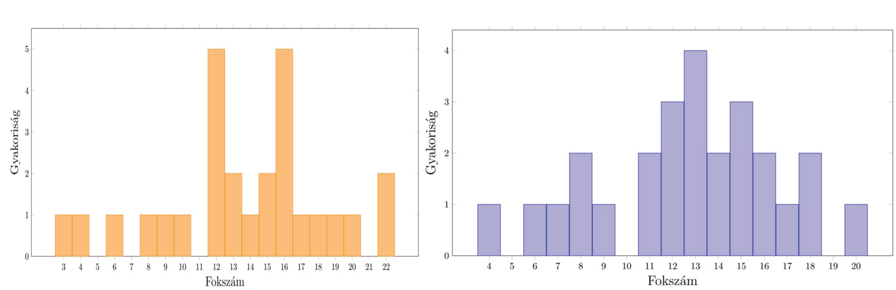
4. ábra. Az ATP ranglistáit vezető férfi teniszezők egymás elleni mérkőzéseinek gráfja

A két reprezentáló gráfra (női–WTA és férfi–ATP) vonatkozó legfontosabb tulajdonságokat a *6. táblázat* első és második oszlopa mutatja be.

	WTA	ATP	WTAmod	ATPmod
Csúcsszám (alternatívák)	27	26	23	21
Élszám (összehasonlítások)	184	165	169	136
Minimális fokszám	3	4	9	7
Maximális fokszám	22	20	22	20
Átmérő	4	4	2	2
Átlagos legrövidebb út	1,63	1,63	1,33	1,35
Élösszefüggőség	3	4	9	7
Csúcsösszefüggőség	3	4	8	7

6. táblázat. A reprezentáló gráfok tulajdonságai

A teniszezőnők esetén láthatjuk, hogy van olyan játékos (Goolagong), aki a világelsők közül mindössze három másikkal (Evert, Navratilova, Austin) játszott valaha, amit a reprezentáló gráf minimális fokszáma mutat. Ezzel ellentétben a Williams testvérek mindketten 22 világelsővel játszottak (maximális fokszám) az általunk vizsgált 27-ből. Nagyon hasonló a helyzet a férfiaknál is, ahol a legkevesebb (4) teniszezővel Newcombe mérte össze erejét, míg a legtöbbekkel (20) Agassi mérkőzött meg. A két reprezentáló gráf nagyon hasonló összefüggőségi mutatókkal rendelkezik, a nők esetén bármely 2, a férfiak esetén bármely 3 él, illetve csúcs törlése még mindig összefüggő gráfhoz vezetne, míg a leghosszabb legrövidebb út (az átmérő) mindkét esetben 4 hosszúságú. Érdekes, hogy ezek jellege is hasonló, a nők esetében a korai időszak kiemelkedő játékosa, Goolagong, és a fiatal játékosok (Osaka, Muguruza, Barty) között vétetik fel, míg a férfiaknál „a nagy öregek", Newcombe és Borg vannak ilyen távolságra a legfrissebb világelsőktől, Djokovic és Murray személyében. Elmondható tehát, hogy az átmérő az időben egymástól távol lévő teniszezők közt figyelhető meg. A fokszámok eloszlásáról árnyaltabb képet kaphatunk az *5. ábra* segítségével.



5. ábra. Az általunk vizsgált női (balra) és férfi (jobbra) teniszezők gráfjainak fokszám-eloszlása

Ezen eloszlást leggyakrabban a hálózatelméletben szokás használni (Albert, Jeong, Barabási 2000) különböző rendszerek, például pénzügyi hálózatok (Berlinger, Daróczi, Dömötör, Vadász 2017) leírására. Esetünkben az átlagos legrövidebb utak mindkét gráfra alacsony értéket vesznek föl, a fokszámok viszonylag szabályosan oszlanak meg egy átlagos érték körül, illetve jellemző a klaszterezettség is (a csúcsok átlagos klaszterezettsége 0,8 a teniszezőnők és 0,79 a férfiak esetén a szokásos [0,1] skálán), azaz a játékosok egy-egy csoportja leginkább egymással mérkőzött meg a pályafutásuk során. Ezen tulajdonságok alapján úgynevezett kis-világ (vagy Watts-Strogatz) típusú hálózatokkal van dolgunk (Watts, Strogatz 1998).

Érdemes még megemlíteni, hogy a teniszezőnők esetében a Williams-testvérek, akiket a két legnagyobb fokszámmal rendelkező csúcs reprezentál, esetleges törlése mellett a gráf összefüggőségi mutatói lényegében változatlanok maradnak. Azonban, kritikus csúcshármasokat alkot Evert, Navratilova és Graf, valamint Evert, Navratilova és Austin, olyan értelemben, hogy törlésükkel a reprezentáló gráf összefüggősége megszűnik. Másrészről viszont, ha a négy legkorábbi játékost (Goolagong, Evert, Navratilova, Austin) töröljük a gráfunkból, akkor egy sokkal erősebben összefüggő hálózatot kapunk, melynek tulajdonságait a 6. táblázat harmadik oszlopa jellemzi (WTAmod). Hasonló eredményt érünk el a férfiak esetén, ha a legkorábbi (Newcombe, Nastase és Borg) és legkésőbbi (Djokovic és Murray) játékosokat töröljük, az ehhez tartozó eredményeket a 6. táblázat negyedik oszlopa tartalmazza (ATPmod). Az ezen részgráfok, illetve részmátrixok alapján számolt rangsorok szinte teljesen megegyeznek a teljes gráfokból számolt rangsorokkal, így lehetséges, hogy éppen ezek az erősen összefüggő kapcsolatok határozzák meg elsősorban a teljes rangsorokat is. Érdekes az a tény is, hogy a nők esetében a módosított, erősen összefüggő gráf két csillaggráf uniója néhány éllel kiegészítve, ahol a két középpont éppen a Williams testvérek, azaz ebben az esetben ők ketten minden más játékossal játszottak, mindenkivel közvetlenül összehasonlítjuk őket. Hasonló reprezentáló gráfokat figyelhetünk meg a döntéselméletben népszerű best-worst módszer alkalmazásakor (Rezaei 2015). A férfiaknál egy Agassi középpontú csillaggráfot kapunk néhány él kiegészítésével, ha a korábban említett törléseket elvégezzük. Az eredmények robusztusságáért akár ezek a struktúrák is felelősek lehetnek.

Az adatok egy másik rendkívül fontos tulajdonsága, amire részleteiben is szeretnénk kitérni, nem más, mint az ordinálisan intranzitív körhármasok, azaz a körbeverések aránya, illetve ezek hatása az általunk alkalmazott módszertanra. Ennek vizsgálatában is segítségünkre lehetnek a reprezentáló gráfok, amennyiben irányított gráfokká alakítjuk őket. Adott él mindig attól a játékostól indul ki, aki többször győzött az adott párban, és afelé mutat, aki többször szenvedett vereséget.

Kendall és Babington Smith (1940) az ordinálisan intranzitív körhármasok számának eloszlását adták meg alacsony alternatívaszám ($n \leq 7$) esetén, illetve ezek számának szignifikanciájára vonatkozóan javasoltak egy statisztikai tesztet. Az eloszlást több alternatívára ($8 \leq n \leq 10$) is kibővítette Alway (1962), míg az ennél is nagyobb esetekre vonatkozóan az eloszlást közelítő eredmények születtek. Moran (1947) bizonyította, hogy a körhármasok eloszlása az elemszám növekedésével tart a normális eloszláshoz – de ez a konvergencia lassú. Knezek, Wallace és Dunn-Rankin (1998) pedig megvizsgálták a korábban Kendall és Babington Smith (1940) által javasolt khí-négyzet eloszlás segítségével történő közelítést, amit még 15 alternatíva fölött is megfelelőnek találtak. Iida (2009) döntéselméleti környezetbe helyezte a körhármasok statisztikai tesztjét, melyet a bináris és döntetlent nem tartalmazó

PCM-eken alkalmazott. Kulcsfontosságú azonban megemlíteni, hogy az eddig említett eredmények mind teljesen kitöltött és döntetleneket nem tartalmazó páros összehasonlításokra építenek.

Döntetleneket is megengedő vizsgálatok esetén Kułakowski (2018) határozta meg az ellentmondásos körhármasok maximális számát bármilyen alternatívaszám mellett, és ehhez kapcsolódóan javasolt egy indexet is. Azonban ő nemcsak az ordinálisan intranzitív körhármasokat tekintette ellentmondásosnak, hanem azokat az eseteket is, amikor egy triádban például pontosan két egyenlőség (irányítatlan él) van az elemek között. Ez a megközelítés a mi példánkban kevésbé jól interpretálható: a sportpéldákban a körbeverés nem tekinthető logikailag ellentmondásosnak. A javasolt indexre megbízható statisztikai tesztek sem léteznek.

Ezért Iida és Kułakowski vonalának követése helyett úgy döntöttünk, hogy a játszmaarányok segítségével megszüntetjük a döntetleneket az adatainkban, és ezt követően az ismert teszteket alkalmazzuk a körhármasok számának szignifikanciájára. Mivel ekkor még mindig maradtak döntetlenek, ezeket az eseteket a játékarányokkal tettük eldönthetővé, végső esetben, ha ez is döntetlent eredményezett (egyetlen ilyen eset volt rangsoronként), akkor az LLNM rangsorbeli helyezéseket alkalmaztuk. Ekkor azonban még mindig beleütközünk a nem teljesség problémájába, mivel a tesztek csak teljesen kitöltött mátrixokra működnek. Ennek kiküszöbölésére a férfi és női teniszezők esetén is az LLNM módszer által adott rangsort hívjuk segítségül. Ahol két játékos nem játszott egymással, azaz a gráfban nem fut él közöttük, azt vesszük a párban győztesnek, aki az LLNM rangsor szerint előrébb végzett. Így létrehozzuk a megfelelő gráfot (mátrixot), aminek már minden éle irányított és teljes is, s ezen alkalmazzuk a tesztet. Az eredeti és a teljes reprezentáló gráfok körhármasokra és döntetlenekre vonatkozó adatait mutatja be a *7. táblázat.*

	WTA	ATP	WTAteljes	ATPteljes
Kitöltöttség	184/351	165/325	351/351	325/325
Döntetlenek	20	19	0	0
Ordinálisan intranzitív triád	81	57	293	273
Ordinálisan intranzitív triád maximum	819	728	819	728

7. táblázat. A körhármasokra vonatkozó adatok

Láthatjuk, hogy mindkét esetben nagyjából a megfelelő PCM elemeinek a fele ismert (52,4% a nők és 50,8% a férfiak adataira), valamint a döntetlenek száma mindkét esetben 10 százalék körüli (10,9% és 11,5%). Az ordinálisan intranzitív triádok száma azonban a nem teljesen kitöltött mátrixokból származik, így ezt a teljes kitöltésre vonatkozó maximumokkal nem hasonlíthatjuk össze. Az LLNM rangsor alapján teljessé tett, és a játszmaarányok segítségével a döntetlenektől megtisztított női eset adatait láthatjuk a 7. táblázat harmadik oszlopában. A megfelelő χ^2 teszt (mely a normális közelítésnél szigorúbb az esetünkben) értéke 185, a hozzá tartozó p érték pedig praktikusan 0, ami azt jelenti, hogy a női teniszezők adatai az ordinálisan intranzitív triádokat tekintve az összes szokásos szignifikanciaszinten nagyon konzisztensek, azok száma nem befolyásolja szignifikánsan az eredményt. A 7. táblázat negyedik oszlopa tartalmazza a hasonlóan teljessé és döntetlenektől mentessé tett esetet a férfiakra. Ekkor a χ^2 teszt értéke 169, a hozzá tartozó p érték pedig ismét praktikusan 0, tehát az ordinálisan intranzitív triádok száma itt sem befolyásolja szignifikánsan az eredményeket.

8 Következtetések, nyitott kérdések

A nem teljesen kitöltött páros összehasonlítási mátrixokkal foglalkozó kutatásainkban a férfi és női tenisz nemzetközi éljátékosaira vonatkozó alkalmazásunk meggyőző empirikus bizonyítékokat szolgáltat arra, hogy ez a modell alkalmas jól interpretálható rangsorok előállítására. Tanulmányunk a kiválasztott játékosok egymás elleni mérkőzéseinek adatbázisára épül. A teljes adatrendszerrel és annak részhalmazaival végzett rangsorolások arra is rávilágítanak, hogy ez az adatrendszer robusztus abban az értelemben, hogy bár az elemszám változásával a rangsorok megváltoznak, az eltérések az adatmátrix elemzésével jól magyarázhatók és logikailag konzisztensek.

Új eredményeink közül a részhalmazokon történt vizsgálatok elsősorban döntéselméleti szempontból relevánsak, bár a részrangsorok a tenisz világát jól ismerők számára önmagukban is érdekesek lehetnek. Elemzéseink másik újdonsága a gráfreprezentációk vizsgálata. Itt kezdeti eredményeket értünk el: különösen fontosnak tartjuk a körhármasok (körbeverések) szerepének további elemzését. A további kutatások új irányait már nem feltétlenül az empirikus alkalmazás elmélyítése jelenti (például újabb tenisz rangsorok más játékosokkal, más típusú adatkezeléssel), hanem a nem teljesen kitöltött mátrixok tulajdonságainak pontosabb leírása. Ebben a gráfreprezentációk nyithatnak új utakat.

Irodalom

1. Albert, R., Jeong, H. & Barabasi, A. L. (2000): Error and Attack Tolerance of Complex Networks. *Nature,* 406, 378–382.

2. Alway, G. G. (1962): The distribution of the number of circular triads in paired comparisons, *Biometrika,* 49, 265–269.

3. Anholcer, M., Fülöp, J. (2019): Deriving priorities from inconsistent PCM using network algorithms, *Annals of Operations Research,* 274(1-2), 57–74.

4. Baker, R. D, McHale, I. G. (2014): A dynamic paired comparisons model: Who is the greatest tennis player? *European Journal of Operational Research,* 236(2), 677–684.

5. Baker, R. D, McHale, I. G. (2017): An empirical Bayes model for time-varying paired comparisons ratings: Who is the greatest women's tennis player? *European Journal of Operational Research,* 258(1), 328–333.

6. Berlinger, E., Daróczi, G., Dömötör, B., Vadász, T. (2017): Pénzügyi hálózatok mag–periféria szerkezete. A magyar bankközi fedezetlen hitelek piaca, 2003–2012. *Közgazdasági Szemle,* 64, 1160–1185. DOI: 10.18414/KSZ.2017.11.1160.

7. Bozóki, S., Csató, L., Temesi, J. (2016): An application of incomplete pairwise comparison matrices for ranking top tennis players, *European Journal of Operational Research,* 248(1), 211–218.
8. Bozóki, S., Fülöp, J., and Rónyai, L. (2010): On optimal completion of incomplete pairwise comparison matrices. *Mathematical and Computer Modelling,* 52(1), 318–333.
9. Bradley, R. A., Terry, M. E. (1952): Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons, *Biometrika,* 39(3), 324–345.
10. Brunelli, M. (2018): A survey of inconsistency indices for pairwise comparisons. *International Journal of General Systems,* 47(8), 751–771.
11. Chao, X., Kuo, G., Li, T., Peng, Y. (2018): Jie Ke versus AlphaGo: A ranking approach using decision making method for large-scale data with incomplete information, *European Journal of Operational Research,* 265(1), 239–247.
12. Crawford, G., Williams, C. (1985): A note on the analysis of subjective judgment matrices. *Journal of Mathematical Psychology,* 29(4), 387–405.
13. Csató, L. (2017): On the ranking of a Swiss system chess team tournament, *Annals of Operations Research,* 254(1-2), 17–36.
14. Csató, L. (2021): Tournament Design. How Operations Research Can Improve Sports Rules? Palgrave Pivots in Sports Economics
15. Csató, L., Tóth, Cs. (2020): University rankings from the revealed preferences of the applicants, *European Journal of Operational Research,* 286(1), 309–320.
16. Elo, A. E. (1978): The rating of chessplayers, past and present, Arco Pub. Gu, W., Saaty, Th. L. (2019): Predicting the Outcome of a Tennis Tournament: Based on Both Data and Judgments, *Journal of Systems Science and Systems Engineering,* 28(3), 317–343. DOI: 10.1007/s11518-018-5395-3
17. Iida, Y. (2009): The number of circular triads in a pairwise comparison matrix and a consistency test in the AHP, *Journal of the Operations Research Society of Japan,* 52(2), 174–185.
18. Kendall, M. G., Babington Smith, B. (1940): On the method of pairwise comparisons, *Biometrika,* 31, 324–345.
19. Knezek, G., Wallace, S., Dunn-Rankin, P. (1998): Accuracy of Kendall's Chi-square Approximation to Circular Triad Distributions, *Psychometrika,* 63(1), 23–34.
20. Kovalchik, S. A. (2016): Searching for the GOAT of tennis win prediction, *Journal of Quantitative Analysis in Sports,* 12(3), 127–138.
21. Kovalchik, S. A., Bane, M. K., Reid, M. (2017): Getting to the top: an analysis of 25 years of career rankings trajectories for professional women's tennis, *Journal of Sports Sciences,* 35(9), 1904–1910, DOI: 10.1080/02640414.2016.1241419
22. Kovalchik, S. A., Reid, M. (2018): A calibration method with dynamic updates for within match forecasting of wins in tennis, *International Journal of Forecasting,* 35(2), 756–766.
23. Kułakowski, K. (2018): Inconsistency in the ordinal pairwise comparisons method with and without ties, *European Journal of Operational Research,* 270(1), 314–327. DOI: 10.1016/j.ejor.2018.03.024.
24. Lisi, F., Zanella, G. (2017): Tennis betting: can statistics beat bookmakers? *Electronic Journal of Applied Statistical Analysis,* 10(3), 790–808. DOI: 10.1285/i20705948v10n3p790

25. Moran, P. A. P. (1947): On the method of paired comparisons, *Biometrika,* 34, 363–365.
26. Orbán-Mihálykó, E., Mihálykó, Cs., Koltay, L. (2019): A generalization of the Thurstone method for multiple choice and incomplete paired comparisons, *Central European Journal of Operations Research,* 27, 133–159.
27. Ramón, N., Ruiz, J. L., Sirvent, I. (2012): Common sets of weights as summaries of DEA profiles of weights: With an application to the ranking of professional tennis players, *Expert Systems with Applications,* 39(5), 4882–4889, DOI: 10.1016/j.eswa.2011.10.004
28. Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega,* 53, 49–57.
29. Saaty, T. (1977): A scaling method for priorities in hierarchical structures, *Journal of Mathematical Psychology,* 15, 234–281.
30. Shiraishi, S., Obata, T., Daigo, M. (1998): Properties of a positive reciprocal matrix and their application to AHP. *Journal of the Operations Research Society of Japan,* 41(3), 404–414.
31. Temesi, J. (2017): Páros összehasonlítási mátrixok elemeinek interaktív meghatározása verbális skála esetén, *Szigma,* 58(3-4), 111–131.
32. Temesi, J., Csató, L., Bozóki, S. (2012). Mai és régi idők tenisze – A nem teljesen kitöltött páros összehasonlítás mátrixok egy alkalmazása. In T. Solymosi, J. Temesi (szerk.) *Egyensúly és optimum. Tanulmányok Forgó Ferenc 70. születésnapjára,* Aula Kiadó, Budapest, 213–245.
33. Vaughan-Williams, L., Liu, Ch., Gerrard, H. (2019): How well do Elo-based ratings predict professional tennis matches? Nottingham Trent University, Discussion Papers in Economics, No. 2019/3, 20 p.
34. Watts, D. J., Strogatz, S. H. (1998): Collective dynamics of 'small-world' networks, *Nature,* 393, 440–442.
35. Zhou, Y., Wang, R, Zhang, Y-C., Zeng, A., Medo, M. (2020): Limits of PageRank-based ranking methods in sports data, Cornell ArXiv, https://arxiv.org/abs/2012.06366

INCOMPLETE PAIRWISE COMPARISON MATRICES: RANKING TOP WOMEN TENNIS PLAYERS

The method of pairwise comparisons is frequently used for ranking purposes. The aim of the paper is to apply the method for ranking top women tennis players with the help of their win/lose ratios. Incomplete pairwise comparison matrices can be constructed from the data obtained from WTA (Women's Tennis Association) homepage. The database contains head-to-head results from the period between 1973 and 2020 for those 27 players who had the position No. 1 in the official rankings of WTA for shorter or longer periods. Their names and the length of their professional tennis carrier can be seen in Figure 1, the W/L ratios are in Table 1. If two players have not met on the court, the element of the matrix is unknown (empty cell). Our goal is to create historical ranking including players even from disjoined time periods as we did with men top tennis players in Bozóki, Csató, Temesi (2016).

A great many of studies analysed sport results. Statistical analysis of performance data in tennis can be done for various purposes. Some articles use the data

for forecasting certain results of sport events (Kovalchik, 2016). Lisi and Zanella (2017), for instance, estimate the probability of wins with a logistic regression model. Their example is the Grand Slam championships' results in 2013. A special approach for creating good forecasts applies Elo-model (Elo, 1978). Vaughan-Williams et al. (2019) confirm the fitness of the model for Wimbledon 2018 results especially for top women players. Gu and Saaty (2019) apply descriptive indicators (e.g., age, right- or left-handedness, ranking position) and performance indicators (e.g., number of aces, winning or losing service games, winning or saving break-points). Their Analytic Network Process model is based on the factor analysis of the key indicators; it was tested on the results of US Open 2015. They reported very good fitness with the real results (85%) compared to the usual forecasts (70%). Ramon et al. (2012) used similar data, but they applied Data Envelopment Analysis to rank tennis players. Baker and McHale (2014) used Grand Slam data of more than four decades to estimate the power value of tennis players with the probabilistic dynamic model of paired comparisons. Their final rankings for women and men players gave similar results to our rankings. Orbán-Mihálykó et al. (2019) applied WTA Head-to-Head results (as we did) in order to rank women tennis players using the Thurstone model to estimate parameters with the maximum likelihood method. Their ranking is similar to our results, too.

Our analysis used the pairwise comparison method. If $P_1, P_2, \dots, P_n$ are the players, then we can build a pairwise comparison matrix $\mathbf{P} = [p_{ij}]$, $i, j = 1, \dots, n$ with the properties: (i) $p_{ij} > 0$, $i, j = 1, 2, \dots, n$ (positivity); (ii) $p_{ji} = 1/p_{ij}$, $i, j = 1, 2, \dots, n$ (reciprocity), thus $p_{ii} = 1$, $i = 1, 2, \dots, n$ (homogeneity). Meaning of p_{ij} element of the matrix is that „how much stronger/better/preferred P_i than P_j"? Our data can be described as follows: z_{ij} $(i, j = 1, \dots, n, i \neq j)$: the number of matches played between P_i and P_j, $(z_{ij} = z_{ji})$; x_{ij} $(i \neq j)$: the number of matches between P_i and P_j where the winner was P_i; $y_{ij} = z_{ij} - x_{ij}$ $(i \neq j)$: the number of matches between P_i and P_j, where P_i was the loser.

The incomplete PCM matrix has $n \times n$ elements. Let p_{ij} denote the W/L ratio between P_i and P_j, then the elements of matrix $\mathbf{P}$ are

$$p_{ij} = \begin{cases} x_{ij}/y_{ij}, & \text{if } i > j \text{ and } x_{ij} \neq 0,\ y_{ij} \neq 0; \\ y_{ji}/x_{ji} = 1/p_{ji}, & \text{if } i < j \text{ and } x_{ij} \neq 0,\ y_{ij} \neq 0; \\ 1, & \text{if } i = j, \\ \text{missing} & \text{otherwise.} \end{cases}$$

We made data correction as shown in (3) to avoid the case of 0 as denominator in W/L ratios. Moreover, as it can be seen from Table 1, the range of z_{ij} elements is large, therefore the range of the values of p_{ij} is large, too. We have used the transformation (4) to extract the range, where the transforming exponent z_{ij}/K is the ratio of the number of matches played between the given two players and the maximum of the number of matches played between any two players. The initial matrix of the calculations is matrix $\mathbf{T}$ with elements t_{ij}. The weight vector calculated from matrix $\mathbf{T}$ will give the ranking of the players. The most frequently used techniques to estimate the weight vectors are the eigenvector method (SV) (Saaty 1977), and the logarithmic least squares method (LLNM) (Crawford, Williams 1985):

a) SV: $$\mathbf{P}\mathbf{w} = \lambda_{\max} \cdot \mathbf{w},$$

b) LLNM: $$\sum_{i=1}^{n} \sum_{j=1}^{n} \left(\ln p_{ij} - \ln \frac{w_i}{w_j} \right)^2 \underset{\mathbf{w}}{\rightarrow} \min .$$

where $\mathbf{w}$ is the estimated weight vector with elements w_i $(i = 1, \dots, n)$, and $\lambda_{\max}$ is the dominant eigenvalue of $\mathbf{T}$.

Our initial data can be represented by an undirected graph as it is illustrated in Figure 2. Vertices are the players, edges denote that at least one match has been completed by the connected players. An important property of the graph is that it is connected. Bozóki, Fülöp, Rónyai (2010) proved, that the SV and LLNM methods extended for incomplete PCMs give exactly one solution in case of the connectedness of the representing graph. The weight vector of the women players' incomplete PCM gave the ranking of the first column in Table 2. The result is not surprising: Serena Williams, Graf and Navratilova stand in the first three positions, Hingis, Clijsters and Henin follow them. The new generation is represented by Barty in the 10th place. The second column of the table demonstrates that the W/L rates prove to be a good proxy of our ranking. The Spearman rank correlation value is 0.957. The next step of our calculations was to analyse submatrices of the initial matrix. How do subrankings behave? PCM1 column of Table 3 shows a ranking without the first eight players of the total ranking. PCM2 is a ranking without the last eight players of the total ranking. PCM3 is the ranking of the four great stars of the seventies. PCM4 is a ranking of 12 players from the next era. Finally, in PCM5 there is a ranking of 19 players selected randomly from our pool of women players. There are two important conclusions from these calculations. The first one is that – as we expected – changes in the set of players changed the rankings. However, these changes did not blow over the original rankings entirely, the new positions could be explained with the new patterns of the modified PCM. When we made the same kind of analysis with men players (Table 5), similar conclusions were drawn, demonstrating the robustness of our data system – robustness of rankings based on pairwise comparisons.

Using graph representation gave us the possibility to have a deeper insight to the properties of the incomplete pairwise comparison matrices, as it can be seen in the table below.

	WTA	ATP	WTAmod	ATPmod
Number of vertices (players)	27	26	23	21
Number of edges (players compared)	184	165	169	136
Minimal vertex degree	3	4	9	7
Maximal vertex degree	22	20	22	20
Diameter	4	4	2	2
Average shortest path	1.63	1.63	1.33	1.35
Edge connectivity	3	4	9	7
Vertex connectivity	3	4	8	7

The representing graphs have similar connectivity indicators. The longest shortest path (diameter) is 4 in both cases, and it can be determined between players far from each other in time. Figure 5 illustrates the degree of vertices in both graphs. That distribution is used in network theory (Albert, Jeong, Barabási 2000) for analysing various types of systems. In our case the average shortest paths have small values in both graphs, and the degrees are distributed relatively orderly around the average, clustering values are 0.8 for men, and 0.79 for women (certain groups of players had results from the same group). We can state that our networks are small-world type ones (Watts, Strogatz 1998).

Another line of our research referred to the nontransitive triads. In order to apply the known tests we had to correct the data with the help of the set rates to remove ties. The χ^2 test value is 185 for women and 169 for men, the corresponding p values are practically 0s. The conclusion is that the number of ordinally nontransitive triads does not have an effect on our results.