Knowledge-Based Systems 241 (2022) 108168

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Improving PageRank using sports results modeling

Yuhao Zhou [a], Ruijie Wang [a], Yi-Cheng Zhang [a], An Zeng [a,b], Matúš Medo [c,d,a,*]

[a] *Department of Physics, University of Fribourg, Fribourg 1700, Switzerland*
[b] *School of Systems Science, Beijing Normal University, Beijing 100875, PR China*
[c] *Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu 610054, PR China*
[d] *Department of Radiation Oncology, Inselspital, University Hospital of Bern, and University of Bern, Bern 3010, Switzerland*

A R T I C L E I N F O

Article history:
Received 16 September 2021
Received in revised form 1 December 2021
Accepted 4 January 2022
Available online 17 January 2022

Keywords:
Networks
Modeling sports results
Ranking in sport
Stochastic modeling
PageRank

A B S T R A C T

How to rank participants of a sports tournament is of fundamental importance. While PageRank has been extensively used for this task, the algorithm's superiority over simpler ranking methods has never been clearly demonstrated. We address this knowledge gap by comparing the performance of multiple ranking methods on synthetic datasets where the true ranking is known and the methods' performance can be thus quantified by standard information filtering metrics. Using sports results from 18 major leagues, we calibrate a state-of-art model, a variation of the classical Bradley–Terry model, for synthetic sports results. We identify the relevant range of parameters under which the model reproduces statistical patterns found in the analyzed empirical datasets. Our evaluation of ranking methods on the synthetic datasets shows that PageRank outperforms the benchmark ranking by the number of wins only early in a tournament when a small fraction of all games have been played yet. Increased randomness in the data due to home team advantage, for example, further reduces the range of PageRank's superiority. We propose a new PageRank variant that combines forward and backward propagation on the directed network representing the input sports results. The new method outperforms PageRank in all evaluated settings and, when the fraction of games played is sufficiently small and the sport is not too random, it outperforms also the ranking by the number of wins. Beyond the presented comparison of ranking methods, our work paves the way for designing optimal ranking algorithms for sports results data.

1. Introduction

Humans are born to compete: We thrive by measuring ourselves against the others [1,2]. Sport in particular provides ample opportunities for competition with high significance to the economy [3,4] and the society [5,6]. Once a sport competition is over, it remains to decide who has won. What is a simple question for a single match between two participants (individuals or teams) becomes a difficult one for a structured tournament with multiple games between several participants. The design of sport tournaments is therefore of crucial importance. Effective tournament design helps the participants to perform well during the tournament, produces match-ups that are interesting to the fans and, crucially, allows to identify the best-performing participant with high probability [7] (see [8] for a survey of results on sport tournaments and beyond).

We focus here on sports leagues, such as a soccer league, for example, where each participating team plays against every other team once or several times. In such a league, points are traditionally assigned to teams for every win or tie that they achieve. The teams are then ranked by their point totals (from the highest to the lowest; additional criteria can be used to break ties). This benchmark ranking method is so simple that it inevitably prompts the question: Can we rank the teams better? Of particular appeal is the idea to consider the strength of the opposing team. In particular: Can we improve the ranking if a win against a strong team counts more than a win against a weak team? A closely related idea has been previously formalized by the PageRank algorithm which has been originally designed to rank web pages and later used in a broad range of systems [9] and found applications in other domains such as community detection [10], for example. Unsurprisingly, PageRank and its various modifications have been also applied to results from various sports: Tennis [11,12], mixed martial arts [12], international football [13], cricket [14], wrestling [15], and boxing [16], for example.

* Corresponding author at: Department of Physics, University of Fribourg, Fribourg 1700, Switzerland.
E-mail addresses: yuhao.zhou@unifr.ch (Y. Zhou), ruijie.wang@unifr.ch (R. Wang), yi-cheng.zhang@unifr.ch (Y.-C. Zhang), anzeng@bnu.edu.cn (A. Zeng), matus.medo@unifr.ch (M. Medo).

https://doi.org/10.1016/j.knosys.2022.108168

While PageRank is frequently used on sports results data, its superiority over simple point-based schemes has not been established yet. That is not surprising as empirical sports results lack robust ground truth (*i.e.*, a correct team ranking) against which different rankings could be compared. We fill this gap by setting up a model for synthetic sports results. This model is calibrated using empirical sports results and we show that it closely matches the empirical winning probabilities. The ground truth is known for synthetic data which allows us to comprehensively evaluate three distinct rankings methods (points ranking, PageRank, and a new PageRank variant) on the synthetic datasets.

1.1. Research hypotheses and contributions

Applying PageRank on sports results data has a simple and understandable motivation: A win against a strong opponent should intuitively carry more weight than a win against a weak opponent. This idea, together with a self-consistent definition of what "strong" and "weak" means, is behind the PageRank score [17] where a link from a highly-valued node increases the target node's reputation more than a link from a little-valued node. As PageRank is a network algorithm, sports results need to be first represented with a network where each result is represented with one link. In particular, the loss of team i against team j is represented by an edge from i to j. Once PageRank is applied on the resulting network, this edge causes some part of i's score to contribute to the score of team j (see Section 2.3 for details). If team i has many wins, it has high score and team j is thus rewarded more for beating it, in line with our opening motivation.

At the same time, the necessary condition for PageRank to work well can be formulated in mathematical terms. As shown in [18], PageRank score of a node is on average proportional to its indegree when degree–degree correlations are absent in the network. As a consequence, PageRank then carries no more information than indegree. Of particular interest are positive indegree correlations (commonly referred to as indegree assortativity) whose presence is an indication that weak teams tend to win against weak opponents whilst strong teams tend to win against weak *and* strong opponents. In such a situation, to value a win against a strong opponent more can indeed help to better identify the best teams. This leads us to the first hypothesis which aims to verify whether the use of PageRank on sports results data is justified:

H1: Upon representing sports results with directed links, win of team i over team j as a directed edge from j to i, node indegree is assortative, *i.e.*, it is positively correlated among neighboring nodes in the sports results network.

However, an assortative indegree pattern alone is not a sufficient condition for PageRank to outperform the benchmark ranking by the number of wins. To this end, we set up a model to generate synthetic sport results and calibrate it on a broad range of sports results data. This allows to probe our second hypothesis:

H2: PageRank outperforms the benchmark ranking by the number of wins on synthetic sports results data.

Our main contributions can be summarized as follows. We test degree assortativity patterns in empirical sports results data. We set up a state-of-art model for synthetic data, show that it accurately matches empirical sports results, and establish the range of model parameters that corresponds to the analyzed empirical datasets. We establish, for the first time, the conditions that have to be met for PageRank to perform better than the benchmark ranking by the number of wins. We propose a novel PageRank variant that outperforms PageRank on synthetic data in all evaluated settings. We reproduce our key result on empirical data, thus showing that our main findings are not model-dependent. Beyond the presented results, our work establishes a robust framework for assessing ranking methods for sports results in the future.

1.2. Related work

Modeling sports results. Results in a sport where players or teams play against each other can be seen as outcomes of paired comparisons between the participants. The Bradley–Terry model [19] is one of the first efforts to model outcomes of such paired comparisons (the authors themselves do not mention applying their model on sport specifically). The model is based on assigning a non-negative winning propensity, π, to each participant and postulating that the probability of i winning against j in the form $\pi_i/(\pi_i + \pi_j)$. A model generalization to include ties (which is important for sports such as soccer) has been introduced in [20]. The Bradley–Terry model was used in a broad range of problems. In [21], for example, the authors used the model to show that sports data have a high degree of randomness. In [22], the winning propensities π of tennis players were shown to be directly proportional to players' ranks in official tennis rankings. The rank of individuals instead of a continuous-valued winning propensity was used also in [23].

In [24], the authors propose a simpler model based on a fixed upset probability parameter that directly specifies the probability that a weaker team beats a stronger team. The model's simplicity makes it possible to derive analytical results for the probability that the weakest team wins an elimination tournament, for example. O'Malley [25] goes in the opposite direction by proposing a model for tennis match outcomes based on the detailed structure of the game. See [26] for detailed reviews of models for paired comparison data.

Ranking algorithms for sports results data. The most elementary method to aggregate results of multiple sports games is to compute each team's winning percentage, $n_w/(n_w + n_l)$, where n_w and n_l are the team's numbers of wins and losses, respectively. In this way, each team is assigned a quantity in the range [0, 1]. Taking into account a starting uniform prior in the same range, a modified estimate has the form $(1 + n_w)/(n_w + n_l + 2)$. In [27], this estimate was used as a basis for a ranking scheme which is particularly useful when many teams have not played against each other (early in a season or in a more complicated setup where teams are divided in multiple divisions or conferences). The information provided by respective wins and losses is limited in low-scoring sports such as football where a single lucky shot can greatly influence the outcome of a match. In [28], the authors propose to limit this randomness by estimating the number of expected goals.

Indirect comparisons (comparing teams A and C based on team A beating team B and team B beating team C) have been considered in [29]. Indirect wins and indirect losses have been quantified in [30] where they have been ultimately combined in a score which is, in fact, a generalization of the well-known Katz centrality metric.

A popular line of research considers the use of eigenvector-based methods for sports rankings [31]. In particular PageRank, a seminal ranking algorithm/centrality metric for nodes in a directed network [17], has been widely applied to sports data such as tennis [11] or cricket [14,32]. PageRank-like algorithms seem well-suited for a sports ranking as they value a win against a strong opponent more than a win against a weak opponent. How to transform input sports results in a directed network on which PageRank is computed is open. The simplest approach is to represent a win of i over j with a directed link from j to i.

In [33], the authors proposed to assign link weights based on the score difference in the corresponding games. A comparison of various edge-weighting methods for the football World Cup data is presented in [13]. Time-aware PageRank variants [12, 34] take the time of each game into account to capture the player/team capability that varies in time. In [35], PageRank score of teams was used together with other features in a machine-learning framework for predicting football results. See [36] for a recent work where several distinct ranking approaches are used to assess the performance of track athletes.

2. Materials and methods

2.1. Network theory methods

To create the directed network on which PageRank is applied, results of available games are represented with directed links. In particular, when team/player i wins against team/player j, a directed link from j to i is formed and its weight is set to one. For each additional win of i over j, the link weight increases by one. The weight of link from j to i, e_{ji}, is thus the number of wins of i over j. The in-strength of node i, $s_i^{in} = \sum_j e_{ji}$, is a generalization of node indegree for weighed networks [37]; in our case it corresponds to the total number of wins of team/player i. Analogously, the out-strength of node i, $s_i^{out} = \sum_j e_{ij}$, is a generalization of node outdegree and corresponds to the total number of losses of team/player i. Note that $s_i^{out}+s_i^{in} = k_i$ where k_i is the number of games of team/player i and $\sum_i s_i^{in} = \sum_i s_i^{out} = G$ where G is the number of games that are used to construct the network. The number of nodes in the network, N, is given by the number of teams/players in the used games.

To assess the strength of degree correlations, we adopt the methodology of Ren et al. [38] which is based on assortativity plots where the average in-strength of neighbors of a node is plotted as a function of the node's in-degree. We use upstream neighbors (nodes that point to a given node) in particular; results are quantitatively similar when downstream neighbors are considered. For a given node i, we measure the average in-strength of its upstream neighbors,

$$\overline{s_i^{in,up}} = \sum_j e_{ji} s_j^{in} / \sum_j e_{ji}. \tag{1}$$

The level of indegree assortativity can be quantified using least-squares linear regression of $\overline{s_i^{in,up}}$ against s_i^{in} in the assortativity plot. The sign of the slope then determines if indegree correlation is assortative (positive slope) or disassortative (negative slope). The p-value of the regression decides if the identified pattern is significant or not. As in [38] we compare the result with the result obtained on a randomized network where links are randomized whilst preserving node out- and in-degree values [this is referred to as the configuration network model in the literature [37]].

The assortativity plot for one season of a sports league comprises N data points corresponding to N teams/players participating in the season. To increase the statistical power of our assortativity analysis, we combine all seasons of one league in one assortativity plot. When N varies between the seasons, the in-strength values are not directly comparable between the seasons. We amend this by normalizing the in-strength by the number of games played by the team, thus obtaining the normalized in-strength, $w_i := s_i^{in}/k_i$. As s_i^{in} is the number of games won by team/player i, w_i is the win ratio of team/player i. In a direct analogy with Eq. (1),

$$\overline{w_i^{up}} = \sum_j e_{ji} w_j / \sum_j e_{ji} \tag{2}$$

is the average win ratio of upstream neighbors of node i, which we use to construct the assortativity plot.

2.2. Sports results model

We assume a competition setting where N teams play against each other once or multiple times. In the model, we assume that the outcome of a match between teams i (home team) and j (away team) is stochastic. The probability that the home team wins, $i \succ j$, is assumed in the form of the logistic function

$$P(i \succ j) = 1/\left[1 + e^{-(f_i - f_j + H)/\delta}\right] \tag{3}$$

where f_i and f_j are the intrinsic fitness values of the two competing teams, H is an additive term which represents the typical home team advantage and δ is a fitness "weighting" parameter which helps to translate a difference in team fitness in the winning probability of the home team. The model assumes that there are only two possible outcomes: Team i wins or team j wins. This choice is motivated by the absence of ties in three of the four analyzed sports (baseball, ice hockey, and basketball; in the football data, ties account for roughly 25% of all results). The appendix shows that when ties are introduced in the synthetic data, our main conclusions still hold.

We assume for simplicity that team fitness remains the same throughout the whole competition; allowing for fitness variations is yet another interesting direction for future research. Home advantage has been documented for a wide variety of sports [39,40]. While it may seem as an auxiliary issue that has been ignored in [22], for example, we find H to be significantly positive in a vast majority of the sports results sets that we analyze in this paper. We also find that home advantage strongly affects the ranking ability of respective algorithms.

It is helpful to study closer the implications Eq. (3) before proceeding. When $(f_j - f_i)/\delta \gg 1$ (that is, the away team is much stronger than the home team), we get $P(i \succ j) \ll 1$ as expected. If δ increases, the same fitness difference affects the winning probability $P(i \succ j)$ less and the match outcomes thus become more random (in the limit $\delta \to \infty$, $P(i \succ j) = 1/2$ for any f_i, f_j, and H). We thus refer δ as the sport randomness parameter: When δ is large in comparison with fitness differences among the teams, $P(i \succ j) \approx 1/2$ for all i and j. That δ alone is not a measure of sport randomness can be seen from the invariance of Eq. (3) to simultaneously multiplying f_i, H, δ with λ; the value of δ is indeed important only in relation to the fitness values. The effective sport randomness can be thus measured by $\delta/\sigma(f)$ where $\sigma(f)$ is the standard deviation of the fitness values. Similarly, while home team advantage increases with H, outcomes are random in the large δ limit and no home team advantage ensues. The effective strength of home team advantage is thus determined by the ratio H/δ.

2.3. Algorithm to generate synthetic sports results

The algorithm has the following parameters: The number of teams N, fitness sensitivity δ, home advantage H, and the fraction of games that have been played P. $P = 1$ corresponds to $N(N-1)/2$ games played—all teams playing once against each other. It is also possible to consider $P > 1$: That would correspond to a league where the teams play more than once against each other.

Synthetic data are then created in three main steps:

1. Fitness of team i is set to $f_i = (i - 0.5)/N$ where $i = 1, \dots, N$. In this way, the fitness values range from $0.5/N$ to $1 - 0.5/N$ and they are regularly distributed in the range [0, 1] (we investigate other fitness distributions later).

2. Each team is assigned to play $P(N-1)$ games against opponents chosen at random without any two teams playing against each other more than once (in practice, we used the random_degree_sequence_graph function from the package NetworkX in Python). If $P(N-1)$ is not an integer, teams are assigned to play either $\lfloor P(N-1)\rfloor$ or $\lceil P(N-1)\rceil$ games in such a way that the total number of played games is $PN(N-1)/2$. The home team is chosen at random for each game.
3. Determine the outcome of each game by Eq. (3).

By varying the model parameters, we can create synthetic results corresponding to a broad range of sports. Note that the algorithm can be easily modified to encompass more complicated settings such as the regular season followed by playoffs, for example.

2.4. Ranking algorithms

In sports leagues, teams are typically ranked by the number of points that they obtain (such as two points for a win, one point for a draw, zero points for a loss). Since we focus here on sports where draws are not possible, we directly count the number of wins by each team and rank the teams by their win ratio (a team's number of wins divided by its number of games played). Due to its simplicity, WinRatio is our benchmark method.

To apply the PageRank algorithm, we create a directed network of participating teams where all games are represented with directed links. In particular, when player/team i wins against player/team j, a directed link from j to i is formed along which "sports prestige" flows: A win against a highly-valued team contributes highly to the winner's own evaluation. The process can be mathematically represented by the formula [11]

$$P_i = (1-\alpha)\sum_{j:\, s_j^{out}>0} P_j \frac{e_{ji}}{s_j^{out}} + \frac{\alpha}{N} + \frac{1-\alpha}{N}\sum_{j:\, s_j^{out}=0} P_j \tag{4}$$

where P_i is the prestige score of team/player/node i, N is the number of nodes in the network, e_{ji} is weight of the link from j to i which is equal to the number of wins of i over j, $s_j^{out}=\sum_i e_{ji}$ is the out-strength of node j (the number of losses of j) and α is the algorithm parameter (often referred to as the teleportation probability). The last term in Eq. (4) makes the algorithm robust against the nodes with $s_j^{out}=0$ ("dangling nodes") that would otherwise act as score sinks. In line with [11] and other PageRank literature, we use $\alpha = 0.15$.

In addition to standard PageRank, we consider here a new method closely based on PageRank which we refer to as bi-directional PageRank (BiPageRank). The bi-directional PageRank score, S_i, is defined as

$$S_i = P_i - Q_i \tag{5}$$

where P_i is the previously introduced PageRank score (computed on the original directed network) and Q_i is given by

$$Q_i = (1-\alpha)\sum_{j:\, s_j^{in}>0} Q_j \frac{e_{ij}}{s_j^{in}} + \frac{\alpha}{N} + \frac{1-\alpha}{N}\sum_{j:\, s_j^{in}=0} Q_j \tag{6}$$

In agreement with the previous definition of e_{ji}, e_{ij} is the number of losses of i against j and $s_j^{in}=\sum_i e_{ij}$ is the in-strength of j (the number of wins of j). In this way, both the winner of a match is assigned a part of the losing team's score (through P_i) as well as the team that loses is assigned a part of the winning team's negative score (through Q_i). Bi-directional PageRank is then a simple combination of the two scores. The motivation for this modification is straightforward: While P_i allows us to award team i "positive" score based on which teams it won against, Q_i allows us to award team i "negative" score based on which teams it lost against. As a practical illustration, take teams i and j that lost all their matches so far, hence their PageRank scores are the same. If team i lost against good teams (teams that lose rarely) and team j lost against bad teams (teams that lose often), then $S_i > S_j$. The new algorithm thus allows us to distinguish the two teams. Note that separate win and loss scores have been considered also in [30].

2.5. Evaluation metrics

On synthetic data, the resulting ranking of teams produced by a ranking algorithm can be directly compared with their fitness values which are thus used as the ground truth. The computed ranking of teams that is closest to the ranking of teams by their intrinsic fitness, on average, is then considered to be the best one. Intrinsic fitness is a natural choice of the ground truth as it is a hidden variable that affects the game results. Of course, a team with low fitness can be lucky and beat a team with high fitness. In the long run, however, the lucky and unlucky outcomes tend to compensate each other, so it is indeed the fitness that we are interested in.

Denote the computed and ground-truth ranking of team i as x_i and g_i, respectively. We use the following distinct metrics to quantify the ranking performance of an algorithm:

1. The Kendall correlation coefficient, τ, is defined as

$$\tau(\boldsymbol{x},\boldsymbol{g}) = \frac{|\{(i,j):\ (x_i-x_j)(g_i-g_j)>0\}| - |\{(i,j):\ (x_i-x_j)(g_i-g_j)<0\}|}{\frac{1}{2}n(n-1)} \tag{7}$$

where $\boldsymbol{x}$ and $\boldsymbol{g}$ are vectors of the computed rankings and the ground-truth rankings, respectively. Kendall's τ ranges from $+1$ when rankings $\boldsymbol{x}$ and $\boldsymbol{g}$ are identical to -1 when one ranking is the reverse of the other. Note that tied ranking positions in the computed ranking (the ground-truth ranking has no ties by construction) do not contribute to τ. A degenerate ranking that would assign the same rank to all teams would thus achieve $\tau = 0$.
2. While Kendall's τ takes all teams into account, the other two metrics explicitly focus on how well the top teams are ranked. The first metric is the average computed ranking of the top 5 ground-truth teams. The smaller the value, the better the computed ranking. The best value, $(1+2+3+4+5)/5 = 3$, is achieved when the top 5 ground-truth teams are in the top five positions of the computed ranking (not necessarily in the right order).
3. The second metric is the area under the ROC-curve, commonly referred to as *AUC* in the statistics literature. We again use the top 5 ground-truth teams as our target set; the other teams constitute the ordinary set. To compute *AUC*, we use the probabilistic approach [41] where we pick n pairs of teams, one from the target set and the other from the ordinary set. If the target team is ranked higher than the ordinary team n' times and tied n'' times, the *AUC* value can be computed as $AUC = (n' + n''/2)/n$.

2.6. Empirical sports results data

We analyze here sports results data obtained from websites http://www.win007.com/ and https://www.sports-reference.com/. Except for soccer, all games have only two possible outcomes: The home team wins or the away team wins. As we consider an outcome model without draws, all draws (20%–25% of games in one season for all six considered leagues) are ignored. In one season, the participating teams play against each other two or

Table 1
Basic description of the analyzed sports results sets. The league size is the number of competing teams (a range is provided if the number varies in the considered period).

Sport	Country	League Name	Label	Size	Years
Baseball	USA	Major League Baseball – American League	AL	14–18	1997–2016
	USA	Major League Baseball – National League	NL	15–30	1997–2016
	Japan	Nippon Professional Baseball	NPB	12–14	2009–2018
	Mexico	Ligue Mexicaine de Baseball	LMB	14–18	2007–2019
Ice hockey	USA	National Hockey League	NHL	28–31	2000–2019
	Swizerland	Ligue Nationale A	LNA	12–25	2009–2020
	Germany	Deutsche Eishockey Liga	DEL	14–16	2007–2020
Soccer	Germany	Deutsche Futball Liga	Bundesliga	18	2000–2019
	Italy	Lega Serie A	Serie A	20	2005–2019
	Spain	Primera division de Liga	La Liga	20	1998–2017
	England	England Premier League	EPL	20	1999–2018
	USA	Major League Soccer	MLS	10–24	2000–2019
	France	Championnat de France de football Ligue 1	Ligue 1	18–20	2000–2019
	China	Chinese Football Association	CSL	12–16	2004–2019
Basketball	Italy	Lega Basket Serie A	LBSA	15–18	2008–2019
	China	Chinese Basketball Association	CBA	16–20	2007–2017
	Spain	Asociacion de Clubes de Baloncesto	ACB	17–18	2007–2019
	USA	National Basketball Association	NBA	29–30	2001–2020

more times. We analyze only results from regular seasons, playoff matches are ignored. See Table 1 for an overview of the analyzed datasets and their basic statistical properties.

3. Results

3.1. Analysis of empirical datasets

Before assessing performance of respective ranking methods, we calibrate the synthetic model in the following section and address the degree assortativity hypothesis H1 in the next one.

Model calibration on sports results data. To determine realistic values of all $N+2$ model parameters (team fitness values, H, and δ), we use maximum likelihood estimation for sets of results from various sports. For a given set of games with outcomes, $\mathcal{G}$, and a single game $g\in\mathcal{G}$, we denote the home team as h_g, the away team as a_g, and the game result as R_g where $R_g=1$ means that the home team won game g and $R_g=0$ means that the home team lost. The data likelihood given the model then has the form

$$\mathcal{L}(\mathcal{G}|f_1,\dots,f_N,H,\delta)=\prod_{g\in\mathcal{G}}\Big\{R_gP(h_g,a_g)+(1-R_g)\big[1-P(h_g,a_g)\big]\Big\}. \tag{8}$$

Likelihood maximization for the empirical datasets described in Section 2.6 reveals that maximum likelihood estimates (MLE) of team fitness values, $\hat{f}_i$, are close to the fraction of wins, w_i, of each respective team in the analyzed dataset. In particular, the difference between the likelihood maximized over all $N+2$ model parameters and the likelihood maximized only over δ and H (with f_i replaced by w_i) is not sufficient to justify the higher number of parameters in the former model as judged by the Akaike information criterion for model selection [42]. This allows us to write the simplified winning probability of the home team as

$$P(i\succ j)=1/\big[1+\mathrm{e}^{-(\Delta w_{i,j}+H)/\delta}\big] \tag{9}$$

where $\Delta w_{i,j}:=w_i-w_j$ is the difference between the win ratios of the home and away teams.

When home advantage is absent ($H=0$), denoting $\exp[f_i/\delta]:=p_i$ and $\exp[f_j/\delta]:=p_j$ transforms Eq. (3) to $P(i\succ j)=p_i/(p_i+p_j)$ which is precisely the form assumed by the classical Bradley–Terry model [19]. We see now that the model formulation presented by Eq. (3) is still advantageous as: (1) Unlike the "winning propensities" p_i in the Bradley–Terry model, the team fitness values f_i in Eq. (3) directly correspond to the team win ratios, (2) H introduces home advantage in a scale that can be directly compared with the teams' win ratios and their differences ($H=0.1$ is as important as a 0.1 difference in win ratios between the teams).

Using four sample results sets, Fig. 1 shows a comparison between Eq. (9) with maximum likelihood estimates for H and δ and the empirical winning probability plotted as a function of the win ratio difference between the competing teams. The good agreement that can be observed in the whole range of win ratio difference confirms that Eq. (3), on which Eq. (9) is based, can model the empirical data well. The sigmoid curve's steepness in the figure is in a direct relation with the fitness sensitivity parameter δ (smaller δ yields higher steepness). Note also that $P(i\succ j)>0.5$ when $\Delta w_{i,j}=0$ which is a direct consequence of a positive home advantage in all four results sets.

The maximum likelihood estimates for δ and H can be used to compute the effective randomness of results, $\delta/\sigma(f)$, and the effective home advantage, H/δ. As we found that win ratios w approximate well fitness values f, we measure sport randomness directly with $\delta/\sigma(w)$. Fig. 2 summarizes the results with one panel for each sport. We see that the sports substantially differ in their level of randomness as characterized by $\delta/\sigma(w)$ (baseball and basketball are the most and the least random sport, respectively). Different leagues in the same sport have mostly similar $\delta/\sigma(w)$ values except for the CBA basketball league which is significantly less random than NBA and LBSA (in fact, CBA is the least random league on average among the analyzed 18 leagues). The home advantage value is distributed between 0 and 0.25, and the home advantage effect of basketball and football is more significant than baseball and hockey. In agreement with Eq. (3), the effective strength of the home advantage is characterized by H/δ which is shown in Fig. 2 on the vertical axes. The values of H/δ differ significantly between the sports with soccer and basketball showing considerably higher effective home advantage than ice hockey and baseball. CBA is again outstanding by having the highest average effective home advantage. By contrast, baseball leagues have average effective home advantage 5.4-times smaller than CBA.

Indegree assortativity in sports results data. We now turn to indegree assortativity. Following the methodology described in Section 2.1, we obtain one data point $(w_i,\overline{w_i^{up}})$ for each season and each team. Data points from all teams in all studied seasons

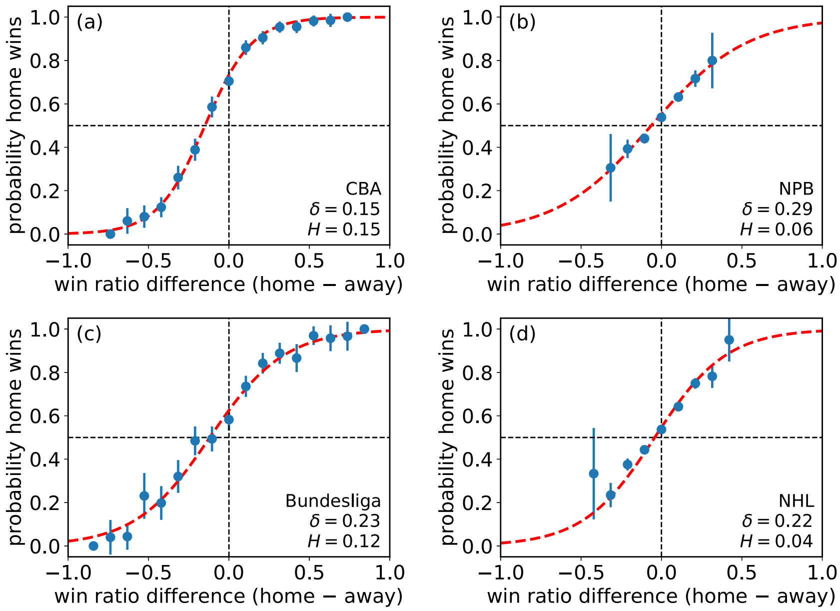

Fig. 1. Relationship between the win ratio difference and the home team winning probability for four distinct leagues. Symbols show empirical winning probabilities for a given end of season win ratio difference (data points based on less than four games have been omitted), error bars show double of the standard error of the mean, and lines show the model probability of winning given by Eq. (9) for the maximum likelihood estimates of δ and H (shown in each panel). The horizontal and vertical dashed lines show the zero win ratio difference and the baseline win probability of 1/2, respectively.

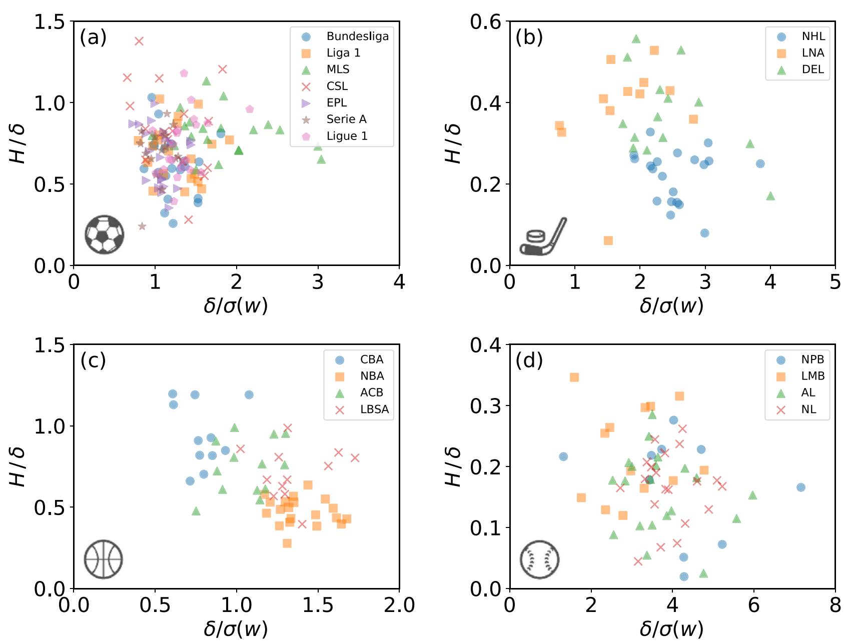

Fig. 2. Effective sport randomness, $\delta/\sigma(w)$, and effective home advantage, H/δ for the sports results sets from Table 1 (each panel shows a different sport). Each symbol is based on estimates of δ and H in a single season. The underlying maximum likelihood estimates of δ range from 0.13 to 0.31.

(we always use a complete season) of a league are then used as input for the linear regression between w_i and $\overline{w_i^{up}}$. Using the p-value threshold of 0.01, 7 leagues show significant indegree assortativity, 6 leagues show significant indegree disassortativity, and the remaining 5 leagues do not exhibit a significant indegree assortativity pattern. The leagues with an assortative pattern are: EPL, Bundesliga, Serie A, and Liga 1 (all four soccer) and CBA, NBA, and ACB (all three basketball). The strongest assortative pattern, as measured by the slope of linear regression between w_i and $\overline{w_i^{up}}$, is found for the CBA league which we previously found to be the least random. A similar relation can be found between high randomness of results and disassortativity: From ice hockey and baseball leagues, which show high effective randomness in Fig. 2, only NHL does not have a statistically significant disassortative pattern. We can thus conclude that our hypothesis H1 is confirmed for less random sports leagues where the average indegree of a node's neighbors indeed grows, on average, with the node indegree. However, indegree assortativity itself is no guarantee that PageRank will perform well. A direct assessment of ranking methods, which is the focus of the following section, is necessary

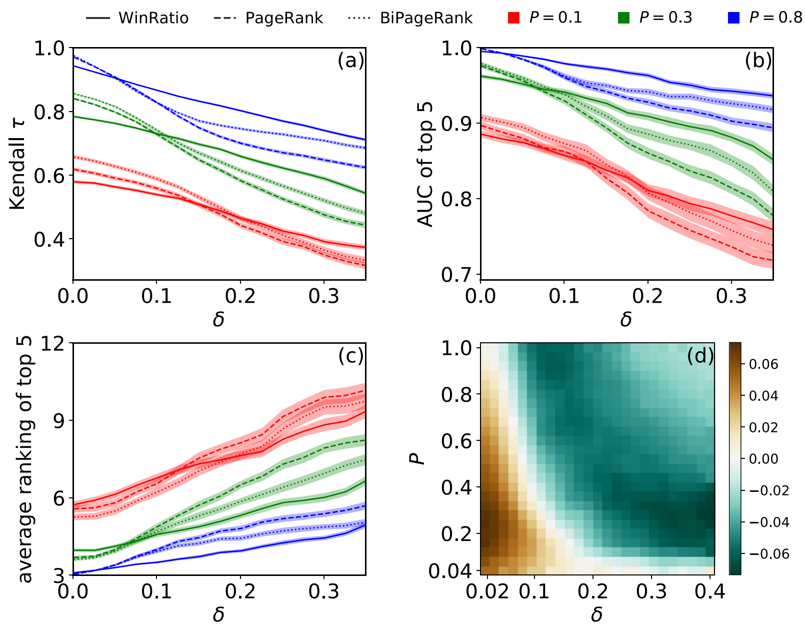

Fig. 3. Performance of the studied ranking algorithms on synthetic data without home advantage for 30 teams. Panels (a)–(c) use three different evaluation metrics and plot the results as a function of the fitness sensitivity parameter, δ, for various fractions of played matches, P. The lines represent mean results and the shaded regions indicate the standard error of the mean, all determined from 100 independently created synthetic datasets. Mean performance differs significantly between the algorithms (as judged by the Kolmogorov–Smirnov test) everywhere except for the regions where the shaded regions substantially overlap. Panel (d) shows the Kendall's τ difference between Bi-directional PageRank and the win ratio as a function of both δ and P.

to establish the conditions under which PageRank outperforms the benchmark ranking by the win ratio.

3.2. Ranking performance on synthetic datasets

Our goal now is to address our hypothesis H2 by evaluating the performance of different ranking algorithms on synthetic sports results generated by the algorithm described in Section 2.3. In simulations, we assume that there are 30 teams ($N = 30$); the results are robust with respect to this choice. We begin by studying the case of no home advantage ($H = 0$) and explore a range of δ values which corresponds to the maximum likelihood estimates of δ used in Fig. 2.

Fig. 3 shows the results of numeric simulations comparing the three considered ranking algorithms as a function of fitness sensitivity, δ, and the fraction of matches played, P. Panels a–c show that the comparison results are remarkably similar for the three evaluation metrics (Kendall's tau, *AUC* and average ranking). In particular, the results show that: (1) As P grows, the ranking performance improves as expected. (2) PageRank outperforms the win ratio only when δ is sufficiently small and the range of PageRank's superiority shrinks as P grows. The threshold δ below which PageRank outperforms the win ratio is considerably stable with respect to the number of teams, N (results not shown). (3) The newly-proposed Bi-directional PageRank is always an improvement (or a tie) over standard PageRank (according to the Kolmogorov–Smirnov, the improvements are statistically significant). (4) Fig. 2 shows that a vast majority of the analyzed datasets have $\delta > 0.15$ which together with Fig. 3 implies that the use PageRank brings no improvement in sport tournament rankings. Even more, PageRank is significantly inferior for sports with high randomness (high δ) later in a season. Finally, the heat map in Fig. 3(d) compares bi-directional PageRank and the win ratio using Kendall's τ for a broad range of the key parameters δ and P. We can see here well that a tie between the two ranking algorithms occurs at δ which progressively decreases as P grows. With respect to the estimates of δ presented in Fig. 2, we see that BiPageRank can outperform WinRatio for less random sports when $P \lesssim 0.2$.

Based on Fig. 3, we can conclude that PageRank and bi-directional PageRank are both more sensitive than the win ratio to increasing randomness of outcomes (represented by increasing δ). This increased sensitivity can be explained by the algorithms' network nature: While a "surprise" outcome of a single match has only a local impact on the win ratio (only the two competing teams are affected), PageRank propagates its scores further over the whole network of teams. When δ is sufficiently large, the surprising outcomes are numerous and their network propagation and accumulation are ultimately detrimental to the ranking ability of PageRank and Bi-directional PageRank. We can thus conclude that as hypothesis H1, hypothesis H2 is also confirmed only partially for sports that are sufficiently little random.

The top row of Fig. 4 evaluates the ranking performance of algorithms when the home advantage parameter, H, is positive. We see that as H grows, the performance of all three ranking algorithms deteriorates. At the same time, the win ratio is more robust to increasing H than the other two ranking methods, which is in line with its higher robustness to increasing δ. In particular, the number of unexpected results (a weaker team wins against a stronger team) increases as H grows and these unexpected results negatively affect the ranking results of PageRank and Bi-directional PageRank. The home advantage thus further reduces the limited range of applicability of PageRank (the range in which PageRank outperforms the win ratio).

In synthetic data so far, we assumed the team fitness values to be uniformly assumed. That this is not the case in empirical data can be easily illustrated as we have already shown that the win ratio is a good approximation for team fitness. Fig. 4d shows the win ratio in four different datasets and shows distinct non-linearity for two of them. This motivates us to consider a non-linear assignment of fitness in the form

$$f_i = \beta \left(\frac{i - 0.5}{N} \right)^{\alpha} + \gamma \tag{10}$$

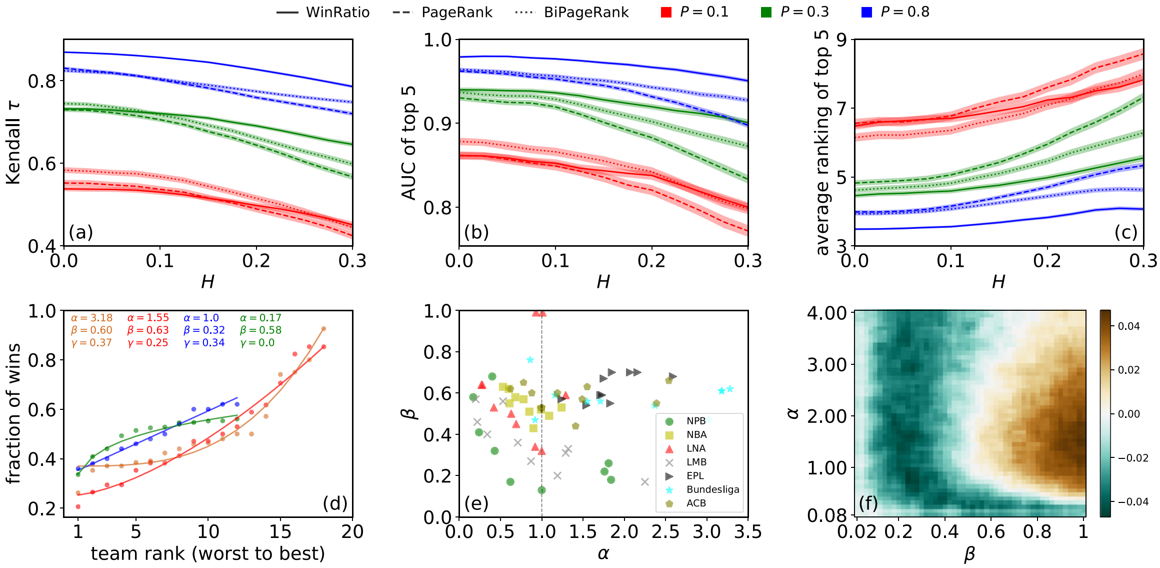

Fig. 4. Performance of the studied ranking algorithms on synthetic data with home advantage (top row) and with a non-uniform distribution of team fitness (bottom row). Panels (a)–(c) show the ranking performance vs. the home advantage parameter, H, for fixed $\delta = 0.1$ (all other parameters as in Fig. 3). (d) The relation between the fraction of wins and the team rank (from the worst to the best) for four chosen result sets (NHL, 2011; ACB, 2011; LMB, 2012; Bundesliga, 2010). Panel (e) Fits of Eq. (10) for all seasons of seven different leagues. (f) The difference $\tau(\text{BiPageRank}) - \tau(\text{WinRatio})$ for synthetic data with various values of α and β ($\delta = 0.1$, $H = 0$, and $P = 0.1$, results are averaged over 100 realizations).

which fits well most of the considered datasets (see least-squares fits in Fig. 4d). Note that a power-law fitness distribution has been considered before in [43]. In Eq. (10), α controls the heterogeneity of the fitness distribution and β determines the difference between the worst and the best team. Once α and β are chosen, γ is fixed by the relation $\sum_i f_i/N = 1/2$ (the average win ratio must be one half as someone's win is always someone else's loss). At the same time, the absolute term γ does not influence the fitness difference $f_i - f_j$ that determines the winning probability in Eq. (3); this difference depends only on α, β, i, and j.

Fig. 4e further shows the fitted values α and β for seven representative leagues and helps us identify $\alpha \in (0, 3.5)$ and $\beta \in (0, 1)$ as the relevant ranges for these two parameters. Fig. 4f shows the difference between the win ratio and bi-directional PageRank for synthetic data generated with α and β in the identified range. We choose here by purpose parameters that favor bi-directional PageRank: Small randomness ($\delta = 0.1$), no home advantage ($H = 0$) and few games played ($P = 0.1$). In agreement with the results presented in Fig. 3, BiPageRank outperforms WinRatio when $\alpha = 1$ and $\beta = 1$ (fitness values are then uniformly distributed in the range $[0, 1]$). We see now that this is essentially the ideal setup for BiPageRank as its advantage decreases when α substantially differs from 1 as well as when β is lower than 1. This is because the average fitness difference between the teams then decreases which, in agreement with Eq. (3), increases the probability of unexpected outcomes. Thus-introduced randomness is detrimental to the performance of PageRank and bi-directional PageRank which is well visible in Fig. 4e. When δ, H, or P increase, the behavior is similar, only the region where BiPageRank outperforms WinRatio shrinks.

In summary, we identified the sensitivity of PageRank and bi-directional PageRank to unexpected results as the main factor limiting their performance. Unexpected results are due to intrinsic randomness of sport, home advantage, and similarity of team fitness values (in reality, many other factors contribute—weather, immediate form of individual players, injuries, and others). As our initial empirical analysis shows, all these factors are common to sports results data. If substantial randomness of results is inevitable, one can ask if we can at least suppress the unexpected results to help PageRank/BiPageRank perform better and possibly outperform the win ratio. To explore the feasibility of this idea, we benefit from the use of synthetic data where team fitness values are known. We can thus identify the unexpected results (wins of weaker teams against stronger teams) and either remove them from the dataset (see Fig. 5a) or reverse them (see Fig. 5b). In Fig. 5, we remove or correct a gradually increasing fraction of unexpected outcomes ($\eta = 1$ means that all unexpected outcomes have been treated) which naturally benefits all three evaluated algorithms. However, PageRank and bi-directional PageRank require large η for their performance to improve substantially whereas the win ratio improves uniformly in the whole range of η. As a result, there is no η for which PageRank or bi-directional PageRank perform better than the win ratio. In empirical data where team fitness values are not directly known, one would first have to identify the unexpected results, which would further lower the efficiency of this approach. We can thus conclude that the removal or correction of unexpected results cannot help PageRank and bi-directional PageRank outperform the win ratio.

3.3. Ranking performance on empirical datasets

After comparing the ranking performance on synthetic datasets in the previous section, we now present a similar comparison on empirical datasets. Since team fitness values are not known in empirical data, we use the ranking of all teams at the end of the season as the ground truth against rankings produced by respective ranking algorithms in earlier parts of the season. This choice is motivated by the earlier observation that the team win ratio is a good proxy for team fitness [see the discussion before Eq. (9)]. Denoting the number of games in season s as N_S, we then use first PN_S games as input for algorithm A and quantify the algorithm's performance using Kendall's τ between the computed ranking and the end of season number of wins, thus obtaining $\tau_S(P, A)$. This is then averaged over seasons to produce $\tau(P, A)$. In this way, we compare the performance of the win ratio with that of bi-directional PageRank by evaluating $\tau(P, \text{BiPageRank}) - \tau(P, \text{WinRatio})$.

The results are shown in Fig. 6a where each horizontal bar represents one league with left and right sides representing the start

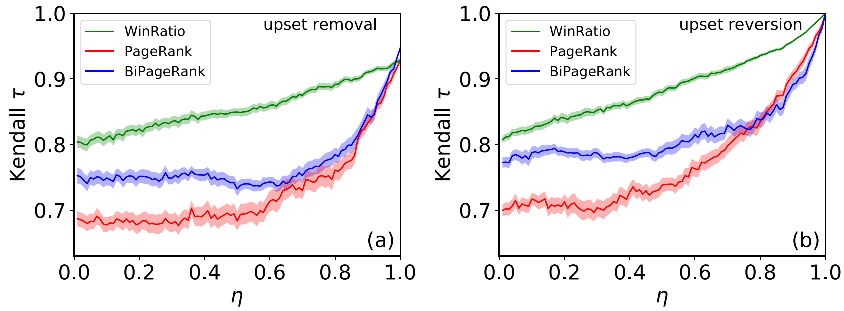

Fig. 5. Ranking performance of the evaluated algorithms when fraction η of unexpected outcomes (i wins over j when $f_i < f_j$) are: (a) removed, (b) reverted. Simulation parameters: $N = 30$, $\delta = 0.25$, $H = 0.08$, $P = 1.0$, $\alpha = 1.5$, $\beta = 0.5$. The results are averaged over 100 independent realizations of the model, shaded and the shaded regions indicate double of the standard error of the mean.

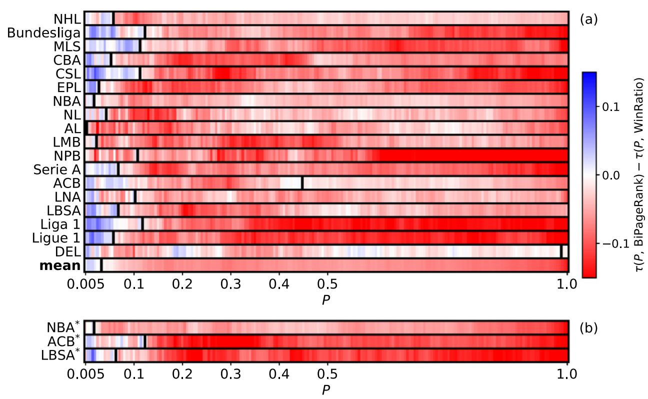

Fig. 6. The performance difference $\tau(P, \text{BiPageRank}) - \tau(P, \text{WinRatio})$ in empirical sports data where the final win ratio in each season is used as the ground truth (the results are averaged over the last 10 available seasons). Individual rows represent different leagues from the beginning ($P = 0$, left) until the end ($P = 1$, right). Vertical markers highlight the highest P value where BiPageRank outperforms WinRatio. Panels (a) and (b) are obtained using the game outcomes and scores data, respectively.

and the end of the season, respectively, and the computed differences between BiPageRank and WinRatio are color-coded. We see that despite using the final fraction of wins in a season as the ground truth (which obviously favors WinRatio), PageRank is still able to outperform WinRatio when P is small. In a direct parallel with previously presented results on synthetic data, the win ratio outperforms bi-directional PageRank almost always except for the very beginning of the season (small P). To highlight the transition between the early part of the league when BiPageRank is best and the later part when the win ratio is best, we mark the largest P at which $\tau(P, \text{BiPageRank}) - \tau(P, \text{WinRatio}) > 0$ with a vertical line for each league. These threshold P values are around 0.1 or lower except for two leagues (ACB and DEL) for which narrow ranges with weakly positive $\tau(P, \text{BiPageRank}) - \tau(P, \text{WinRatio})$ appear also for large P. The overall behavior is best visible in the last row where the difference between BiPageRank and WinRatio is averaged over all considered leagues. The threshold P value here is 0.035 and WinRatio never outperforms BiPageRank by more than 0.03 (as measured by Kendall's τ). This confirms in a model-free way that PageRank and BiPageRank are beneficial for sports results data only when the information is scarce (P is low). When sufficiently many teams have already played against each other, the win ratio generally ranks the teams better.

To conclude, we test if replacing binary match outcomes with the corresponding scores alters the presented results. To this end, we weigh the link representing a match between teams i and j with the score difference. This reflects the hypothesis that a match won by a narrow margin discerns the competing teams' abilities less than a match where the score difference is large. Results obtained by repeating the above-described experiments for three basketball leagues (Fig. 6b) agree well with the results obtained using binary match outcomes (Fig. 6a). This suggests that our results are robust for different representations of the input sports results data.

4. Conclusions

In this paper, we have focused on sports results data from regular leagues where a fixed number of teams play against each other. Results of the games can be represented as a directed network where a directed link from i to j is formed when team j has won over i. We have found that results from less random sports, in our case soccer and basketball, result in networks with indegree assortativity which is the necessary condition for PageRank to perform well. We have calibrated a model for synthetic sports results, a variant of the classical Bradley–Terry model [19]. The model uses only two parameters, home advantage H and sport randomness δ, yet it produces excellent agreement with empirical sports results.

We have assessed the ranking performance of three distinct methods to rank the competing teams: Their win ratio, their PageRank score, and their newly proposed bi-directional PageRank score. Bi-directional PageRank combines two different scores: One positive which accumulates mainly through winning over good opponents (as in PageRank), the other negative which accumulates mainly through losing against bad opponents. The ranking algorithms have been first evaluated on synthetic data. The main finding is that PageRank only outperforms the win ratio when a small fraction of all games have been played *and* randomness of results are sufficiently small. In particular, PageRank yields for the levels of randomness found in empirical sports data (we considered baseball, ice hockey, soccer and basketball). Note that while [44] reports that incompleteness of the network is harmful to PageRank's performance, which is natural, we find that incompleteness of the network is actually favorable to PageRank when its performance is judged relatively to the win ratio benchmark.

The newly proposed BiPageRank outperforms PageRank for all parameter settings, yet it outperforms WinRatio only for sports with the lowest randomness when a small fraction of all games (20%) have been played. The added value of introducing BiPageRank thus lies in demonstrating that PageRank alone, despite its wide use, is not necessarily the best algorithm for sports results data. The reason why BiPageRank outperforms PageRank is simple: It uses more information by taking into account against whom the teams have lost (by comparison, PageRank considers only against whom the teams have won). As a result, two teams that lost all their games so far can be distinguished by who were their opposing teams, for example.

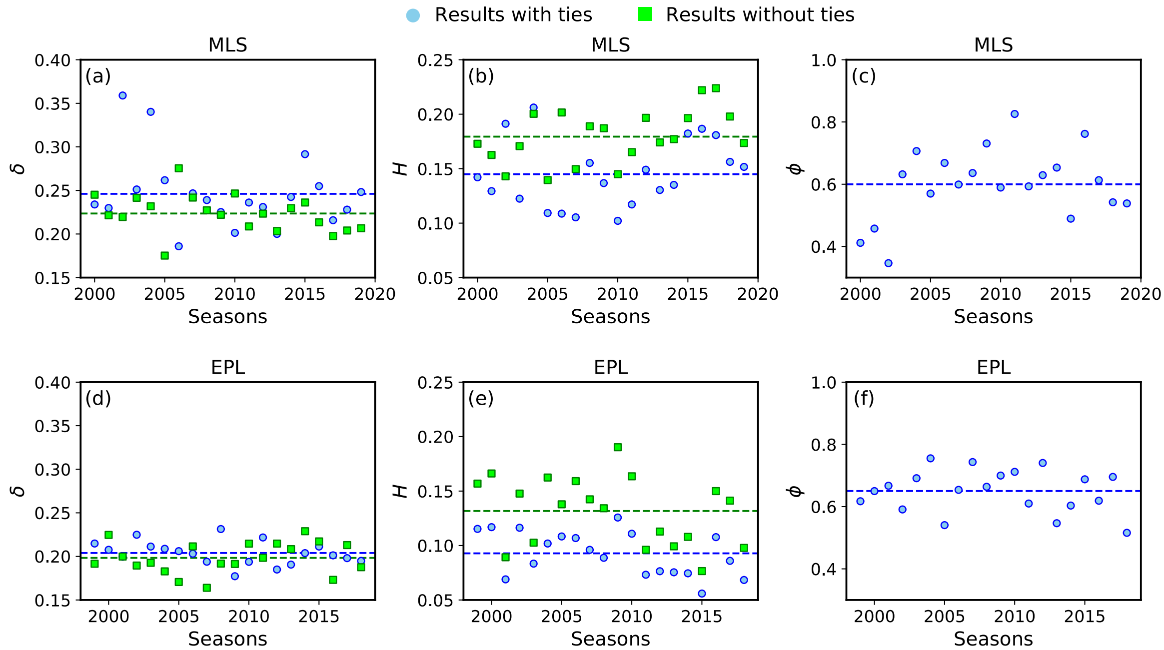

Fig. A.7. Parameter estimation results with and without ties for two soccer leagues, MLS and EPL. The dashed lines show the average parameter value for the two respective models over the analyzed seasons.

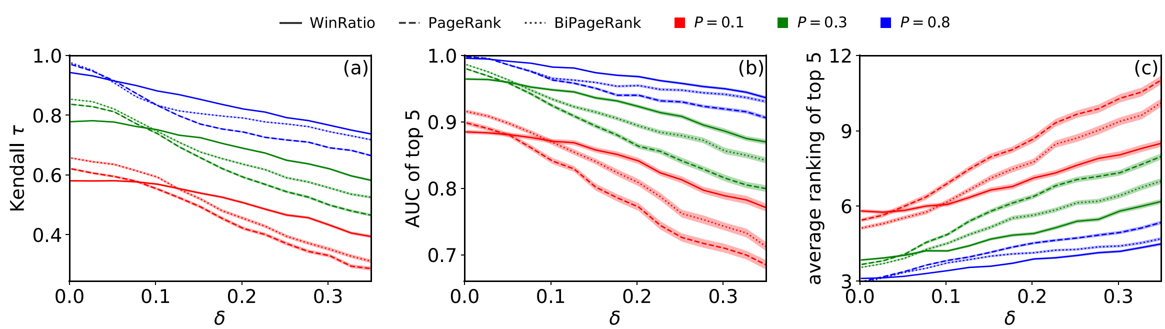

Fig. A.8. Performance of the studied ranking algorithms on synthetic data for 30 teams with ties ($\phi = 0.6$) and without home advantage ($H = 0$). As in Fig. 3, results are shown for three different fractions of played matches, P.

Both PageRank and BiPageRank further suffer when other sources of randomness—in our case home advantage and non-uniform distribution of team abilities—are considered. The sensitivity of PageRank, and closely related BiPageRank, to changes in the data has been already discussed in, for example, [45]. We demonstrate here, for the first time, that this sensitivity combined with the natural randomness of sport renders PageRank of little use on results from a sport tournament. By contrast, the ranking of teams by their win ratio turns out to be comparatively robust to various sources of randomness in results. Nevertheless, our framework for generating synthetic data can be used to continue the search for an algorithm that would outperform the simple ranking of teams by the number of wins.

To keep the model for synthetic sports results simple, we neglected further factors that can be addressed in future research. The assumption of fixed team fitness can be relaxed to allow for modeling a variable sport level or temporary adverse effects of injuries, for example. The simple tournament setup can be generalized to irregular games between the teams as is the case for national teams in soccer or players in tennis, for example. In tennis, in particular, each player has recently played only a small fraction of other players. Using the terminology of our model, the effective P is small, which suggests that PageRank might have some merit for tennis data. To set up a model for tennis-like synthetic results, and evaluate ranking algorithms on such data, is another exciting open direction.

Besides providing specific results on the use of PageRank on sports results data, our work highlights the need to carefully assess the actual performance and limitations of network metrics. This need is exacerbated by the complexity of systems that produce the data, which makes it difficult to judge *ex-ante* if an algorithm is a good match for the data. In citation data, for example, PageRank has been frequently used yet [46] shows that the natural growth of the citation network makes PageRank scores difficult to interpret. If a ground truth set is available, a comparative assessment on empirical data is possible. This can be made more robust by considering multiple empirical datasets and multiple ground truth sets as done recently in [47] to compare ranking metrics for citation data. If a ground truth set is not available but a credible model for a given system exists, an assessment using synthetic data (as we have used here) is a practical alternative. Using a network metric without understanding its scope and limitations directly induces the risk of obtaining unreliable or inferior results.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Swiss National Science Foundation (grant No. 182498). MM and AZ acknowledge support from the National Natural Science Foundation of China (grant Nos. 11850410444 and 71731002, respectively). RJW acknowledges support from the China Scholarship Council (CSC).

Appendix. Ranking performance on synthetic datasets with ties

Similarly as proposed by Rao and Kupper [20], our data-generating model can be modified to produce ties by writing the probabilities of home team i winning and losing against away team j, respectively, as

$$P(i \succ j) = 1/\left[1+\mathrm{e}^{(f_j-f_i-H)/\delta+\phi}\right], \quad P(j \succ i) = 1/\left[1+\mathrm{e}^{(f_i-f_j+H)/\delta+\phi}\right]. \tag{A.1}$$

The tie probability is then $P(i \sim j) = 1 - P(i \succ j) - P(j \succ i)$. The new parameter $\phi \geq 0$ determines the fraction of tied matches; when $\phi = 0$, our original model is recovered as $P_0(i \sim j) = 0$.

We first estimate the model parameters, δ, H, and ϕ using MLE as before. We find that taking ties into account has little effect on the estimated values of δ and H (Fig. A.7). With ties, the estimated home advantage is systematically lower by 0.04, on average. The estimated values of the ties-related parameter ϕ mostly lie between 0.5 and 0.7 for both leagues.

Except for ties introduced through Eq. (A.1), the algorithm to generate synthetic data is the same as before. In the network representation, every tied result is represented by a pair of edges of opposing directions with weights 0.5 each. Using $\phi = 0.6$ motivated by the parameter estimation above, approximately 25% of all results in the synthetic data are ties (the exact fraction varies with δ). We find that the ranking performance of different algorithms on data with ties (Fig. A.8) is qualitatively consistent with the results presented before for data without ties (Fig. 3). As a result of improved performance of WinRatio, especially for larger δ, the range of δ where PageRank and BiPageRank are superior to WinRatio is reduced by introducing ties in the data. This can be due to half-points awarded for ties making the WinRatio ranking less degenerate, and thus improving its ranking ability. In summary, our main finding holds also when ties are introduced in the data.

References

[1] C.M. Frederick-Recascino, H. Schuster-Smith, Competition and intrinsic motivation in physical activity: A comparison of two groups, J. Sport Behav. 26 (3) (2003) 240–254.
[2] R. Le Bouc, M. Pessiglione, Imaging social motivation: Distinct brain mechanisms drive effort production during collaboration versus competition, J. Neurosci. 33 (40) (2013) 15894–15902.
[3] P. Hone, R. Silvers, Measuring the contribution of sport to the economy, Aust. Econ. Rev. 39 (4) (2006) 412–419.
[4] S. Rosner, K. Shropshire, The Business Of Sports, Jones & Bartlett Publishers, 2011.
[5] B.D. McPherson, J.E. Curtis, J.W. Loy, The Social Significance of Sport: An Introduction to the Sociology of Sport, Human Kinetics Publishers, 1989.
[6] R. Giulianotti, Sport: A Critical Sociology, John Wiley & Sons, 2015.
[7] S. Szymanski, The economic design of sporting contests, J. Econ. Lit. 41 (4) (2003) 1137–1187.
[8] E. Dechenaux, D. Kovenock, R.M. Sheremeta, A survey of experimental research on contests, all-pay auctions and tournaments, Exp. Econ. 18 (4) (2015) 609–669.
[9] D.F. Gleich, PageRank beyond the Web, SIAM Rev. 57 (3) (2015) 321–363.
[10] Y. Gao, X. Yu, H. Zhang, Overlapping community detection by constrained personalized PageRank, Expert Syst. Appl. 173 (2021) 114682.
[11] F. Radicchi, Who is the best player ever? A complex network analysis of the history of professional tennis, PLoS One 6 (2) (2011) e17249.
[12] P.S.P. Júnior, M.A. Gonçalves, A.H. Laender, T. Salles, D. Figueiredo, Time-aware ranking in sport social networks, J. Inf. Data Manag. 3 (3) (2012) 195.
[13] V. Lazova, L. Basnarkov, PageRank approach to ranking national football teams, 2015, arXiv preprint arXiv:1503.01331.
[14] A. Daud, F. Muhammad, H. Dawood, H. Dawood, Ranking cricket teams, Inf. Process. Manage. 51 (2) (2015) 62–73.
[15] K.G. Bigsby, J.W. Ohlmann, Ranking and prediction of collegiate wrestling, J. Sports Anal. 3 (1) (2017) 1–19.
[16] A.G. Tennant, C.M. Smith, J.E. Chen C, Who was the greatest of all-time? A historical analysis by a complex network of professional boxing, J. Complex Netw. 8 (1) (2020) cnaa009.
[17] S. Brin, L. Page, The anatomy of a large-scale hypertextual Web search engine, Comput. Netw. ISDN Syst. 30 (1) (1998) 107–117.
[18] S. Fortunato, M. Boguñá, A. Flammini, F. Menczer, Approximating PageRank from in-degree, in: International Workshop On Algorithms And Models For The Web-Graph, Springer, 2006, pp. 59–71.
[19] R.A. Bradley, M.E. Terry, Rank analysis of incomplete block designs: I. The method of paired comparisons, Biometrika 39 (3/4) (1952) 324–345.
[20] P. Rao, L.L. Kupper, Ties in paired-comparison experiments: A generalization of the Bradley-Terry model, J. Am. Stat. Assoc. 62 (317) (1967) 194–204.
[21] R.Y. Aoki, R.M. Assuncao, P.O. Vaz de Melo, Luck is hard to beat: The difficulty of sports prediction, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017, pp. 1367–1376.
[22] W. Deng, W. Li, X. Cai, A. Bulou, Q.A. Wang, Universal scaling in sports ranking, New J. Phys. 14 (9) (2012) 093038.
[23] R. Chetrite, R. Diel, M. Lerasle, The number of potential winners in Bradley–Terry model in random environment, Ann. Appl. Probab. 27 (3) (2017) 1372–1394.
[24] E. Ben-Naim, N. Hengartner, S. Redner, F. Vazquez, Randomness in competitions, J. Stat. Phys. 151 (3–4) (2013) 458–474.
[25] A.J. O'Malley, Probability formulas and statistical analysis in tennis, J. Quant. Anal. Sports 4 (2) (2008).
[26] M. Cattelan, Models for paired comparison data: A review with emphasis on dependent data, Statist. Sci. 27 (3) (2012) 412–433.
[27] W.N. Colley, Colley's Bias Free College Football Ranking Method: The Colley Matrix Explained, Princeton University, Princeton, 2002.
[28] M. Brechot, R. Flepp, Dealing with randomness in match outcomes: How to rethink performance evaluation in European club football using expected goals, J. Sports Econ. 21 (4) (2020) 335–362.
[29] C. Redmond, A natural generalization of the win-loss rating system, Math. Mag. 76 (2) (2003) 119–126.
[30] J. Park, M.E. Newman, A network-based ranking system for US college football, J. Stat. Mech. Theory Exp. 2005 (10) (2005) P10014.
[31] J.P. Keener, The Perron–Frobenius theorem and the ranking of football teams, SIAM Rev. 35 (1) (1993) 80–93.
[32] S. Mukherjee, Identifying the greatest team and captain: A complex network approach to cricket matches, Physica A 391 (23) (2012) 6066–6076.
[33] A.Y. Govan, C.D. Meyer, R. Albright, Generalizing Google's PageRank to rank national football league teams, in: Proceedings Of The SAS Global Forum, vol. 2008, 2008, Paper 151.
[34] S. Motegi, N. Masuda, A network-based dynamical ranking system for competitive sports, Sci. Rep. 2 (2012) 904.
[35] O. Hubáček, G. Šourek, F. Železný, Learning to predict soccer results from relational data with gradient boosted trees, Mach. Learn. 108 (1) (2019) 29–47.
[36] C.B. Beggs, S.J. Shepherd, S. Emmonds, B. Jones, A novel application of PageRank and user preference algorithms for assessing the relative performance of track athletes in competition, PLoS One 12 (6) (2017) e0178458.
[37] M. Newman, Networks, Oxford University Press, 2018.
[38] Z.-M. Ren, M.S. Mariani, Y.-C. Zhang, M. Medo, Randomizing growing networks with a time-respecting null model, Phys. Rev. E 97 (5) (2018) 052311.
[39] A. Nevill, N. Balmer, S. Wolfson, The extent and causes of home advantage: Some recent insights, J. Sports Sci. 23 (4) (2005) 335–445.
[40] H.V. Ribeiro, S. Mukherjee, X.H.T. Zeng, The advantage of playing home in NBA: Microscopic, team-specific and evolving features, PLoS One 11 (3) (2016) e0152440.

[41] L. Lü, T. Zhou, Link prediction in complex networks: A survey, Physica A 390 (6) (2011) 1150–1170.
[42] G. Claeskens, N.L. Hjort, Model Selection and Model Averaging, Cambridge University Press, 2008, Cambridge Books.
[43] S. Da Silva, R. Matsushita, E. Silveira, Hidden power law patterns in the top European football leagues, Physica A 392 (21) (2013) 5376–5386.
[44] G. Ghoshal, A.-L. Barabási, Ranking stability and super-stable nodes in complex networks, Nature Commun. 2 (1) (2011) 1–7.
[45] T.P. Chartier, E. Kreutzer, A.N. Langville, K.E. Pedings, Sensitivity and stability of ranking vectors, SIAM J. Sci. Comput. 33 (3) (2011) 1077–1102.
[46] M.S. Mariani, M. Medo, Y.-C. Zhang, Ranking nodes in growing networks: When PageRank fails, Sci. Rep. 5 (2015) 16181.
[47] S. Xu, M.S. Mariani, L. Lü, M. Medo, Unbiased evaluation of ranking metrics reveals consistent performance in science and technology citation data, J. Informetr. 14 (1) (2020) 101005.