



The effects of scheduling network models in predictive processes in sports

Marc Garnica Caparrós¹ · Daniel Memmert¹ · Fabian Wunderlich¹

Received: 3 March 2022 / Revised: 24 May 2022 / Accepted: 17 September 2022 / Published online: 1 October 2022
© The Author(s) 2022

Abstract

In many sports disciplines, the schedule of the competitions is undeniably an inherent yet crucial component. The present study modeled sports competitions schedules as networks and investigated the influence of network properties on the accuracy of predictive ratings and forecasting models in sports. Artificial networks were generated representing competition schedules with varying density, degree distribution and modularity and embedded in a full rating and forecasting process using ELO ratings and an ordered logistic regression model. Results showed that network properties should be considered when tuning predictive ratings and revealed several aspects for improvement. High density does not increase rating accuracy, so improved rating approaches should increasingly use indirect comparisons to profit from transitivity in dense networks. In networks with a high disparity in their degree distribution, inaccuracies are mainly driven by nodes with a low degree, which could be improved by relaxing the rating adjustment functions. Moreover, in terms of modularity, low connectivity between groups (i.e., leagues or divisions) challenges correctly assessing a single group's overall rating. The present study aims to stimulate discussion on network properties as a neglected facet of sports forecasting and artificial data to improve predictive ratings.

Keywords Network science · Network entropy · Sports forecasting · Sports scheduling networks · Predictive methods

1 Introduction

The use of predictive models in sports has been widely addressed by several disciplines as an opportunity to develop, evaluate and display new analytical methods. Among other predictive tasks related to, for example, tactical behavior (Seidl et al. 2018) or injury prevention (Rossi et al. 2018), accurately predicting the outcome of sports events received significant attention due to knowledge transferability, data availability and economic reasons (McHale and Swartz 2019; Wunderlich and Memmert 2021). Despite the seemingly limited field of sports, the subject of predicting the outcome of sports events does not only vary in terms

of the sport: American football (Baker and McHale 2013), basketball (Manner 2016), European football (Koopman and Lit 2019), horse racing (Lessmann et al. 2010), or tennis (Kovalchik 2016); but also the competition: national leagues (Angelini and de Angelis 2019), international tournaments (Groll et al. 2020), or Olympics (Forrest et al. 2010); and the forecasting level of detail (i.e., winner of the event (Pachur and Biele 2007), set of probabilities for the possible outcomes (Koopman and Lit 2019), etc.). Besides domain-specific knowledge from sports and forecasting, methods stemming from computer science such as data mining and machine learning have gained importance in recent years to handle the complex available sports data for predictive purposes (Berrar et al. 2019; Bunker and Thabtah 2019; Horvat and Job 2020; Hubáček et al. 2019a; Miljković et al. 2010).

The majority of predicting models targeting the outcome of sports events can be encapsulated into a set of components that shape the *sports forecasting process* (Garnica Caparrós et al. 2021). Implicitly or explicitly, mathematical models use numerical representations of the competitors' strength or quality, so-called ratings. Ratings aim to provide a sound assessment of the team's or player's performance. Competitors ranking systems are usually implied from a basic rating

✉ Marc Garnica Caparrós
m.garnica@dshs-koeln.de

Daniel Memmert
d.memmert@dshs-koeln.de

Fabian Wunderlich
f.wunderlich@dshs-koeln.de

¹ Institute of Exercise Training and Computer Science in Sports, German Sports University Cologne, 50933 Cologne, Germany

sorting. Popular sports are based on trivial point-rewarding systems such a football league table or tournament results like the ATP tennis world ranking. From a forecasting point of view, these rankings are not optimal for predictive scenarios. On the other side, forward-looking ratings gathered more attention from the scientific community thanks to their predictive properties (Stefani 2011). As a second component in the sports forecasting process, these ratings are the primary basis of a prediction model, referred to in this paper as the forecasting model. The forecasting model generally yields the final probability distribution among all the possible outcomes of a sports event. The model usually considers the ratings a systematic effect within other factors (i.e., home advantage) and unsystematic effects such as randomness. After an observation (i.e., a match), the ratings are updated by observing the real and expected outcomes.

Finally, statistical or economic measures can assert the forecasting model quality. Model's accuracy measures how efficient the prediction goal was executed, i.e., if the events predicted were indeed observed. On the other side, the model's quality might also be evaluated by generating benefits in the sports betting market. Accuracy and profitability of the forecasting model are usually correlated but should not be generally equated when it comes to model training and evaluation (Wunderlich and Memmert 2020). The inherent nature of this process includes a high degree of uncertainty as only the result of the sport event is observed. Yet, it is not easy to distinguish whether inaccuracy in the models can be attributed to unsystematic effects (i.e., randomness), forecasting inefficiencies or incorrect assessment of competitors strength.

Each step of the fore-mentioned process has been a productive area of application of statistical methods and data mining algorithms. The assessment of sports competitors (i.e., players or teams) can be performed from different approaches. Ratings can be derived from ELO Rating variants (Glickman and Jones 1999) or comparable mathematical models. Recent advances on the availability of highly detailed data in the sports domain boosted the research in determining key performance indicators of sports competitors to achieve accurate predictions (Goes et al. 2019; Jayanth et al. 2018; Miljković et al. 2010). With regard to the forecasting models, deriving predictions from ratings is extensively studied in several sports (Wunderlich and Memmert 2021). However, the current literature seems to oversee the study of a fundamental and cross-discipline attribute of every sport, its schedule. The schedule, usually called the calendar, of a sports competition can be represented as a complex network of participants and pairwise confrontations.

Any sports competition has a predefined structure that will eventually evolve following scheduling rules. The sports forecasting literature has an extended application

to all those sports disciplines governed by a sequence of pairwise comparisons or confrontations. This list of sports disciplines includes all team sports where there are never more than two teams competing against each other and individual sports such as chess and tennis. In these cases, the tournament derives the results from sequences of pairwise confrontations. Tournaments differ in scheduling these comparisons; national football leagues in Europe derive a final winner from a double round-robin scheduling. In US sports, teams are organized in conferences with a higher frequency of games. Tennis tournaments are usually defined as a single-elimination competition. Many tournaments often have schedules that combine non-elimination stages with a final knockout stage (e.g., NBA regular season and playoffs or UEFA Champions League group stage and double-leg knockout stage). If the forecasting model learns and predicts from every event of the tournament and ratings update the competitor's strength depending on every confrontation expected and observed outcomes, it is intuitive to expect that the density, time order, and distribution of these events would influence the accuracy of the process and its evolution. Even if not explicitly considered, rating procedures and forecasting models might benefit from the competition structure and follow their temporal order to perform learning and predicting tasks.

The relational nature in pairs between competitors in the sports described below can be translated into a network model with the vertices representing the competitors and the edges representing the set of confrontations of the tournament. A network abstraction of a system allows describing the local relationships pairwise and the macro-patterning above the structure, permitting knowledge to develop. The use of graph theory and network abstraction of sports tournaments to solve scheduling problems (Drexler and Knust 2007; de Werra 1985) is presented in several sports disciplines such as football (Ribeiro 2012) or tennis (Ghoniem and Sherali 2010). Indeed, network science, as defined by Brandes et al. (2013), has experimented an exponential growth in recent years as their applications proved to be insightful in topics such as human migration (Pitoski et al. 2021) or epidemiology (Bansal et al. 2010). In sports, some contributions used similar network abstractions and network science methods. In Park and Newman (2005), the authors analyzed the same network model with win-loss differential additions to investigate how to rank teams on competitions with several conferences. Another study (de Saá Guerra et al. 2012) used the network adjacency matrix to measure the level of competitiveness in basketball leagues. Recently, a rank-based social network analysis on team strength and confrontations was proposed (Shi and Tian 2020) that aimed to synthesize performance metrics of each team in the weight of the network edges. In European Football, network analysis plays a role in performance analysis of teams

and players by the so-called passing networks (Buldú et al. 2018; Duch et al. 2010), a network abstraction with players being the nodes and edges the passes between them as weighted and directed links. A recent study (Medina et al. 2021) explores these passing networks properties and their effects on the match outcome.

1.1 Definitions

1.1.1 The scheduling network model

The network model consists of competitors (i.e., teams or players) as nodes and matches or confrontations as edges (Ribeiro 2012). The existence of an edge e_{ij} between the node i and the node j denotes a match between the two competitors (see Fig. 1). Sports with clear meaning and impact on the venue of the match (i.e., European Football leagues with home and away fair matches between teams) are represented as a directed network where the edge source is the away team, and the target of the edge is the home team. In sports where the venue is irrelevant (i.e., Tennis or Chess tournaments), their network representation does not consider the edges' direction.

Figure 1 exemplifies two states of the network model and showcases its temporality. The network is temporal from two different perspectives. First, edges are scheduled on a certain day and can be timely arranged following the schedule of the competition and the order of matches. Second, some networks might not be fully scheduled at a certain point and receive edge additions depending on the actual outcome of certain confrontations (e.g., in elimination tournaments, the winners move to the next stage, and new games are scheduled). Rating procedures and forecasting models are integrated into this temporal schema and can be accessed in the network as the node and edge attributes, respectively. Ratings are a series of values linked to each node, and the forecasting models add a set of predictions related to each edge. This paper analyzed the forecasting models yielding a distribution of probabilities among all the possible outcomes (i.e., node x win probability vs node y win probability in elimination matches or without draw option or

home-draw-away probabilities for a football match). Thus, rating $r_{x,i}$ refers to the node x rating value before the $i + 1$ th confrontation while r_x is the set of rating values assigned to the node x ordered by time. Similarly, each edge contains a set of forecasts trying to predict the event's outcome. If appropriate, the actual outcome of the match is also added to the edge attributes.

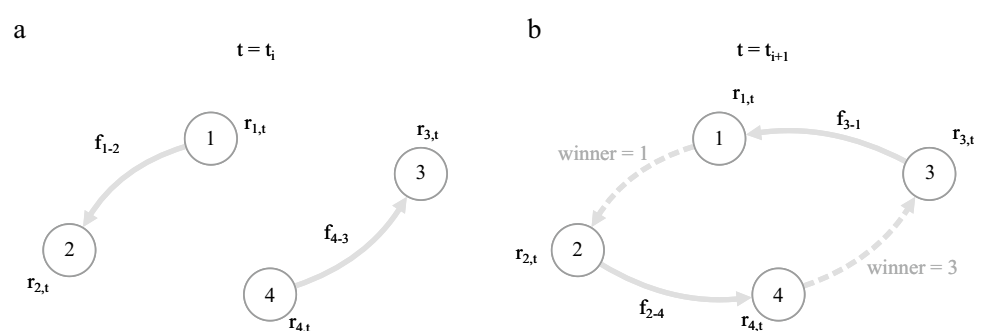
1.1.2 Network properties

The topological study of the networks must be linkable to real-world scenarios and provide new knowledge in current sports domains and future applications. Additionally, the network properties ought to be available at the macro level of the network and allow a feasible random generation of networks (to understand the generation of random networks, refer to Sect. 2.3.). The generation of networks for the study used the following network measures (Newman 2003).

The density of a network is defined as the ratio between the actual number of edges and the maximum number of edges in a network. In our competition network model, this translates into the proportion of matches compared to the maximum number of possible matches. In rating procedures, it could be expected that leagues with a high density, such as national sports leagues, provide an easier environment to assess and compare the strength of each competitor. In contrast, sports tournaments with less density of matches oppose more challenging scenarios. The assumption that the structure and connection of the presented network model contain entropy and, therefore, a knowledge that could be used for predictive tasks is highly related to the transitivity property of a network. Transitivity is a marker of how nodes tend to cluster together.

The degree is a significant feature of a network node. Each vertex degree is equivalent to the number of edges adjacent to the node. The relevance of the vertex degree not only is present in the micro-analysis of complex networks but also in the macro-analysis with the network degree distribution. The study of this network property directly tackles the challenges that might source from leagues with different participation frequencies between competitors and leagues

Fig. 1 Simple network model representing the sports competition network implemented. **a** Ratings and forecasting for each of the nodes and edges at time t_i . **b** Status of ratings, forecasts and observed results at time t_{i+1}



with all competitors participating on the same number of occasions. It is expected that high degree nodes would be more accurately estimated than low degree nodes.

The modularity of a network measures the structure of a network by looking at its divisions and the connections within divisions. Networks divisions are often referenced in the literature as clusters or communities. High modularity structures have dense relationships within clusters but sparse connections between them. Modularity is a notable characteristic of US sports divided by conference or international European leagues gathering teams from different national leagues. Often, predicting the outcome of a match within two competitors from different divisions forms a bigger challenge than predicting matches within the same division. It is expected that the size and density of the groups and the interconnections between groups will affect the predictive ability of matches within divisions.

1.2 Real-world examples

To further contextualize and motivate the focus of this study, four different real-world sports tournaments are visualized as their scheduling network model in Fig. 2. Topologically, large differences among sports disciplines and tournaments become visible. National leagues (Fig. 2a) form dense structures where every competitor faces each other twice in a round-robin schedule. However, these national leagues are also included in bigger networks (Fig. 2b) where international leagues concur (i.e., UEFA Champions League games). In these scenarios, several dense connected components are present in the network model, and just a few connections are formed between components. Figure 2c shows a smaller competition of professional men's European football national teams as an example of a sports competition with modifiable scheduling rules: six connected components of four nodes can be identified referring to the group stage of the competition, only a subset of the most successful nodes

is then participating in the knockout stage that involves all the components. Last but not least, a more unstructured network model is presented in Fig. 2d, representing one full season of a professional men's tennis circuit. This network model involves more nodes than the others and is highly sparse. Moreover, the degree of each node is a determinant aspect as a component of the nodes in the network is involving the vast majority of the edges. In contrast, a lot of other nodes are left to fewer interactions.

An overview of the main characteristics of each network model with regard to the properties presented in the previous section is shown in Table 1. As an initial exploratory analysis, real-world sports competitions networks are generally sparse when taking the full network model where every competitor faces each other, only happening in domestic leagues. Interestingly, domestic leagues are the only ones with a constant degree distribution among all its participants, compared to the tennis circuit that contains high fluctuations among the degree of each node and a highly asymmetric distribution. Additionally, international competitions contain predefined components that are indeed full connected, while edges between these divisions are scarce. Please refer to Sect. 2.3. for further discussion on how these real-world metrics were embedded in the artificial generation of networks. Additionally, Sect. 4 provides a brief use case on how the results of this theoretical study could be validated and implemented in these real-world data scenarios.

1.3 Contribution

The contribution of the present study is three-fold. First, it proposes an extendable network model defining sports tournaments confrontations and integrating predictive rating procedures and forecasting models. Second, realistic networks are simulated by random creation while tweaking the desired predefined network properties. Third, artificial data is introduced to recreate a full forecasting process on these

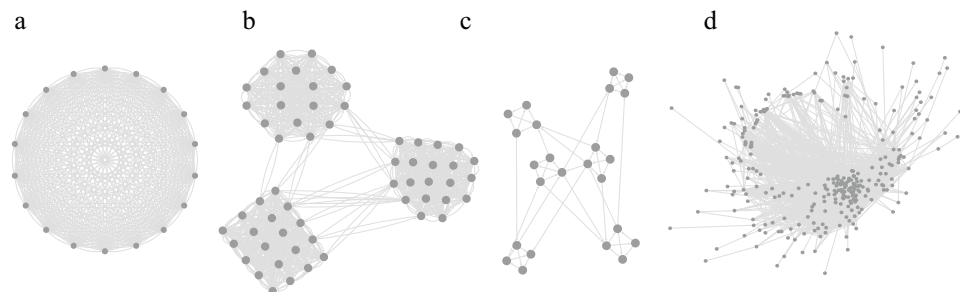


Fig. 2 Real-world sports tournaments visualized as a network of confrontations. **a** One season of the Bundesliga, the professional men's European football league in Germany. **b** Three of the most competitive men's European football leagues in Europe: English Premier League, Spanish First Division and Bundesliga with the two profes-

sional international competitions: UEFA Champions League and Europa League. **c** The 2020 UEFA European Football Championship, a quadrennial international men's European football champions of Europe. **d** The 2019 Tennis ATP Tour, the men's professional tennis circuit

Table 1 Network properties observed in the different sports tournaments

Sports Competition	Domestic Football Bundesliga	International Club Football European Leagues ¹	International Football Men's Eurocup 2020	Tennis ATP tour 2019
<i>Density</i>	1.0	0.021	0.092	0.031
<i>Degree distribution</i>				
Median	34	38	4	8
Q1, Q3	[34,34]	[38,44]	[3,5]	[2,35]
Min,Max	[34,34]	[34,53]	[3,7]	[1,79]
Degree spread	0.0	0.158	0.5	4.125
<i>Modularity</i>				
Number of divisions	1	3	6	1
Inter-division connectivity	–	0.013	0.031	–

Density refers to the actual number of edges divided by the maximum number of edges. Theoretically, tennis seasons allow infinite confrontations between two opponents; however, a double-round robin system is used as the maximum number of edges for all competitions. *Degree spread* refers to the ratio between the interquartile range and the median. *Inter-division connectivity* refers to the actual number of edges between divisions divided by the maximum number of edges between divisions (again assuming a double round-robin system).

¹ *BL* Bundesliga, *PL* English Premier League, *LL* Spanish First Division La Liga, *CL* UEFA Champions League, *EL* UEFA Europa League

networks, enabling a network properties dependent analysis of rating accuracy and predictive quality. To the best of our knowledge, this is the first study systematically analyzing the impact of network structure as a neglected facet of forecasting models in sports. The main advantage of this approach is that—in contrast to applications using real-world data—a fully controlled environment of the sports forecasting process is implemented, yielding improved empirical validation and insights on the impact of network structure.

In summary, this study seeks to investigate the impact of several network properties on the optimal training of models and their rating accuracy and predictive quality. Moreover, it aims to identify the potential for improvement of rating procedures by analyzing the strengths and weaknesses of state-of-the-art rating procedures on different network structures.

2 Methods

2.1 Data simulation

In real-world scenarios, sports scheduling changes are rare. Additionally, real-world tournaments do not provide a big enough range of network differences to analyze the impact of the network characteristics in the forecasting methods. To solve this paradigm, this study introduces artificial data. While artificial data is useful to recreate realistic environments without using personal data (Jahangirian et al. 2010), it is also useful to provide extreme scenarios and generate new hypotheses. The simulation framework (Garnica Caparrós et al. 2021) used in this study creates a large set of competition schedules with certain network properties and analyses their impact on network evolution and forecasting.

The networks' generation, modeling, and analysis were implemented in Python programming language (van Rossum and Drake 2009) and the software NetworkX for complex networks analysis (Hagberg et al. 2008). The visualization of the networks was performed by Cytoscape.js graph theory library (Franz et al. 2016).

To achieve an accurate simulation of the sports forecasting domain, all aspects of the sports forecasting process are also added to the simulation. For every single competition simulated (i.e., network), all competitors (i.e., nodes) receive a simulation of their true strength and their evolution in time during the duration of the competition, from now on referred to as *true ratings*. Similarly, every match (i.e., edge) is properly simulated with a certain probability distribution for each possible match outcome, referred to as the *true forecasts*. The match result is drawn from this denominated true forecast. True ratings are created as a numerical function with a specified starting point and trend, including random daily fluctuations and seasonal changes per node. True ratings are the main input to simulate match possible results by deriving a distribution of outcome probabilities. A ternary match result was assumed (i.e., home win, draw, away win), following previous approaches (Hvattum and Arntzen 2010) in European football, the match probabilities are obtained by an ordered logit regression model (OLR) using the difference of competitors ratings as the single covariate. The model yields the match probabilities concerning the possible outcomes (i.e., probability of a home win, draw and away win). Three parameters c_0, c_1, β are required by the model. The true forecast is configured with $c_0 = -0.9, c_1 = 0.3$ and $\beta = 0.006$. The configuration of the true ratings and true forecasts defines the “reality” of the simulated environment. The scheduled competition is modeled as a network, with

nodes being the competitors with associated time-series attributes as ratings and edges representing the confrontation between competitors with associated attributes including time, result, and forecasts.

2.2 Estimates and evaluation metrics

To properly simulate a forecasting process and the impact of each network property, the simulation needs estimators that try to mimic the true ratings accurately, and true forecast explained in the previous section (as in reality, when a predictor/bettor is proposed to rate teams or predict match results by assuming a certain truth underneath). A frequently used method to estimate competitors strength is the ELO Rating (Glickman and Jones 1999; Hvattum and Arntzen 2010). To each competition simulated, the study integrates a trained ELO Rating with parameters $c = 10$, $d = 400$ and a home advantage $\omega = 80$. The parameter of the ELO Rating is left to optimization and used as an additional study metric. The K -factor is examined and interpreted in the literature as how much a competitor's rating can change after a single match. A too-small K -factor creates slow ratings to converge, while too high K -factors create unstable ratings with wild fluctuations.

In contrast to pure real-world applications, the simulated environment enables to measure how accurate the proposed ELO Rating estimates the true ratings. Thereafter, the distance between the estimator values and the true ratings is reported for every node at every point in time as the rating error. Even though this error metric is impossible to obtain in real-world scenarios, it indicates how accurate the estimators are from the real simulated values. Moreover, measuring only the estimator's accuracy against the simulated environment would lack real applicability. To solve this, the predictive value of these estimators is also analyzed. The ELO rating estimator is used as an input of a proposed forecast. An OLR model equivalent to the one used to simulate the data is implemented using as the covariate the difference of the estimated ratings. Using the same model as the one used to simulate the data, the analysis focuses entirely on the estimator rating predictive value without any noise besides the inherent randomness of the data. Two metrics are reported to measure the predictive power of the ELO Rating-based forecast model, the Rank Probability Score (RPS) (Constantinou et al. 2012) and the forecast error. The RPS measures the accuracy of the forecasting model as a quadratic loss function from the actual result of the game, where lower values are interpreted as more accurate forecasts. The RPS evaluation is constrained by the actual result of the match, which sometimes can be misleading (highly unlikely match outcomes are still possible). Therefore, the forecast error is reported as the distance between the true forecast and the estimator-based forecast. This provides a

fair measurement of how similar the estimator model was to the actual forecast.

The study follows the same training split design at every network constructed. The first 20% of edges served as an initialization for the estimator. The following 30% of edges were used as the in-sample training set to select the ELO rating K -factor by optimizing for RPS. For every data set split, a median value for rating error, RPS and forecast error for all confrontations was calculated. Thereafter, the optimal K was calculated by optimizing RPS. Consequently, the remaining 50% of the edges served as the evaluation set. A median value for each evaluation metric was calculated.

2.3 Random networks generation

The second step in the data simulation is to generate several random network topologies for each of the studied network properties: density, degree distribution, and modularity. To properly study the effect of each property, the random networks generated covered the full range of possible values. Several design constraints are introduced for each property to ensure a proper collection of sample networks.

2.3.1 Networks generation by density

Network density can range from 0 (i.e., a low connected network) values to 1 (i.e., a fully connected network). Random networks with the same degree per node were created with an increasing number of nodes to achieve a list of networks that uniformly traverses the network density spectrum. The networks with a lower number of nodes achieved higher density values. A fully connected network mimics a national sports league with a double round-robin schedule where every team plays each other twice through the season. Therefore, the full connected network is obtained by a network with 50 nodes and a node degree of 98. In contrast, the minimum density of 0.04 is achieved by keeping the node degree of 98 and 1226 nodes. A final set of 25 networks was created with a constant node degree of 98 and densities ranging from 0.04 to 1.0 with approximate increments of 0.04. The K -factor of the ELO rating was optimized for each network, and the evaluation metrics were extracted, as explained in the previous section.

2.3.2 Networks generation by degree distribution

A new measurement is introduced for this part of the study as degree distribution spread. A degree distribution spread of value d_s with a mean degree of μ refers to a network with a degree distribution with mean at μ and values uniformly distributed between $\mu - \mu d_s$ and $\mu + \mu d_s$. Networks of 200 nodes were created with a fixed mean degree per node of 120. Thereafter, the generation of networks ranged from a

degree distribution spread of 0 (i.e., all nodes with a degree of 120) to a degree distribution spread of 1 (i.e., all nodes degree drawn uniformly between 0 and 240). In other words, low values for degree distribution spread generated networks with a similar degree for all nodes. In contrast, higher values in the degree distribution spread created networks with a highly variable degree distribution (existence of highly connected nodes vs poorly connected nodes). The desired distributions were obtained using Python NumPy library (Harris et al. 2020), the generation of random networks given a certain degree distribution employed a configuration model (Newman 2003; van der Hofstad 2017).

A final set of 40 networks was created with a distribution spread ranging from 0 to 1 with increments of 0.025. From each network, the low degree nodes are defined as the decile of the nodes with lowest degree (10% of all nodes with the lowest degrees) and the high degree nodes defined as the decile of the nodes with highest degree (10% of all nodes with the highest degree) were extracted. The optimal K -factor and the evaluation metrics were calculated for each subgroup of nodes on each network. This can be viewed as how the degree distribution spread and competition schedules with variable degrees on its competitors' effect on rating calibration and predictive accuracy. Additionally, the optimization of the K -factor was individually observed by degree groups. In this case, the results were reported by comparing the lowest degree nodes and the highest degree nodes for each trial.

2.3.3 Networks generation by modularity

Finally, the third set of networks was created focusing on their modularity. All networks contained 200 nodes and ten divisions or communities (20 nodes per division). The connectivity between divisions is defined as the probability of an inter-division edge being present. The ten divisions were fully connected at each network, while the connectivity between divisions ranged from 0 (10 independent groups) to 1 (a single full connected network of 200 nodes). A final set of 50 networks was created with connectivity ranging from 0 to 1 with increments of 0.02. The analysis determines how inter-division connectivity affects rating calibration and

predictive accuracy. An optimal K -factor and the evaluation metrics were calculated for each network.

Additionally, the rating error was analyzed at the team and division levels. A division rating can be defined as the average of all node's ratings belonging to a certain division. Thereafter, a certain true rating and an estimated rating are present at the division level. Optimally, a new node-level rating error definition was introduced to differentiate between node-level and division-level rating inaccuracies. A division-dependent rating error was calculated as the difference between the rating error and the division rating error for each node. The division-dependent rating error reflects how accurate the rating error is within the scope of a certain division.

3 Results

The results of the three network properties studies are presented similarly. First, the optimal specification for the ELO rating estimator (i.e., selection of K -factor) for every network simulated is presented. Then, the accuracy of the estimators is explained by the network property considering rating accuracy and predictive value. More detailed findings are reported to understand the internal behavior and motivate the discussions for each property characteristic.

3.1 The effects of network density

In this study, the simulated network size increases exponentially as network density decreases. The network with the highest density value contained 50 nodes and 2450 edges, while the lowest density network, at a density of 0.04, contained 1226 nodes. The optimal K -factor for the proposed estimator on each network with a different density is presented in Fig. 3. This parameter does not significantly impact the density values as it remains constant through all the density range. Figure 4 shows the rating accuracy and the predictive performance of the estimators added to each network. No clear effect of network density on any indicators can be identified.

Fig. 3 Optimal K -factor for each of the constructed networks by network density

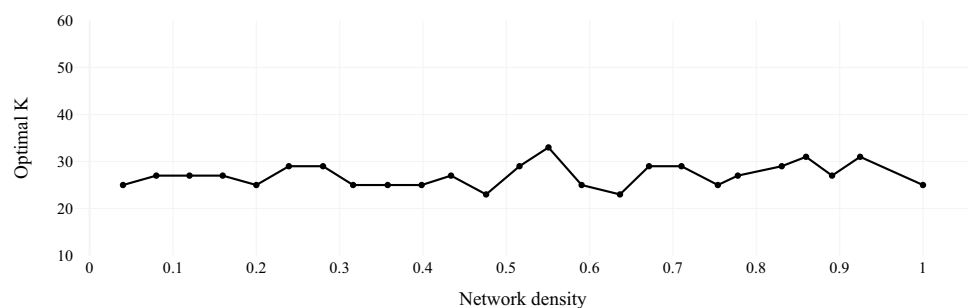
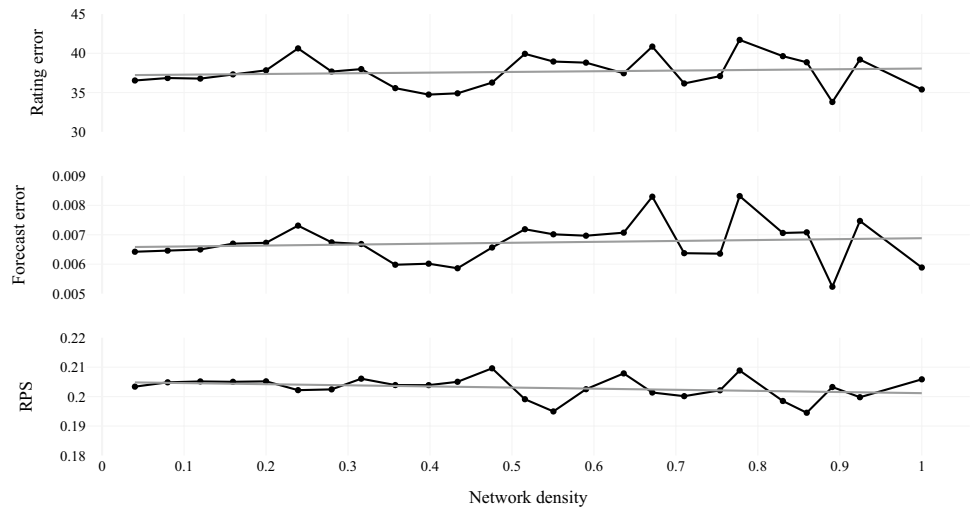


Fig. 4 Rating error and predictive value (actual forecast error and RPS) by network density. Linear trends are presented as a thinner grey line



3.2 The effects of degree distribution spread

This section evaluates the effects of the degree distribution spread in the forecasting capabilities. In this case, the size of the resulting networks remained relatively constant as the main difference between each network was their degree distribution.

With regard to the ELO rating optimal K -factor, Fig. 5 presents how the optimal K resulted by degree distribution spread of the generated networks. While no trend was identified if all nodes are analyzed together, the optimal K behaves differently, when analyzing only the low degree and the high degree nodes. Low degree nodes require higher K values, and their optimal K increases with a high degree distribution spread. In contrast, high degree nodes require lower values of K , specifically in high values of spread, and are not affected by the increase of the distribution spread. Additionally, Fig. 6 shows two individual examples of how the predictive value by RPS differs by certain values of K in high and low degree nodes. At low distribution spread, both functions are similar and result in similar optimal points. In contrast, at high distribution spreads, the functions are

distant and have different shapes, resulting in different optimal points for each group.

The evaluation metrics evolution by degree distribution spread are presented in Fig. 7 considering only the defined low degree and high degree nodes. The rating and forecast error reported for high degree nodes is kept at a constant level while it increases at the low degree nodes as the degree distribution spread is incremented. This effect is not visible if we only observe the RPS value.

3.3 The effects of network modularity

The third collection of studies focused on network modularity. For each network, ten divisions were constructed, and inter-division connectivity was investigated. As shown in Fig. 8, the optimal K for the estimator on each network decreases as we increase the connectivity between divisions. Moreover, Fig. 9 shows a gradual fall in the rating error and forecast error as networks increase their inter-division connectivity. The RPS values also drop; however, given that the RPS depends on the random realized outcomes, the trend appears slightly weaker than the inherent noise in the values.

Fig. 5 Optimal K by degree distribution spread considering only the decile of nodes with lowest degree and highest degree. Linear trend is marked as a thinner grey line

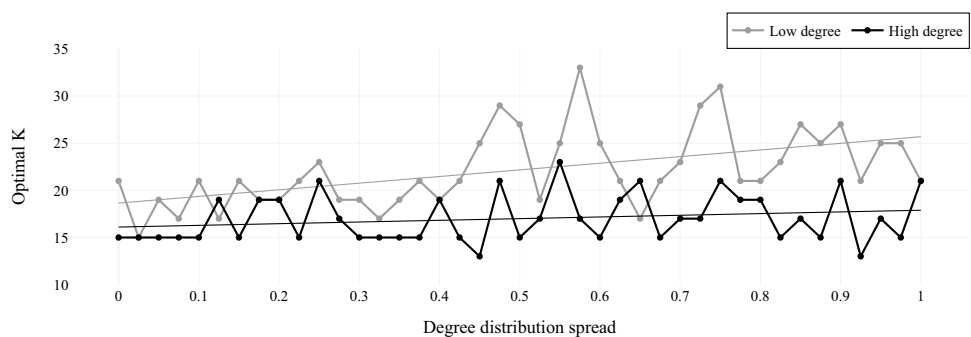


Fig. 6 RPS values for each possible K -factor for the lowest and highest degree groups respectively at distribution spread values of 0.2 and 0.95

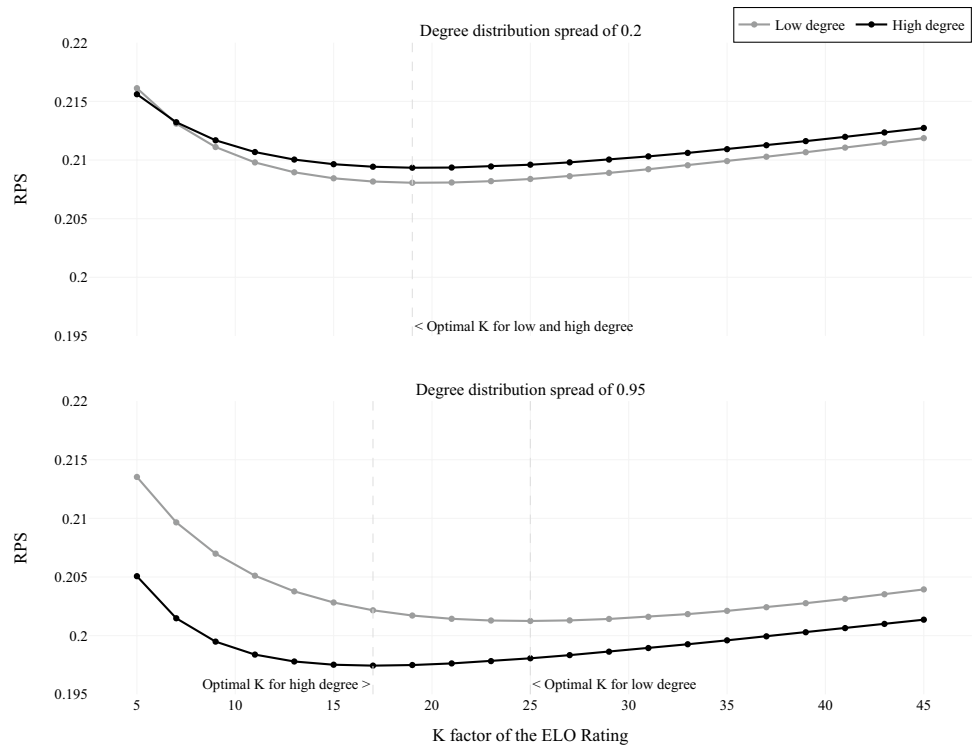
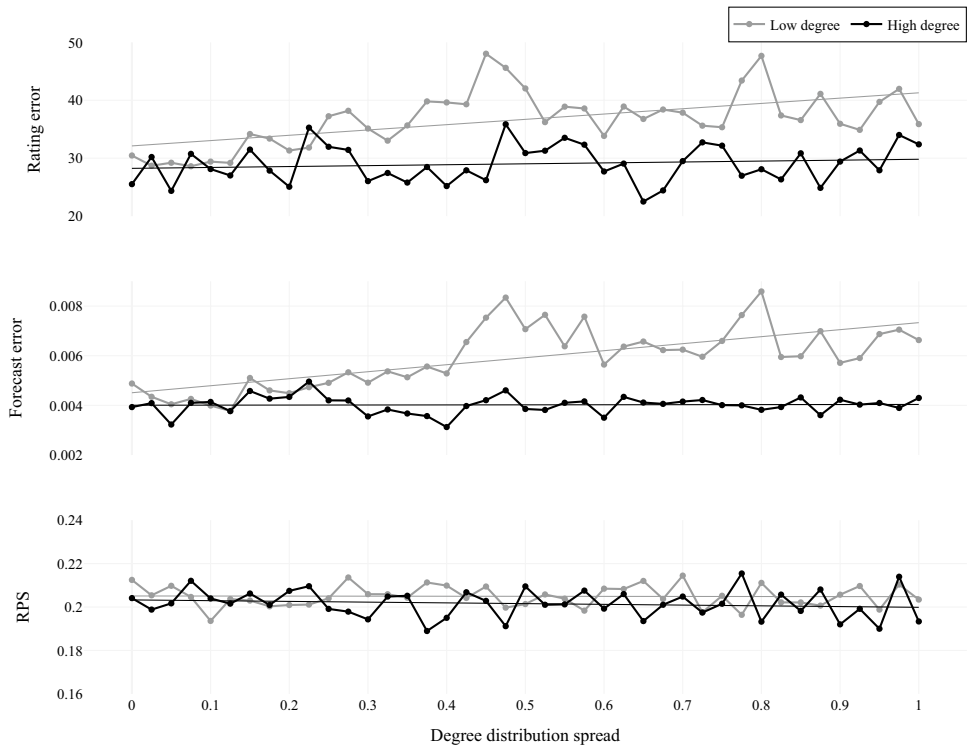


Fig. 7 Rating error, forecast error and RPS by degree distribution spread and degree group. Linear trend is marked as a thinner grey line



Dissemination between node rating error, division rating error and division-dependent rating error is presented in Fig. 10. The graph shows how the estimators achieve higher accuracy at node and division levels in networks

with higher inter-division connectivity. Additionally, in networks with higher inter-division connectivity, the distance between the rating error and the division-dependent rating error is lower.

Fig. 8 Optimal K -factor depending on the connectivity between divisions

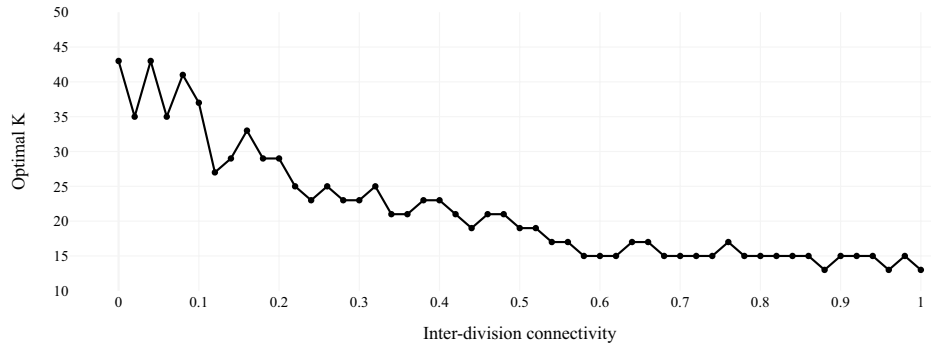


Fig. 9 Rating error and predictive value (actual forecast error and RPS) by inter-division connectivity. Linear trend is marked as a thinner grey line

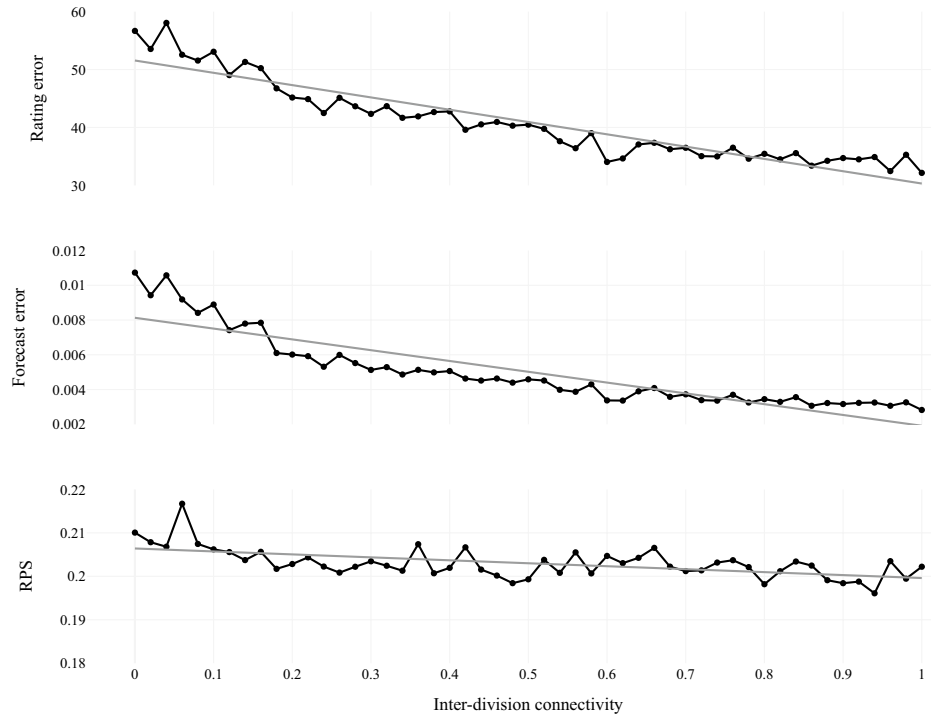
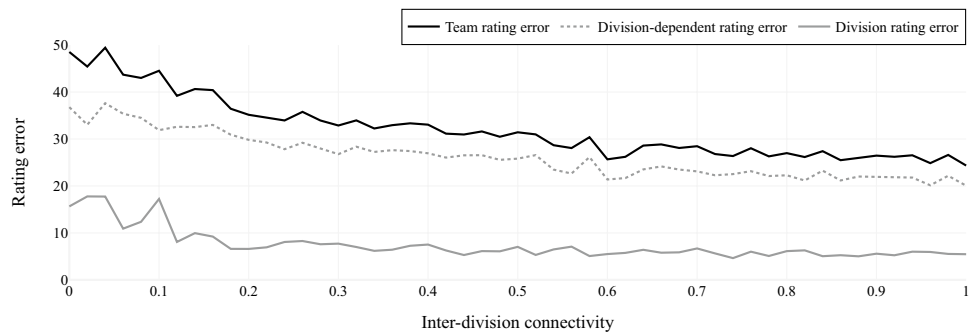


Fig. 10 Rating error at team level (overall team and division-dependent) and at division level by inter-division connectivity



4 Practical applications: a use case on sports divisions

The primary aim of this paper was to present the theoretical results of the simulation-based study to measure

the impact of the sports scheduling network in forecast models in sports. However, it is important to provide transferable insights into real-world data scenarios. In this section, a practical application motivated by the presented study highlights how network modularity could provide important modeling and optimization insights in the

sports forecasting process. Example of real-world datasets were presented in Sect. 1.2. This use case comprises data from European Football national leagues and international competitions in 2020/2021 season; 1826 matches were extracted from the five top leagues; English Premier League, French Ligue 1, German Bundesliga, Italian Serie A and Spanish First Division. In addition, 72 international matches were gathered from the UEFA Champions League and the UEFA Europa League. For this case study, only the international matches confronting teams from the listed national leagues were included, representing only the 4% of the total number of matches in the dataset. All data were obtained through (LLC SR).

The resulting network model contains five connected components. Only a few connections exist between the components which refer to the international confrontations. International matches are generally more challenging to model for a forecasting process in such scenarios due to the potential differences between national competitions and the lowest number of international matches. Network modularity challenges are also present in US sports, where tournaments are divided into divisions with different playing frequencies. While these dimensions are not directly observable in some other unstructured network models (i.e., in tennis), network analysis methods could reveal components within the network and motivate a similar procedure. As shown in Sect. 3.3, network modularity, its components and their inter-component edges seem to affect ratings optimization and accuracy. A possible network-aware modification of the ELO Rating could be based on differentiating between edges within the components (i.e., matches in national leagues) and edges between components (i.e., international competitions matches). Thus, this practical use case adjusts the ELO Rating algorithm to segregate between edge groups (i.e., national league matches and international matches). Finally, this adjusted ELO Rating is compared to the standard version.

The main results of this use case are presented in Table 2. The two ELO Rating implementations were added to the same data. The version ELO_{same} is a basic implementation of an ELO Rating with $c = 10$, $d = 400$, a home advantage

parameter $w = 50$ (Hvattum and Arntzen 2010), and a single K -factor calibrated by RPS. Despite current approaches containing more complex parametrizations, ELO_{same} is used as the benchmark for the state of the art calibration and evaluation of ELO Rating in research and practice. The adjusted version, ELO_{dif} , was implemented with the same parameters as ELO_{same} . However, in this case, ELO_{dif} contained two calibrated K -factors, one for matches within national leagues and one for international leagues matches. Both K -factors were calibrated by RPS. Despite being a basic use case, results already indicate the potential of such network-aware implementations in rating procedures. First, the two different K -factors had different optimal points. This finding validates the results found in Sect. 3.3 where the K -factor was affected by the increase in connections between components. Moreover, while the aggregated RPS of national league matches remained unaffected, results show a slight improvement in the RPS performance of the international league matches when using two K -factors. Thus, the ratings are more accurate in predicting international league matches than a single K -factor.

This practical application is subject to several limitations. The sample size is limited to a single natural year of competitions, while in most sports forecasting scenarios, sample sizes of at least three years are usually gathered. Consequently, the number of international matches only represents the 4% of the total number of matches. Due to the small sample size, no initialization of ratings or data split between the in-sample and the out-of-sample dataset was incorporated. Additionally, the leagues added to the study could be considered of a similar level of competitiveness, reducing the challenge of international matches. However, this use case aimed to demonstrate the applicability of this study's findings, and these limitations should be tackled to further structure real-world data studies. Despite the proposal of network-aware new methodologies being considered out of the scope for the present study, it is expected that the findings of this paper, in conjunction with the brief use case presented, will motivate and boost their research, development and application.

5 Discussion

Predictive models targeting the outcome of sports events frequently raise attention in several research disciplines and often tackle methods to properly estimate competitors' strengths or procedures to predict the result of certain events accurately. In this predictive context, the structure of the competition (i.e., how competitors are scheduled to compete) is yet assumed to be of lesser relevance. The aim of this study was to use a simulation-based approach to analyze the effects of competition schedules in the forecasting

Table 2 Results comparing the basic ELO Rating implementation ELO_{same} and the network-aware proposed implementation with two different K -factors ELO_{dif}

	Optimal K	RPS _{inter} ²	RPS _{intra} ³
ELO_{same}	42	0.2111	0.2349
ELO_{dif}	42,541	0.2111	0.2341

¹The first value is the optimal k for national league matches and the second value is the optimal k for international leagues matches

²Average RPS of the national leagues matches

³Average RPS of the international leagues matches

performance (i.e., model optimization, rating accuracy and predictive value) via network science. The analysis modeled a sports competition schedule as a network model with nodes representing the competitors and edges representing the confrontation between competitors and embedded a full forecasting process in the network model. Separate strands of literature can support the presented results and motivate further research. First, the literature on sports forecasting models (Hvattum and Arntzen 2010; Wunderlich and Memmert 2021) that has been neglecting to systematically study the possible influence of network structures. Second, the literature on optimal scheduling of sports tournaments that have used network science to apply the general conditions of a competition format while optimizing additional aspects such as traveling costs (Drexler and Knust 2007; Fry and Ohlmann 2012; Ribeiro 2012), but has not focused yet on predictive aspects of competition schedules.

Methodologically, the effects on rating error and forecast error were consistent in all analyses, meaning that rating inaccuracies directly translate to forecasting inaccuracies. This can be considered a direct consequence of the fact that in the simulation model, the forecasting process equals the true forecasting process; thus, no additional inaccuracy is introduced when obtaining probabilities from the estimators instead of the true rating values. The same applies to the results with regard to RPS; however, effects on RPS appear to be not as clear as effects on forecast error. This can be attributed to the fact that rating and forecast errors, by definition, are not prone to the inherent randomness in observed results. As such, the additional noise driven by observed results in the RPS can obfuscate the true mechanisms and have been documented as one of the limitations of this evaluation metric in recent studies (Hubáček et al. 2019b). These results highlight three advantages of using the present setup of artificial data creation that are not given in real-world data: First, the data characteristics can be fully controlled, which made it possible to successfully exclude errors in forecasting from ratings in the analysis. Second, the true ratings and true forecasts are known and thus available for comparison. Third, the inherent randomness in observed results is excluded from analysis by using rating error and forecast error values.

Theoretical insights can be drawn from analyzing the three different network properties. Regarding density, no significant effect of the network property on rating or forecast accuracy can be found. The lack of impact from density is considered a surprising result as a higher density is associated with higher transitivity, i.e., if participant A faces participant B in a dense network, there is a good chance both teams already faced participant C, and these results are known. In a less dense network, this is rather unlikely, and as such, less information from transitivity is available in the network. The results suggest that the ELO rating does

not seem to take advantage of higher transitivity in denser networks. Regarding the degree distribution of the networks, results show clear evidence that rating and forecast error are predominantly driven by those nodes (competitors) with a very low degree (low number of matches). This is an understandable result, as a low degree is associated with less information, and as such, there is less possibility for the rating estimation to converge towards the true rating. Moreover, results prove that the optimal specification of the model (in terms of the K -factor) depends on the degree of the nodes, which becomes a large discrepancy in the case of high degree distributions. Driven by the smaller amount of information for low degree nodes, it is comprehensible that any new information should be given a higher weight than new information for high degree nodes. The third network property investigated was the network modularity; results revealed that networks with higher inter-division connectivity require lower K -factor and show lower rating error, forecasting error and RPS. However, in contrast to the other two properties, the number of edges per node increases for increasing modularity. Therefore, the increased information from additional confrontations could also explain the decrease of inaccuracies. It is also explainable that a low number of connections between divisions generates a bias in estimating the overall rating of nodes in the divisions as the overall division rating is assumed to be the same for each division. The present study then further explored the source of rating inaccuracies by comparing the node rating error with its division rating error. Results show that the estimation of overall ratings across divisions and the estimation of node ratings inside a division contributes to improving rating accuracy. In contrast, the overall division ratings seem to be a serious challenge in networks with few inter-division connections.

This combination of findings provides support for improving and developing new forecasting models. This study aimed at identifying the potential for improvement of predictive ratings. In general, the study has confirmed that network characteristics play a role in forecasting and should be considered in the model choice. Analysis of density has shown that an established rating procedure such as the ELO rating does not benefit from transitivity in the networks. Therefore, future rating approaches should try to specifically exploit transitivity by involving indirect comparisons of participants in the rating procedure. A further advantage of using the micro and macro network structures could be to update competitors' values even though the competitors did not directly engage in a confrontation. Analysis of degree distribution has revealed that different parts of a network would require different model specifications in terms of optimal k . These results could relate to sports schedules with a high disparity in competitors' occurrences and how to deal with the new competitors (i.e., new-ranked tennis

players). In the current state of the art in forward-looking rating systems (Coulom 2008), such as the ELO Rating, the K -factor is usually referred to as the rating adjustment. Too high values for the K -factor create sensitive and unstable ratings, while low values create ratings that generally will not respond quickly enough to a competitor's strength evolution. A stable K -factor is the standard approach, although variable K -factors have been already discussed in the literature, e.g., depending on the experiences of the rated competitor (Bester and von Maltitz 2013). Although K -factor appears to have a limited influence on the ranking of competitors in the long term (Albers and de Vries 2001), the present results suggest that a variable K -factor taking the network characteristics into account would be likely to improve rating accuracy. Finally, the analysis of modularity has shed light on the issue of overall divisions rating estimations, which seem to be inaccurate when connections between divisions are very limited. To solve this issue, some potential might lie in using increased K -factors for the sparse connections between leagues or indirectly updating all teams in a league based on the results of an inter-league match. The latter results have practical implications in competitions containing divisions (i.e., US Sports or European tournaments considering national leagues teams). Results support other studies proposing special procedures to deal with different division strengths and inter-division confrontations, e.g., introducing competitive balance coefficients to predict the outcome of international football games (Halicioglu 2009). A practical application is showcased to justify further and motivate the adoption of new network-aware methods in the sports forecasting process. This brief use case using real data illustrates the potential lines of improvement that network-aware rating procedures and predictive models could achieve in concordance with the theoretical findings presented in this paper.

The present study is subject to limitations that could be a fruitful field for further research. First, the study is limited to three basic network properties representing a very small selection of possible metrics from network science that might be worth investigating. Moreover, for feasible data generation and interpretability, these properties have been considered in three separate analyses while not studying joint effects of the properties. The authors, therefore, propose that future studies can investigate further properties and consider the joint evaluation, including network entropy metrics (Omar et al. 2020). The present analysis is also limited to the specifications of the simulated environment, which introduces a single rating and forecasting model. While ELO rating (Glickman and Jones 1999) and ordered logistic regression (Hvattum and Arntzen 2010) are common and established methods, further research should particularly consider additional rating methods to validate the generalizability of results. Due to the use of artificial data generation, the lack of any real-world data evidence might

be another point of criticism. However, artificial data was intentionally chosen as an analysis like this would simply not be possible with real-world data. The available real-world networks from various sports and competitions represent a very small fraction of the artificially generated network specifications in this study. As such, a direct transferability of the results to real-world data by taking a similar approach is impossible. At the same time, theoretical insights have a very limited value if not being transferable to real-world problems. Consequently, further research should build on the present results, focusing on using the network structure for varying K -factors and indirect comparisons to test model improvements based on real-world datasets.

Funding Open Access funding enabled and organized by Projekt DEAL. This research was supported by a grant from the German Research Council (DFG, Deutsche Forschungsgemeinschaft) to D.M. (Grant ME 2678/29.1).

Availability of data and materials Data are available on reasonable request.

Declarations

Conflict of interest (check journal-specific guidelines for which heading to use) No potential competing interest was reported by the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Albers PCH, de Vries H (2001) Elo-rating as a tool in the sequential estimation of dominance strengths. *Anim Behav* 61(2):489–495. <https://doi.org/10.1006/anbe.2000.1571>
- Angelini G, de Angelis L (2019) Efficiency of online football betting markets. *Int J Forecast* 35(2):712–721. <https://doi.org/10.1016/j.ijforecast.2018.07.008>
- Baker RD, McHale IG (2013) Forecasting exact scores in national football league games. *Int J Forecast* 29(1):122–130. <https://doi.org/10.1016/j.ijforecast.2012.07.002>
- Bansal S, Read J, Pourbohloul B (2010) Lauren Ancel Meyers: the dynamic nature of contact networks in infectious disease epidemiology. *J Biol Dyn* 4(5):478–489. <https://doi.org/10.1080/17513758.2010.503376>

- Berrar D, Lopes P, Davis J, Dubitzky W (2019) Guest editorial: special issue on machine learning for soccer. *Mach Learn* 108(1):1–7. <https://doi.org/10.1007/s10994-018-5763-8>
- Bester DW, von Maltitz MJ (2013) Introducing momentum to the ELO rating system. University of the Free State, Department of Mathematical Statistics and Actuarial Science
- Brandes U, Robins G, McCranie A, Wasserman S (2013) What is network science? *Netw Sci* 1:1–15. <https://doi.org/10.1017/nws.2013.2>
- Buldú JM, Busquets J, Martínez JH, Herrera-Diestra JL, Echegoyen I, Galeano J, Luque J (2018) Using network science to analyse football passing networks: dynamics, space, time, and the multi-layer nature of the game. *Front Psychol* 9:1900. <https://doi.org/10.3389/fpsyg.2018.01900>
- Bunker RP, Thabtah F (2019) A machine learning framework for sport result prediction. *Appl Comput Inform* 15(1):27–33. <https://doi.org/10.1016/j.aci.2017.09.005>
- Constantinou AC, Fenton NE (2012) Solving the problem of inadequate scoring rules for assessing probabilistic football forecast models. *J Quant Anal Sports*. <https://doi.org/10.1515/1559-0410.1418>
- Coulom R (2008) Whole-history rating: a Bayesian rating system for players of time-varying strength. In: Hutchison D, Kanade T, Kitter J, Kleinberg JM, Mattern F, Mitchell JC, Naor M, Nierstrasz O, Pandu Rangan C, Steffen B, Sudan M, Terzopoulos D, Tygar D, Vardi MY, Weikum G, van den Herik HJ, Xu X, Ma Z, Winands MHM (eds) *Computers and games*. Lecture notes in computer science, vol 5131, Springer, Berlin Heidelberg, pp 113–124. <https://doi.org/10.1007/978-3-540-87608-3>
- de Saá Guerra Y, Martín González JM, Sarmiento Montesdeoca S, Rodríguez Ruiz D, Garcaí-Rodríguez A, García-Manso JM (2012) A model for competitiveness level analysis in sports competitions: application to basketball. *Physica A* 391(10):2997–3004. <https://doi.org/10.1016/J.PHYSA.2012.01.014>
- de Werra D (1985) On the multiplication of divisions: the use of graphs for sports scheduling. *Networks* 15(1):125–136. <https://doi.org/10.1002/net.3230150110>
- Drexl A, Knust S (2007) Sports league scheduling: graph- and resource-based models. *Omega* 35(5):465–471. <https://doi.org/10.1016/j.omega.2005.08.002>
- Duch J, Waitzman JS, Amaral LAN (2010) Quantifying the performance of individual players in a team activity. *PLoS One* 5(6):1–7. <https://doi.org/10.1371/journal.pone.0010937>
- Forrest D, Sanz I, Tena JD (2010) Forecasting national team medal totals at the summer Olympic games. *Int J Forecast* 26(3):576–588. <https://doi.org/10.1016/j.ijforecast.2009.12.007>
- Franz M, Lopes CT, Huck G, Dong Y, Sumer O, Bader GD (2016) Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics (Oxford, England)* 32(2):309–311. <https://doi.org/10.1093/bioinformatics/btv557>
- Fry MJ, Ohlmann JW (2012) Introduction to the special issue on analytics in sports, part ii: sports scheduling applications. *Interfaces* 42(3):229–231. <https://doi.org/10.1287/inte.1120.0632>
- Garnica-Caparrós M, Memmert D, Wunderlich F (2022) Artificial data in sports forecasting: a simulation framework for analysing predictive models in sports. *Inf Syst E-Bus Manage*, 1–30
- Ghoniem A, Serali HD (2010) Models and algorithms for the scheduling of a doubles tennis training tournament. *J Oper Res Soc* 61(5):723–731. <https://doi.org/10.1057/jors.2008.190>
- Glickman M, Jones A (1999) Rating the chess rating system. *Chance* 12(5):21–28
- Goes F, Kempe M, Lemmink K (2019) Predicting match outcome in professional dutch football using tactical performance metrics computed from position tracking data. pp 2516–2314. <https://doi.org/10.29007/4JJB>
- Groll A, Heiner J, Schauburger G, Uhrmeister J (2020) Prediction of the 2019 IHF world men's handball championship—a sparse Gaussian approximation model. *J Sports Anal* 6(3):187–197. <https://doi.org/10.3233/JSA-200384>
- Hagberg Aric A, Daniel A. Schult, Pieter J (2008) Swart: exploring network structure, dynamics, and function using networkx. In: Varoquaux G, Vaught T, Millman J (eds) *Proceedings of the 7th python in science conference*, Pasadena, CA USA, pp 11–15
- Halicioglu F (2009) Research on the prediction of the likely winners of the Euro 2008 football tournament. *J Quant Anal Sports*. <https://doi.org/10.2202/1559-0410.1137>
- Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, Kern R, Picus M, Hoyer S, van Kerkwijk MH, Brett M, Haldane A, Del Río JF, Wiebe M, Peterson P, Gérard-Marchant P, Sheppard K, Reddy T, Weckesser W, Abbasi H, Gohlke C, Oliphant TE (2020) Array programming with numpy. *Nature* 585(7825):357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Horvat T, Job J (2020) The use of machine learning in sport outcome prediction: a review. *WIREs Data Min Knowl Discov*. <https://doi.org/10.1002/widm.1380>
- Hubáček O, Šourek G, Železný F (2019a) Learning to predict soccer results from relational data with gradient boosted trees. *Mach Learn* 108(1):29–47. <https://doi.org/10.1007/s10994-018-5704-6>
- Hubáček O, Šourek G, Železný F (2019b) Exploiting sports-betting market using machine learning. *Int J Forecast* 35(2):783–796. <https://doi.org/10.1016/j.ijforecast.2019.01.001>
- Hvattum LM, Arntzen H (2010) Using ELO ratings for match result prediction in association football. *Int J Forecast* 26(3):460–470. <https://doi.org/10.1016/j.ijforecast.2009.10.002>
- Jahangirian M, Naseer A, Stergioulas L, Young T, Eldabi T, Brailsford S, Patel B, Harper P (2010) Simulation in health-care: lessons from other sectors. *Oper Res* 12(1):45–55. <https://doi.org/10.1007/S12351-010-0089-8>
- Jayanth SB, Anthony A, Abhilasha G, Shaik N, Srinivasa G (2018) A team recommendation system and outcome prediction for the game of cricket. *J Sports Anal* 4(4):263–273. <https://doi.org/10.3233/JSA-170196>
- Koopman SJ, Lit R (2019) Forecasting football match results in national league competitions using score-driven time series models. *Int J Forecast* 35(2):797–809. <https://doi.org/10.1016/j.ijforecast.2018.10.011>
- Kovalchik SA (2016) Searching for the goat of tennis win prediction. *J Quant Anal Sports*. <https://doi.org/10.1515/jqas-2015-0059>
- Lessmann S, Sung M-C, Johnson JEV (2010) Alternative methods of predicting competitive events: an application in horserace betting markets. *Int J Forecast* 26(3):518–536. <https://doi.org/10.1016/j.ijforecast.2009.12.013>
- LLC SR Football Statistics and History (no date). Retrieved May 9, 2022, from <https://fbref.com/>
- Manner H (2016) Modeling and forecasting the outcomes of nba basketball games. *J Quant Anal Sports*. <https://doi.org/10.1515/jqas-2015-0088>
- McHale I, Swartz T (2019) Editorial: forecasting in sports. *Int J Forecast* 35(2):710–711. <https://doi.org/10.1016/j.ijforecast.2019.01.002>
- Medina Pablo, Carrasco Sebastián, Rogan José, Montes Felipe, Meisel Jose D, Lemoine Pablo (2021) Carlos Lago Peñas, Juan Alejandro Valdivia: Is a social network approach relevant to football results? *Chaos, Solitons Fractals* 142:110369. <https://doi.org/10.1016/j.chaos.2020.110369>
- Miljković D, Gajić L, Kovačević A, Konjović Z (2010) The use of data mining for basketball matches outcomes prediction. In: *IEEE 8th international symposium on intelligent systems and informatics*, pp 309–312. <https://doi.org/10.1109/SISY.2010.5647440>

- Newman MEJ (2003) The structure and function of complex networks. *SIAM Rev* 45(2):167–256. <https://doi.org/10.1137/S003614450342480>
- Omar YM, Plapper P (2020) A survey of information entropy metrics for complex networks. *Entropy (Basel, Switzerland)*. <https://doi.org/10.3390/e22121417>
- Pachur T, Biele G (2007) Forecasting from ignorance: the use and usefulness of recognition in lay predictions of sports events. *Acta Physiol (Oxf)* 125(1):99–116. <https://doi.org/10.1016/j.actpsy.2006.07.002>
- Park J, Newman MEJ (2005) A network-based ranking system for us college football. *J Stat Mech Theory Exp* 2005(10):10014. <https://doi.org/10.1088/1742-5468/2005/10/P10014>
- Pitoski D, Lampoltshammer TJ, Parycek P (2021) Human migration as a complex network: appropriate abstraction, and the feasibility of network science tools. In: Haber P, Lampoltshammer T, Mayr M, Plankensteiner K (eds) *Data science—analytics and applications*. Springer Fachmedien Wiesbaden, Wiesbaden, pp 113–120
- Ribeiro CC (2012) Sports scheduling: problems and applications. *Int Trans Oper Res* 19(1–2):201–226. <https://doi.org/10.1111/j.1475-3995.2011.00819.x>
- Rossi A, Pappalardo L, Cintia P, Iaia FM, Fernández J, Medina D (2018) Effective injury forecasting in soccer with GPS training data and machine learning. *PLoS One* 13(7):0201264. <https://doi.org/10.1371/journal.pone.0201264>
- Seidl T, Cherukumudi A, Harnett A, Carr P, Lucey P (2018) Bhost-gusters: realtime interactive play sketching with synthesized nba defenses. In: *Proceeding of the 12th MIT Sloan sports analytics conference 2018*. MIT, Boston, MA
- Shi J, Tian XY (2020) Learning to rank sports teams on a graph. *Appl Sci (Switzerland)* 10(17):1–10. <https://doi.org/10.3390/app10175833>
- Stefani R (2011) The methodology of officially recognized international sports rating systems. *J Quant Anal Sports*. <https://doi.org/10.2202/1559-0410.1347>
- van der Hofstad R (2017) *Random graphs and complex networks*. Cambridge series in statistical and probabilistic mathematics, vol 43. Cambridge University Press, Cambridge. <https://doi.org/10.1017/9781316779422>
- van Rossum G, Drake FL (2009) *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA
- Wunderlich F, Memmert D (2020) Are betting returns a useful measure of accuracy in (sports) forecasting? *Int J Forecast* 36(2):713–722. <https://doi.org/10.1016/J.IJFORECAST.2019.08.009>
- Wunderlich F, Memmert D (2021) Forecasting the outcomes of sports events: a review. *Eur J Sport Sci* 21(7):944–957. <https://doi.org/10.1080/17461391.2020.1793002>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.